



**Supplemental Fig. S5.** Conservation of XP<sub>1</sub>, XP<sub>3</sub> and XP<sub>4</sub> glycomotifs in MAAB output by 1KP group for class 1: GPI-AGPs, class 2: CL-EXTs and class 24: <15% HRGPs. There is a strong bias for AP<sub>1,3,4</sub>, SP<sub>1,3,4</sub> and TP<sub>1,3,4</sub> glycomotifs in GPI-AGPs, whereas CL-EXTs show a strong bias for SP<sub>3</sub> and SP<sub>4</sub>, and moderate bias for SP<sub>1</sub> over KP and HP motifs (rarely seen in GPI-AGPs). The few class 2 (CL-EXTs) sequences identified in green algae and Chromista are likely sporadic contaminants based on the low mean number of CL-EXTs (0.1 (green algae) and 0.2 (Chromista)) (see Fig. 2C, Johnson et al., companion paper), low detection rates (see Fig. 2C and Supplemental Table S2, Johnson et al., companion paper) and BLASTp analysis revealing high sequence similarity to vascular plant CL-EXTs (data not shown). Class 24 sequences with <15% known HRGP motifs have a broader spectrum of XP<sub>1</sub>, XP<sub>3</sub> and XP<sub>4</sub> motifs, with a weak preference for SP<sub>n</sub> and LP<sub>n</sub> motifs. Refer to Fig. 2B (Johnson et al., companion paper) for the number of datasets per 1KP group. Data for charophycean green algae (CGA; Charales, Coleochaetales, Zygnematales, Klebsormidiales, Chlorokybales, Mesostimatales), a subset of the 1KP green algae group of chlorophytes and algal streptophytes, is also reported. Transformed data (square root) was used for heat maps.