

# Comparative Study of Lectin Domains in Model Species: New Insights into Evolutionary Dynamics

Sofie Van Holle<sup>1</sup>, Kristof De Schutter<sup>1,2</sup>, Lore Eggermont<sup>1</sup>, Mariya Tsaneva<sup>1</sup>, Liuyi Dang<sup>1</sup> and Els J.M. Van Damme<sup>1,\*</sup>

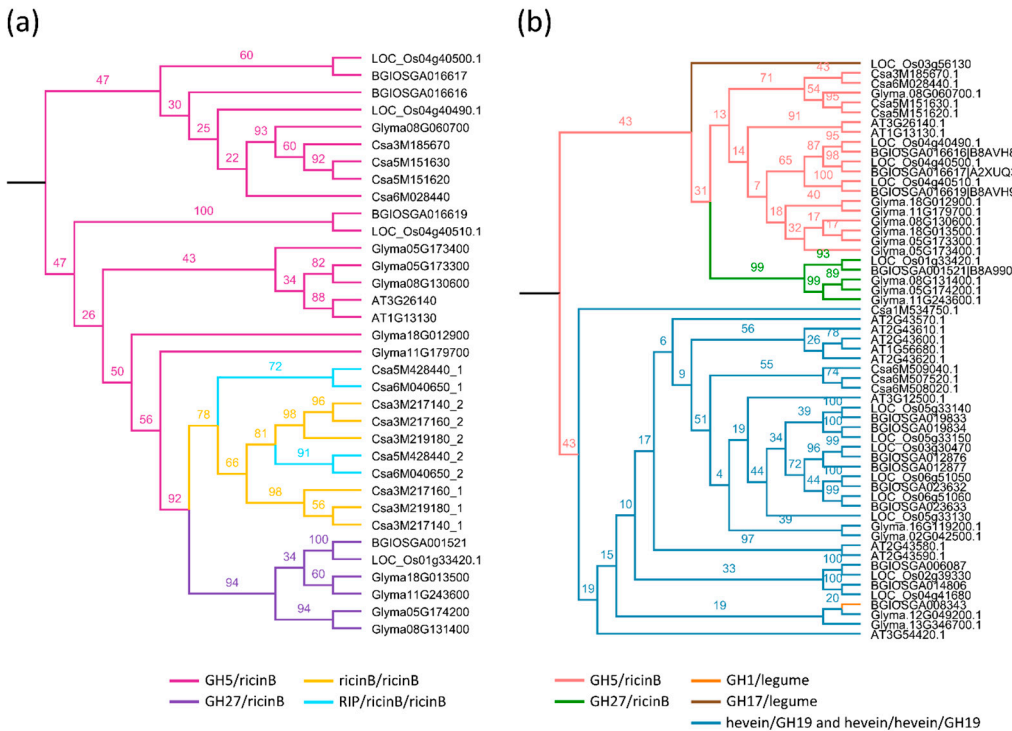
## Supplementary Materials:

**Table S1.** Relationship between the number of lectin genes and genome size, chromosome number and protein-coding transcripts

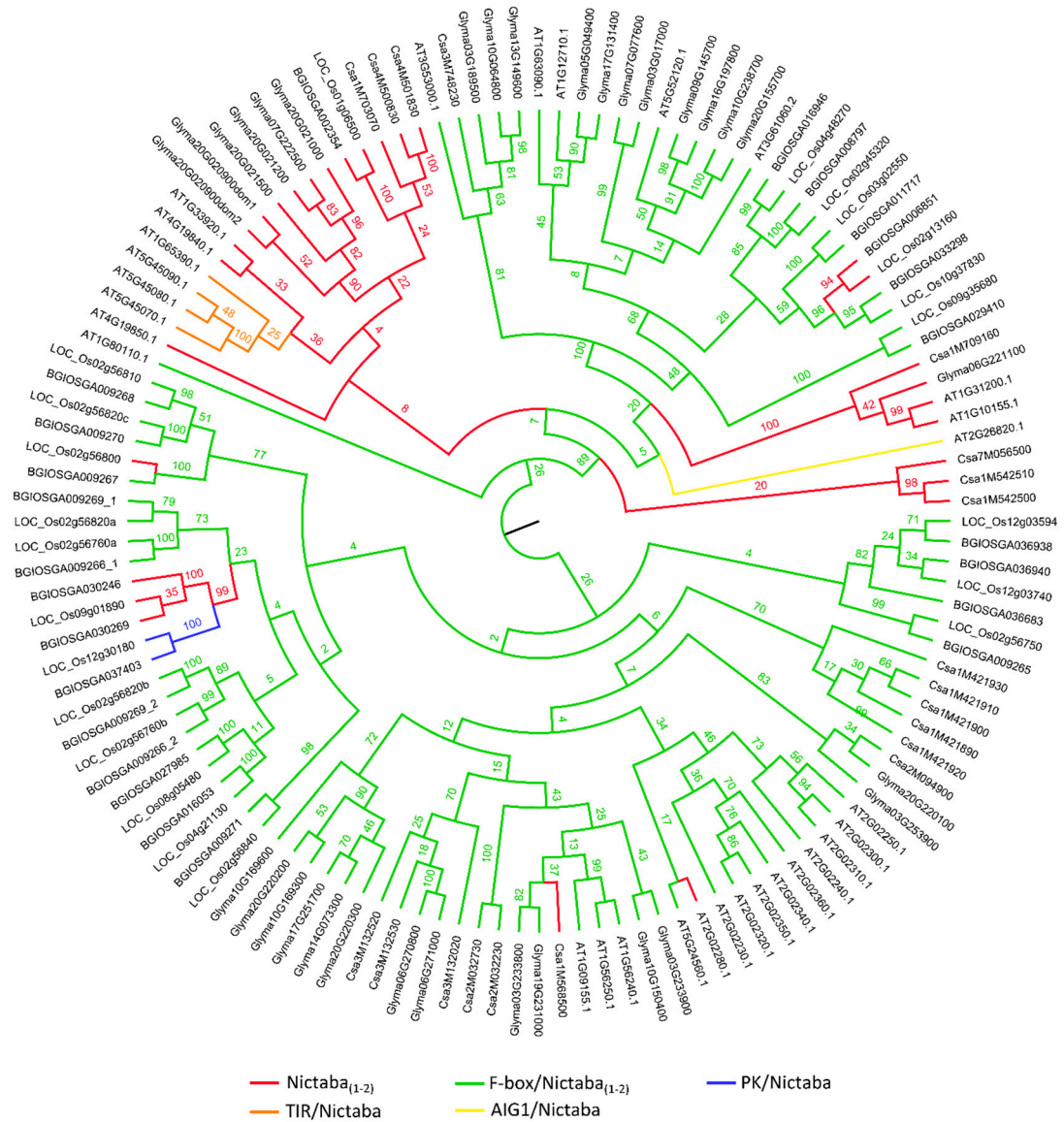
	<i>Arabidopsis</i>	Soybean	Cucumber	Rice ( <i>japonica</i> )	Rice ( <i>indica</i> )
Total number of lectin genes	216	368	146	329	278
Genome size (Mb)	135	1,115	367	385	389
Chromosome number	5	20	7	12	12
Protein-coding transcripts	35,386	88,647	32,528	52,424	56,284
Ratio lectins/genome size	1.60	0.33	0.40	0.85	0.71
Percentage lectins/protein-coding transcripts	0.61	0.42	0.45	0.62	0.49
Number of whole genome duplications	3	3	1	3	3

**Table S2.** Number of all identified domain combinations containing lectin motifs in the species under study. Numbers between brackets indicate the number of repeats.

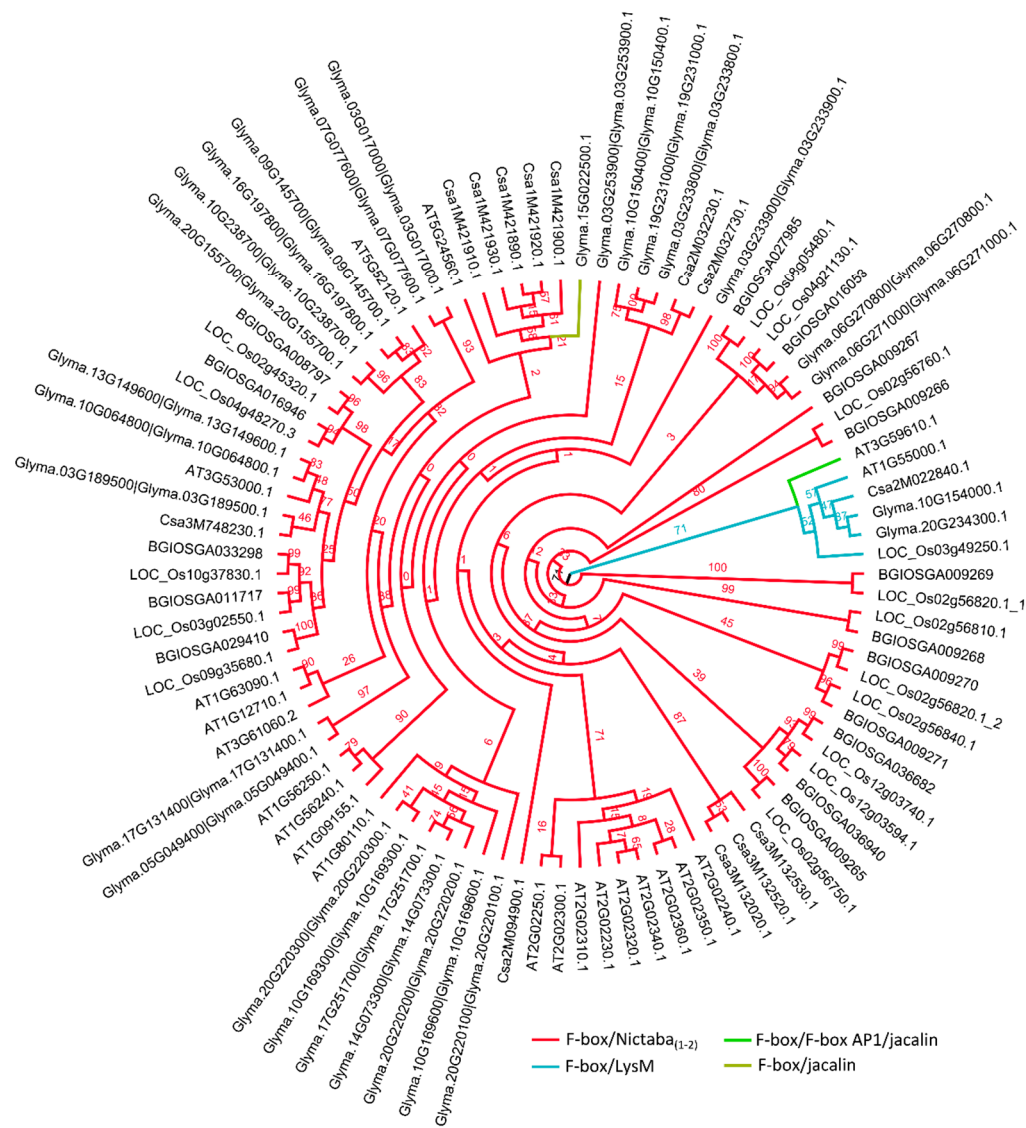
Shown in the excel of Supplementary Table S2.



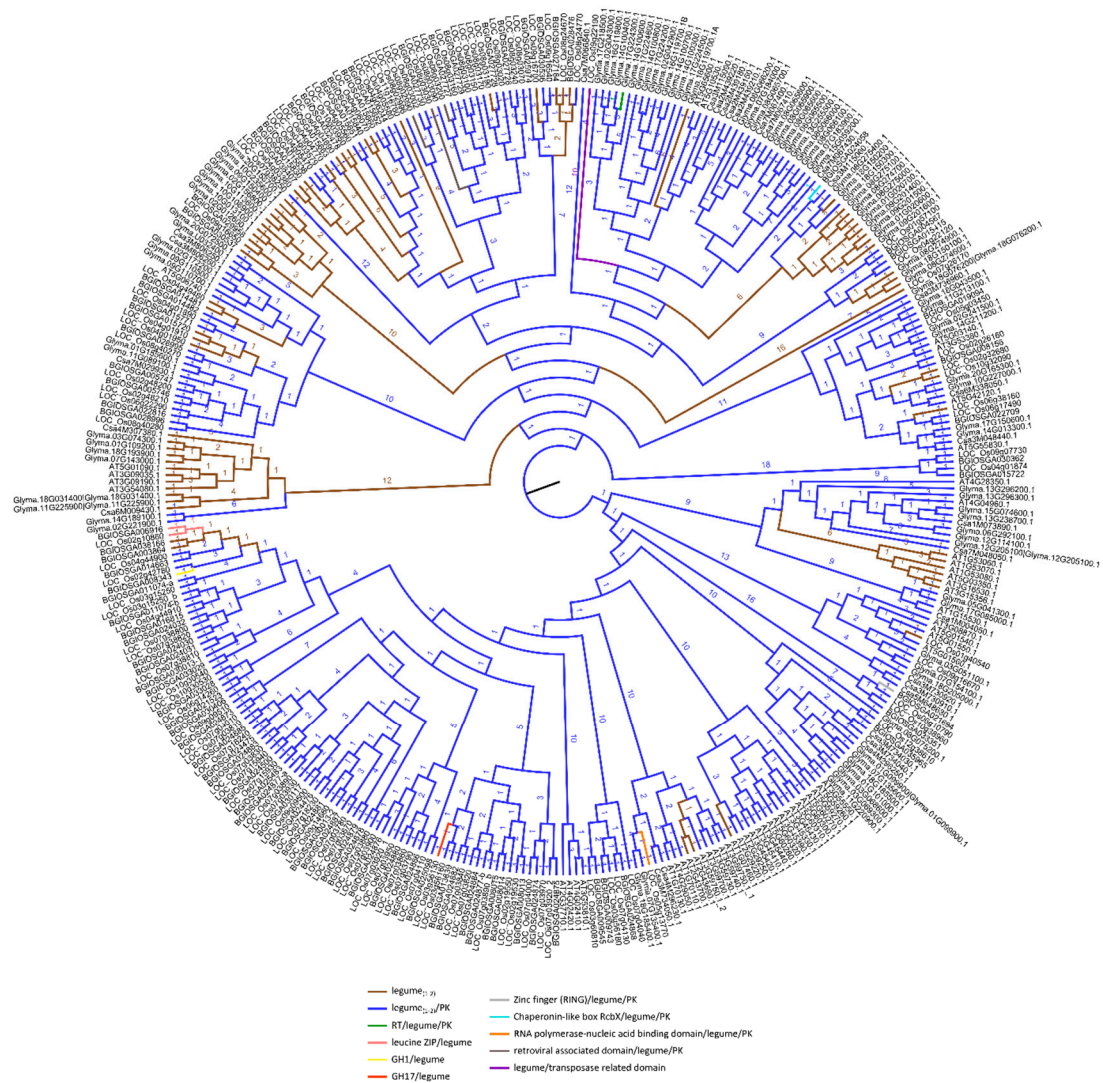
**Figure S1.** Phylogenetic relationships of all ricin B (a) and GH (b) domain sequences from *Arabidopsis* (AT), soybean (Glyma), cucumber (Csa), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length ricin B (a) and GH (b) protein sequences.



**Figure S2.** Phylogenetic relationships of the Nictaba domain sequences from *Arabidopsis* (AT), soybean (*Glyma*), cucumber (*Csa*), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length Nictaba protein sequences.



**Figure S3.** Phylogenetic relationships of the F-box domain sequences from *Arabidopsis* (AT), soybean (*Glyma*), cucumber (*Csa*), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length F-box protein sequences.

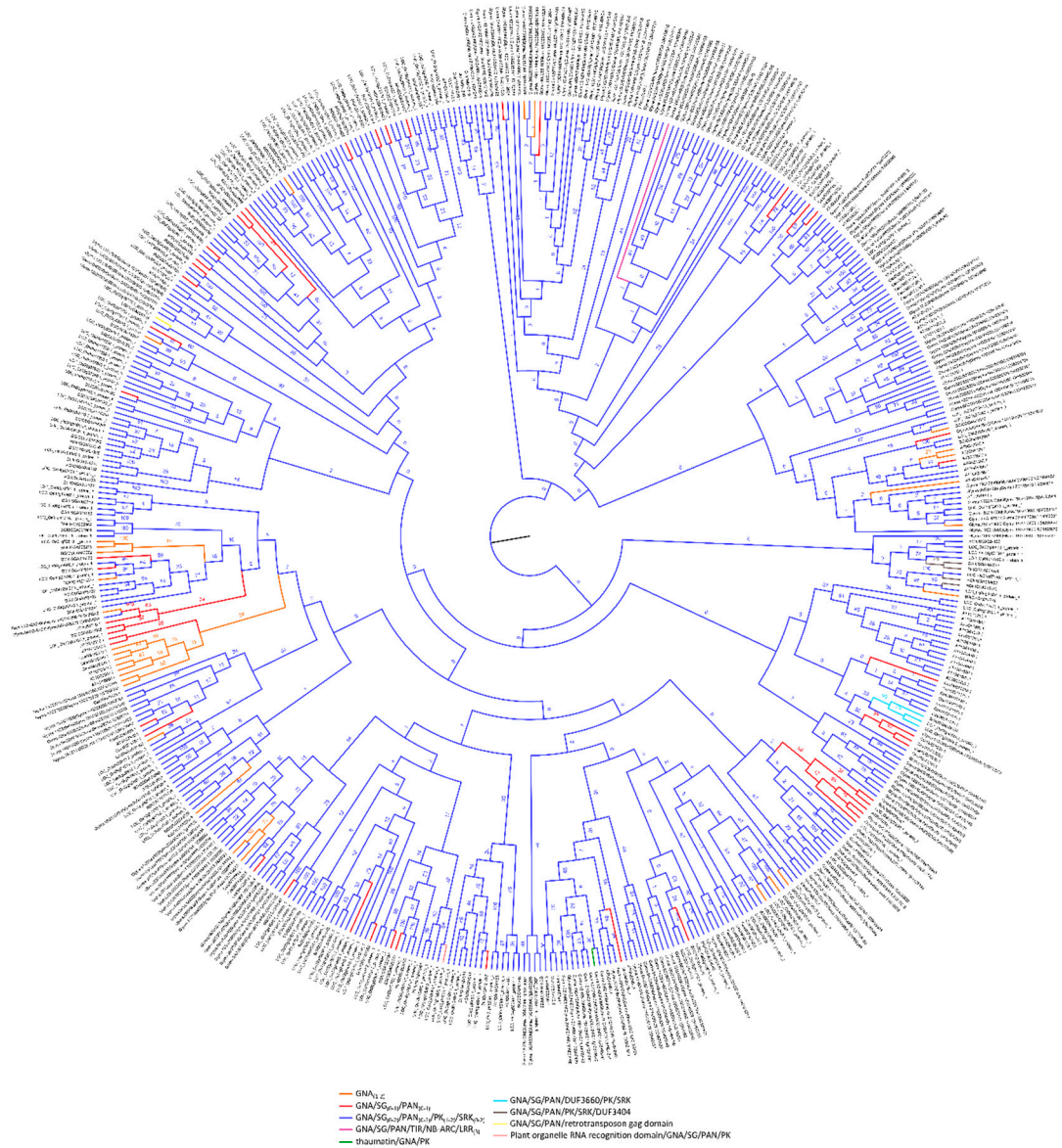


**Figure S4.** Phylogenetic relationships of the legume lectin domain sequences from *Arabidopsis* (AT), soybean (*Glyma*), cucumber (*Csa*), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length protein sequences.



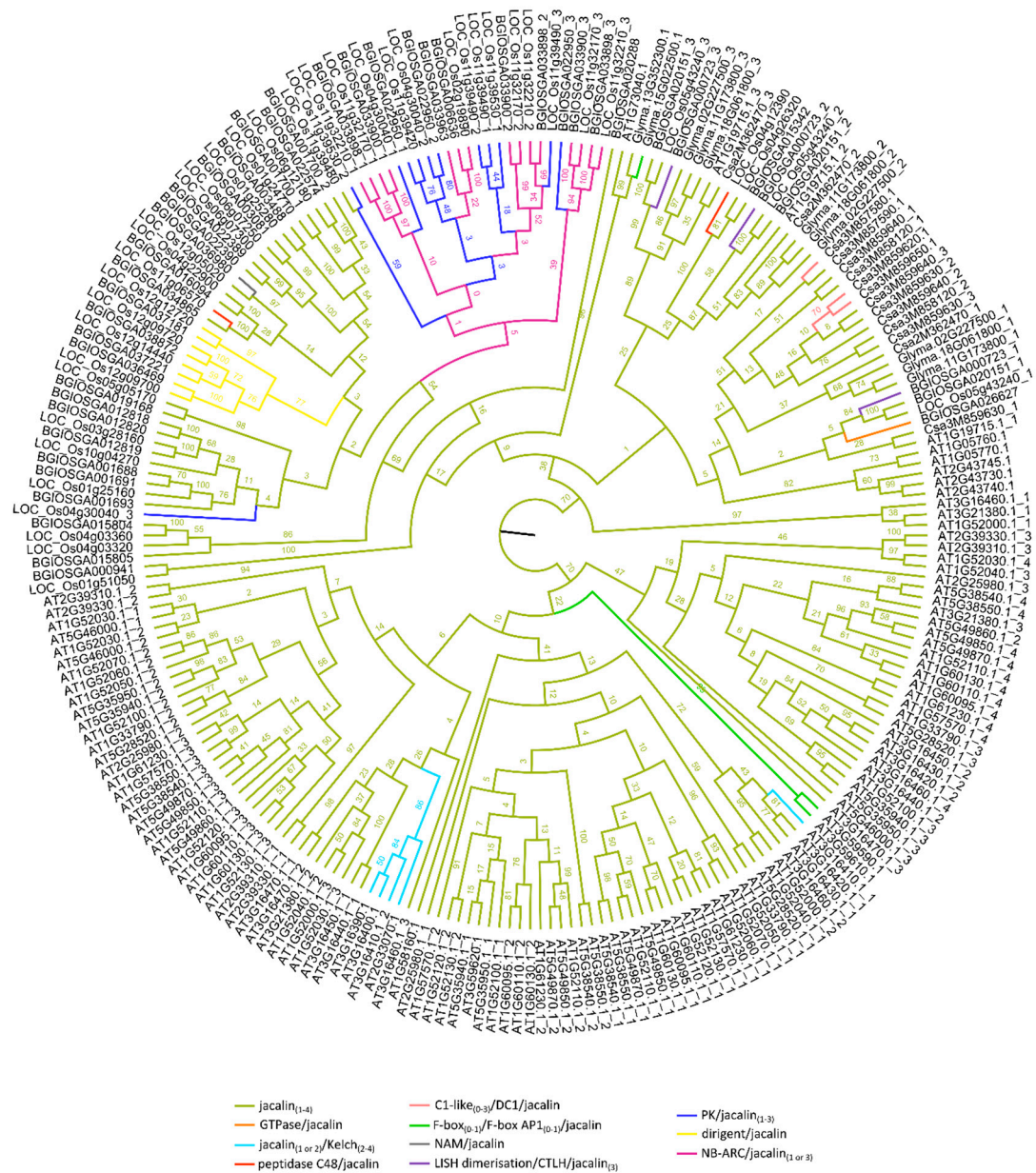


**Figure S5.** Phylogenetic relationships of the PK domain sequences from *Arabidopsis* (AT), soybean (Glyma), cucumber (Csa), and rice (*japonica*: LOC\_Os). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length PK sequences.

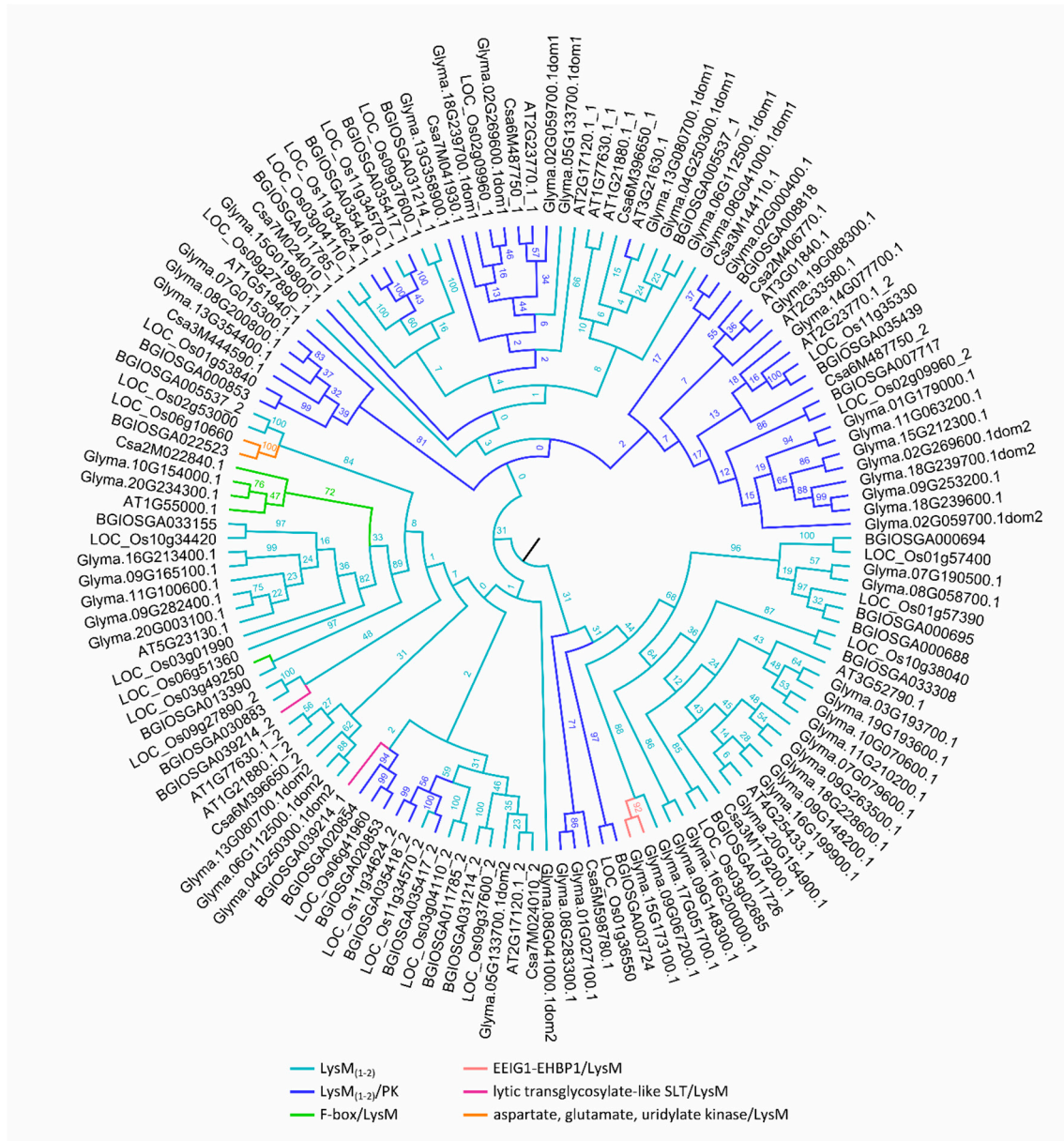


**Figure S6.** Phylogenetic relationships of the GNA domain sequences from *Arabidopsis* (AT), soybean (Glyma), cucumber (Csa), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length GNA lectin sequences.



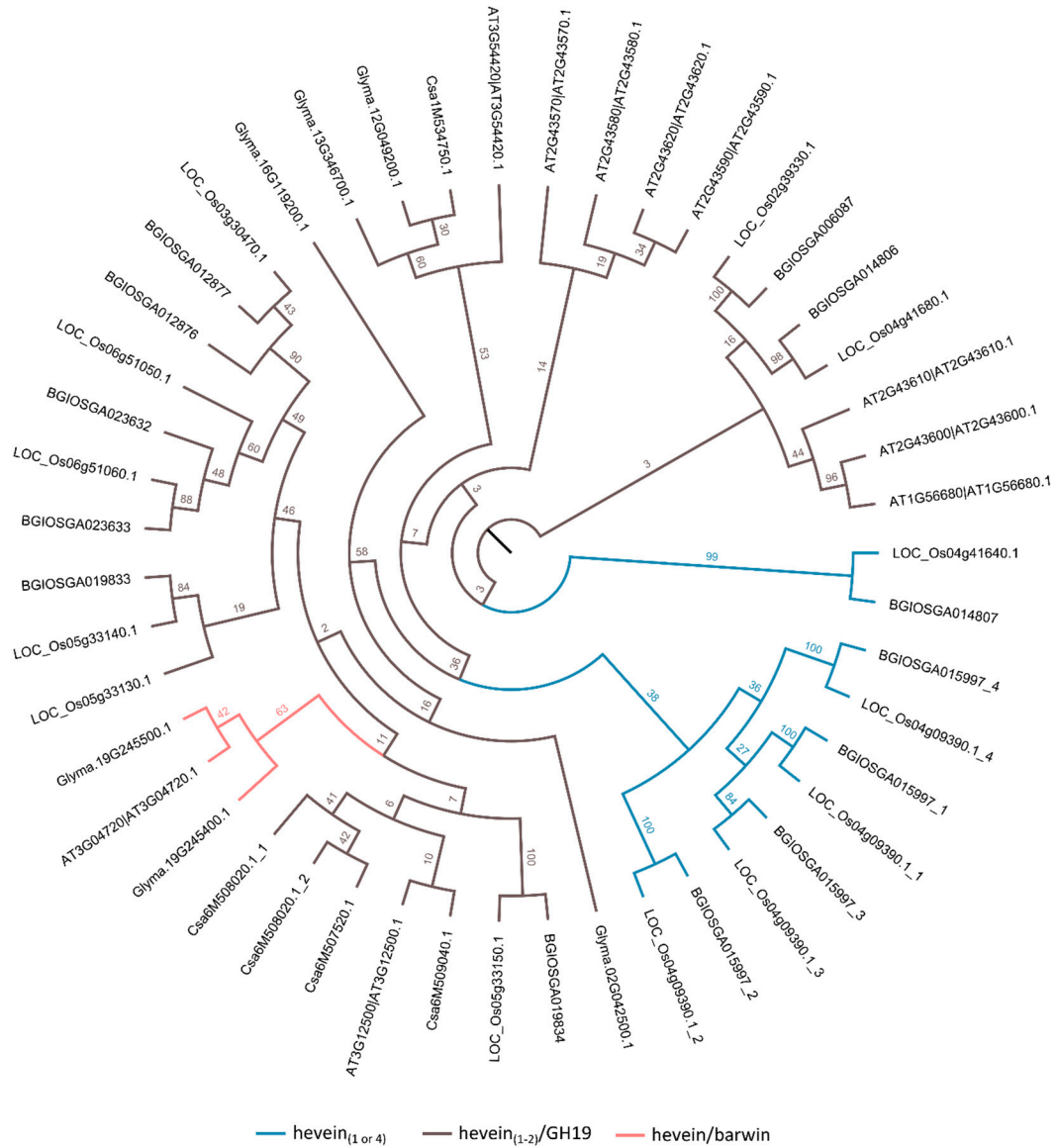


**Figure S7.** Phylogenetic relationships of the JRL domain sequences from *Arabidopsis* (AT), soybean (*Glyma*), cucumber (*Csa*), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length lectin sequences.



**Figure S8.** Phylogenetic relationships of the LysM domain sequences from *Arabidopsis* (AT), soybean (Glyma), cucumber (Csa), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length LysM sequences.





**Figure S9.** Phylogenetic relationships of the hevein domain sequences from *Arabidopsis* (AT), soybean (*Glyma*), cucumber (*Csa*), and rice (*japonica*: LOC\_Os, *indica*: BGIOGA). The numbers indicate the bootstrap values and the colored branches correspond to the different domain architectures of the full-length hevein sequences.