SUPPORTING INFORMATION FOR

The maize pathogen *Ustilago maydis* secretes glycoside hydrolases and carbohydrate oxidases directed towards components of the fungal cell wall

Jean-Lou Reyre[1,2], Sacha Grisel[1,3], Mireille Haon[1,3], David Navarro[1,4], David Ropartz[5-6], Sophie Le Gall[5-6], Eric Record[1], Giuliano Sciara[1], Olivier Tranquet[1], Jean-Guy Berrin[1,3]* and Bastien Bissaro[1]*

[1] INRAE, Aix Marseille University, UMR1163 Biodiversité et Biotechnologie Fongiques, F-13009, Marseille, France
[2] IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, F-92852 Rueil-Malmaison, France
[3]INRAE, Aix Marseille University, 3PE platform, F-13009, Marseille, France
[4]INRAE, Aix Marseille University, CIRM-CF, F-13009, Marseille, France
[5]INRAE, UR1268 BIA, F-44300, Nantes, France
[6]INRAE, PROBE research infrastructure, BIBS facility, F-44300, Nantes, France

*corresponding authors:
Jean-Guy Berrin (jean-guy.berrin@inrae.fr)
Bastien Bissaro (bastien.bissaro@inrae.fr)

This file includes:

1. Supplementary Figure 1 to 21 and Supplementary table 1 to 2
2. Full abbreviations list
3. Supplementary references list

1. Supplementary Figures

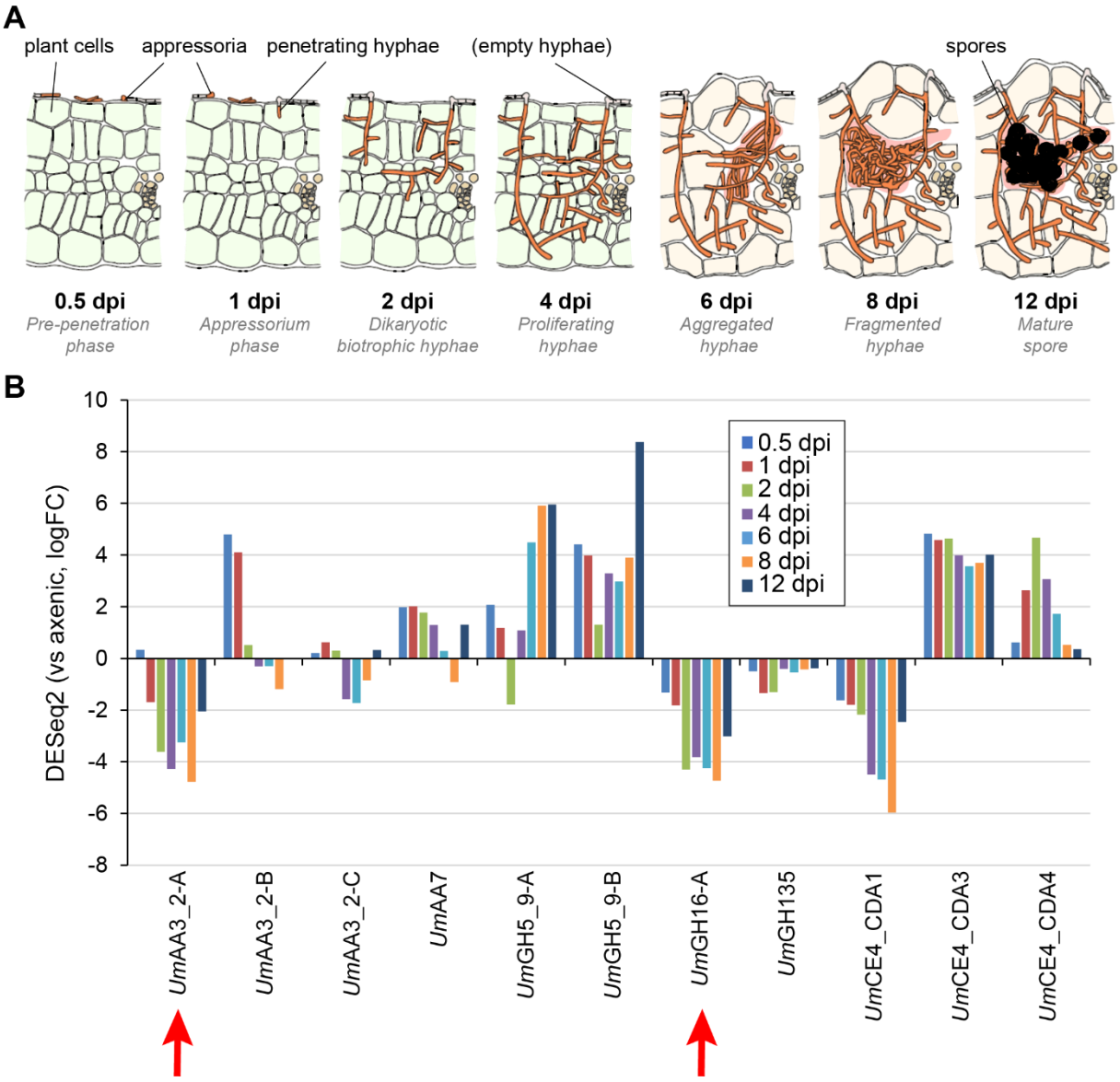


Fig. S1. Time-course expression along *U. maydis* infection cycle on maize of the 11 genes coding for putative FCW-active CAZymes. (A) Schematic representation of *U. maydis* plant infection cycle (dpi, days post inoculation). **(B)** Differential gene expression (*vs* axenic condition). Original transcriptomic data were retrieved from (1) for the following genes (proteins) : UMAG_03551 (*Um*AA3_2-A), UMAG_03256 (*Um*AA3_2-B), UMAG_04044 (*Um*AA3_2-C), UMAG_10861 (*Um*AA7), UMAG_05550 (*Um*GH5_9-A), UMAG_00235 (*Um*GH5_9-B), UMAG_02134 (*Um*GH16_1-A), UMAG_06157 (*Um*GH135), UMAG_00638 (*Um*CE4_CDA1), UMAG_11922(*Um*CE4_CDA3) and UMAG_01788 (*Um*CE4_CDA4). Red arrows indicate the two enzymes characterized in the present study.

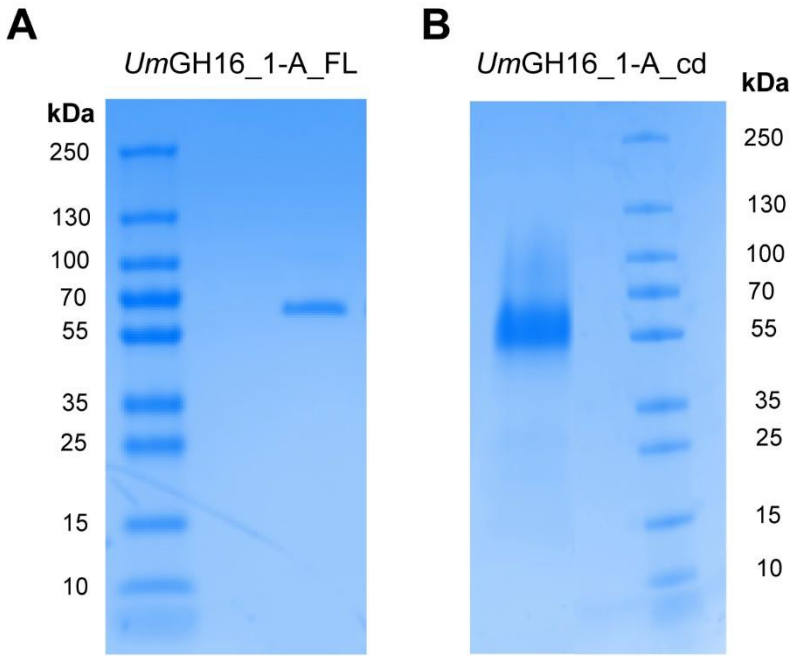


Fig. S2. SDS-PAGE analysis of *Um*GH16_1-A, with (A) and without (B) C-term extension. The enzymes were expressed in *P. pastoris* and purified by IMAC. The theoretical molecular weight of full length (*Um*GH16_1-A_FL) and truncated (*Um*GH16_1-A_cd) versions of *Um*GH16_1-A are 39.168 kDa and 33.439 kDa, respectively. Note that the higher molecular weight observed on SDS-PAGE is most likely due to protein *N*-glycosylation (7 putative glycosylation sites, exposed at the protein surface, are predicted by the online tool NetNGlyc – 1.0).

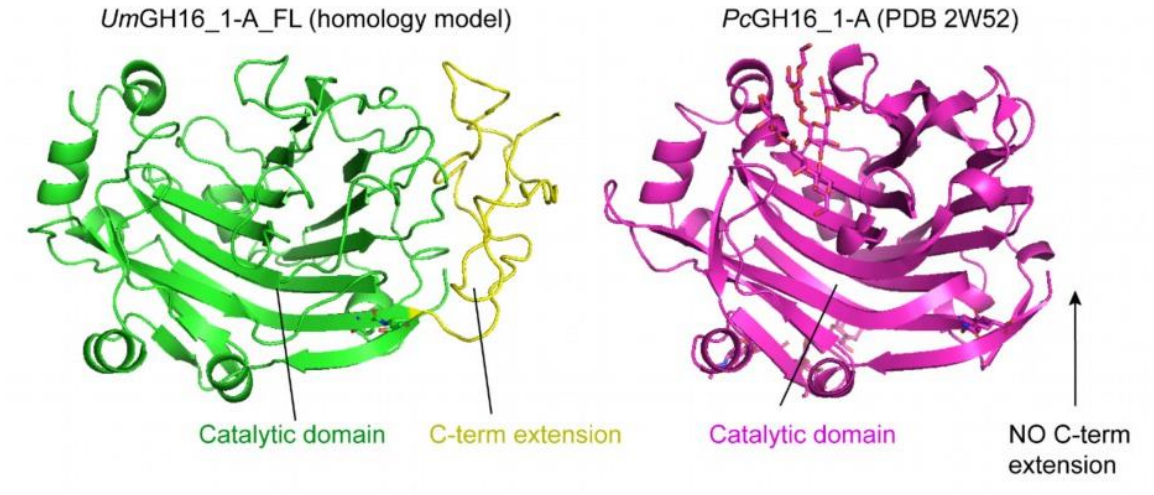
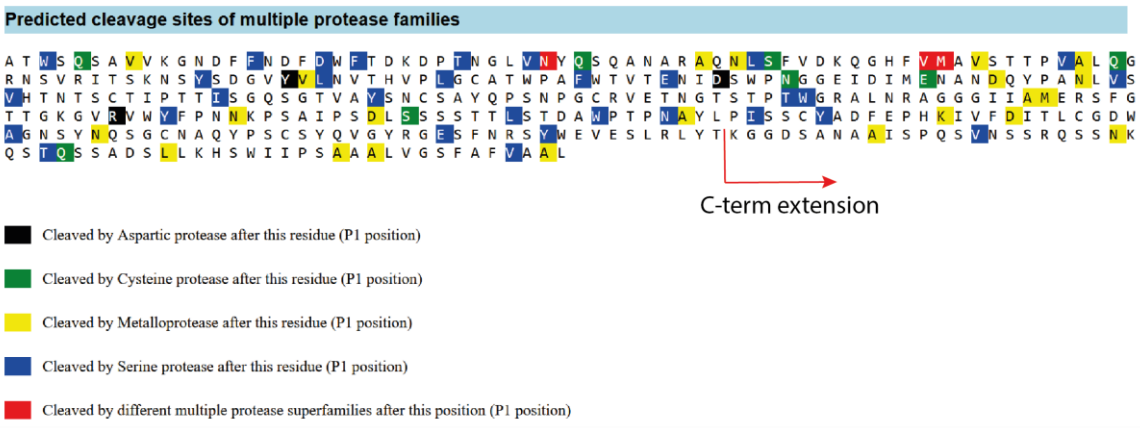


Fig. S3. Comparison of structures of (A) *Um*GH16_1-A_FL (homology model, with AlphaFold) and (B) its closest structural homologue, the GH16_1-A from *Phanerochaete chrysosporium* (PDB 2W52; (2)**).**

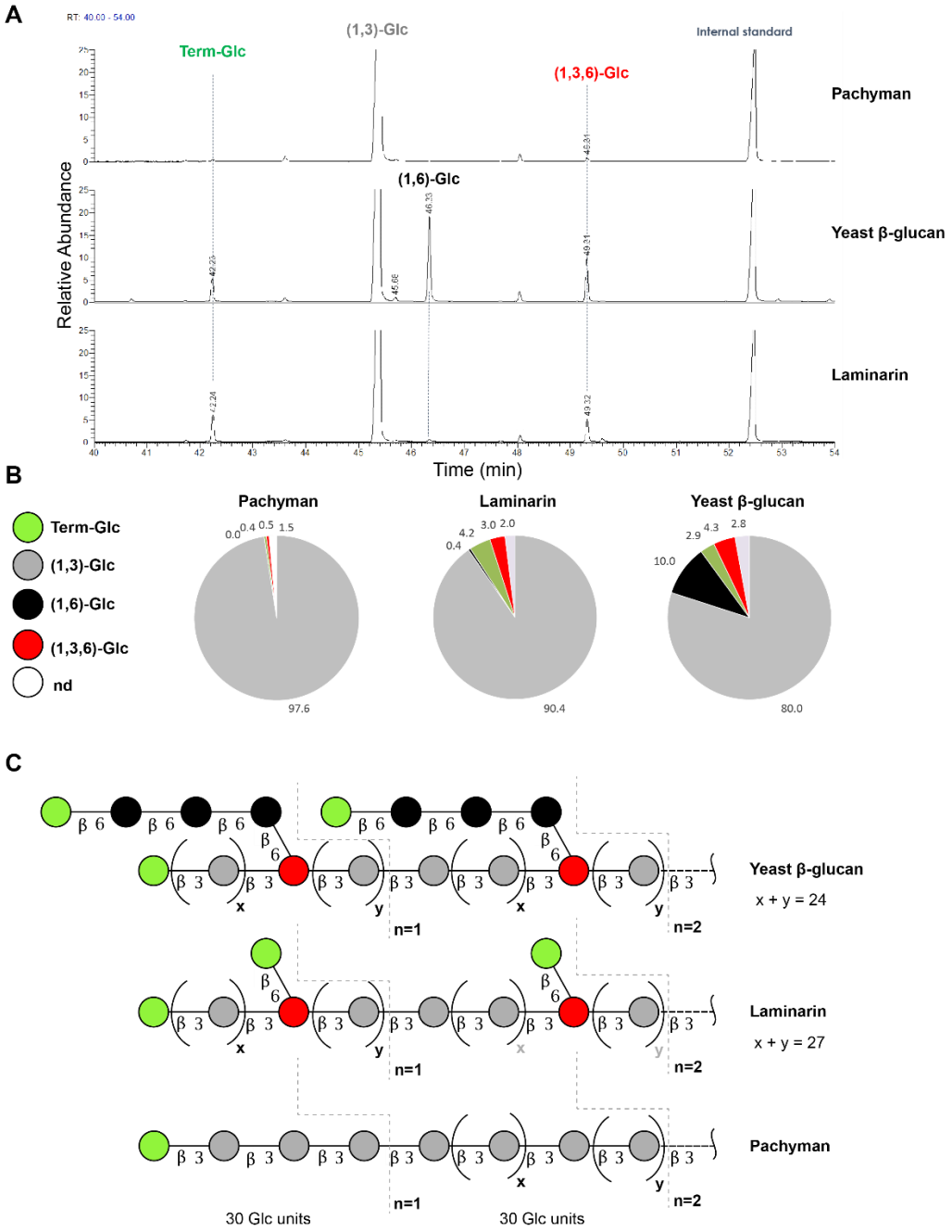
A



B

RANK in secretome	Score	OrthoDB	JGI ProtID	UMAG_ID	Enzyme class	Enzyme family	Protease Type
13	75	254398	jgi|Ustma2_2|9235|	UMAG_11908	PROTEASE	Cathepsin D.	Aspartic peptidase
20	39	254398	jgi|Ustma2_2|9376|	UMAG_02178	PROTEASE	Cathepsin D.	Aspartic peptidase
42	18	226881	jgi|Ustma2_2|13300|	UMAG_06118	PROTEASE	Tripeptidyl-peptidase I.	Serine peptidase
46	17	386362	jgi|Ustma2_2|8783|	UMAG_01888	PROTEASE	Carboxypeptidase D.	Serine peptidase
80	5	138643	jgi|Ustma2_2|8781|	UMAG_01886	PROTEASE	Carboxypeptidase C.	Serine peptidase
81	5	98560	jgi|Ustma2_2|11242|	UMAG_03947	PROTEASE	Carboxypeptidase D.	Serine peptidase
91	3	254398	jgi|Ustma2_2|6857|	UMAG_00064	PROTEASE	Cathepsin D.	Aspartic peptidase

Fig. S4. (A) Prediction of proteolytic cleavage sites in *Um*GH16_1-A and (B) proteases found in the TOP100 secreted proteins. Proteolytic cleavage sites were predicted with PROSPER (3).



	Laminarin			Yeast β-glucan		
Number of blocks	n = 1	n = 2	n = 3	n = 1	n = 2	n = 3
TOTAL Glucose units	30	60	90	30	60	90
Term-Glc	2 (6.6%)	3 (5%)	4 (4.4%)	2 (6.6%)	3 (5%)	4 (4.4%)
(1,3)-Glc	27 (90.0%)	55 (91.7%)	89 (92.2%)	24 (80.0%)	52 (86.7%)	89 (88.9%)
(1,6)-Glc	0 (0%)	0 (0%)	0 (0%)	3 (10%)	3 (10%)	3 (10%)
(1,3,6)-Glc	1 (3.3%)	2 (3.3%)	3 (3.3%)	1 (3.3%)	2 (3.3%)	3 (3.3%)

Fig. S5. Linkage analyses of Laminarin, yeast β-glucan and pachyman. (A) Total Ion Chromatograms (TIC) of partially methylated alditol acetates by GC-MS, **(B)** quantification and **(C)** reconstitution of sugar linkages.

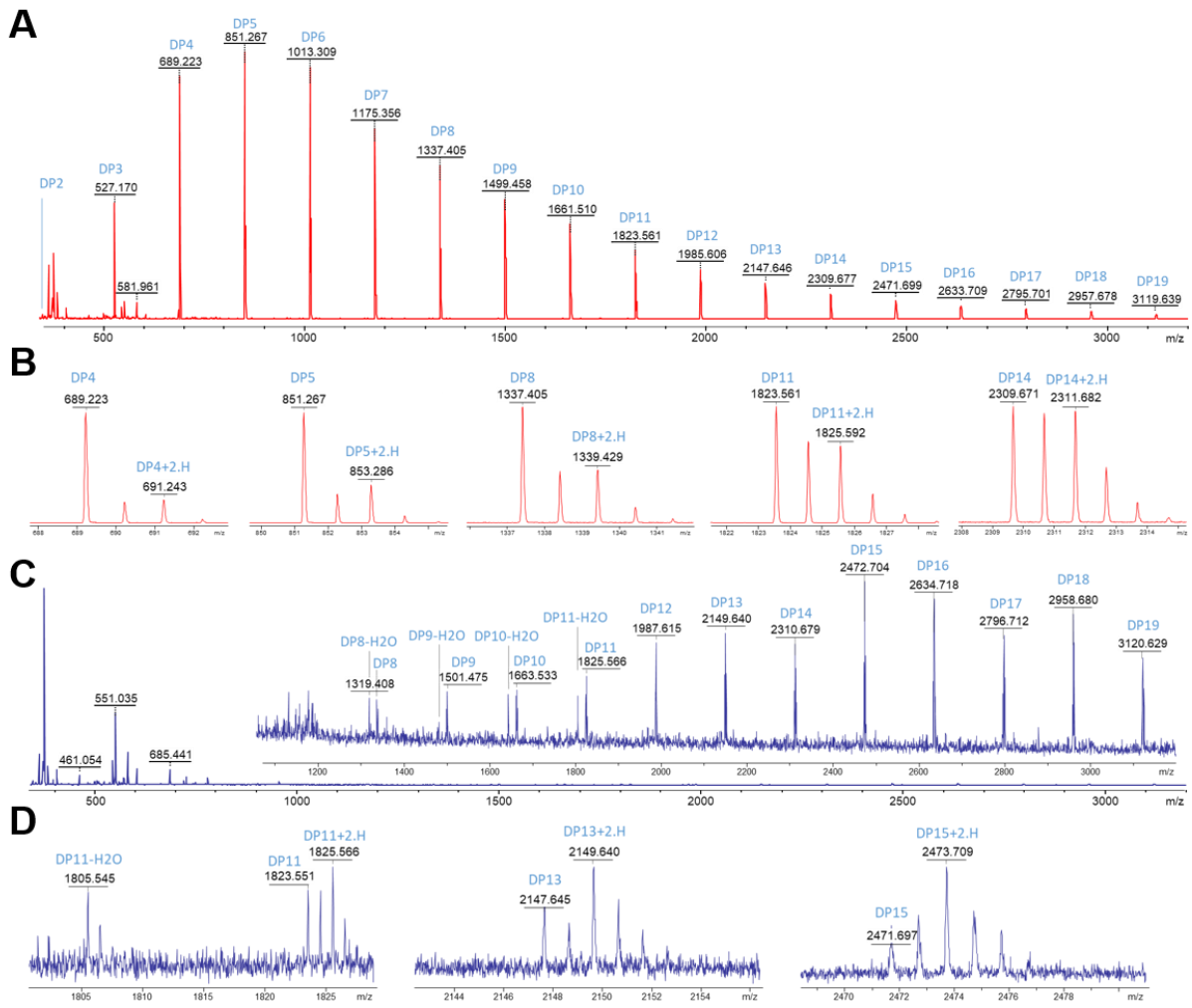


Fig. S6. MALDI-ToF-MS analyses of soluble products released from Laminarin in the presence (A-B, red) or absence (C-D, blue) of *Um*GH16_1-A_cd. Full Spectra are presented in figure A and C. Representative zooms to highlight normal species, reduced species and species with a loss of water are shown figure B and D.

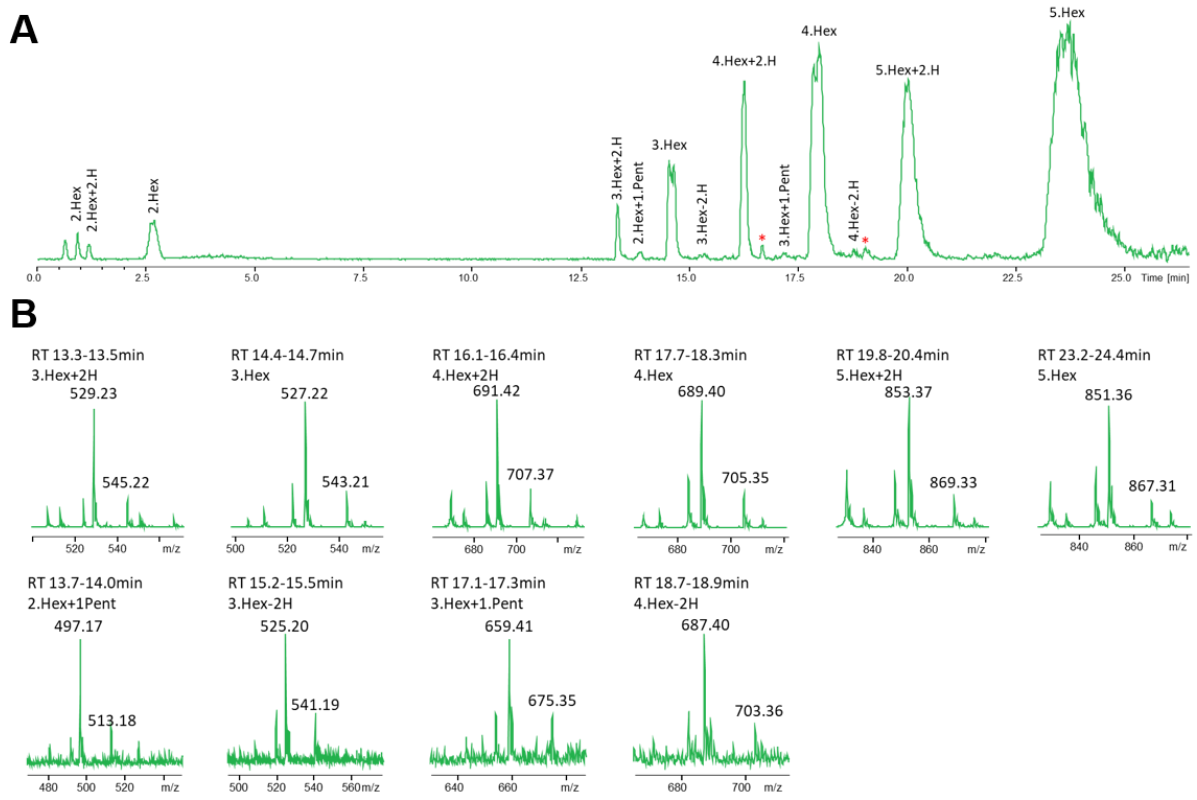


Fig. S7. UPLC-ESI-IT analyses of soluble products released from Laminarin by *Um*GH16_1-A_cd. (A) Base peak chromatography between 0 and 27 min. (red stars indicate a contaminant) and **(B)** ESI-MS spectra of each compounds revealed by the analysis. RT and nature of the oligosaccharides are indicated for each MS spectrum. For the sake of clarity, only the $[M+Na]^+$ and the $[M+K]^+$ ions are labelled with the corresponding m/z values.

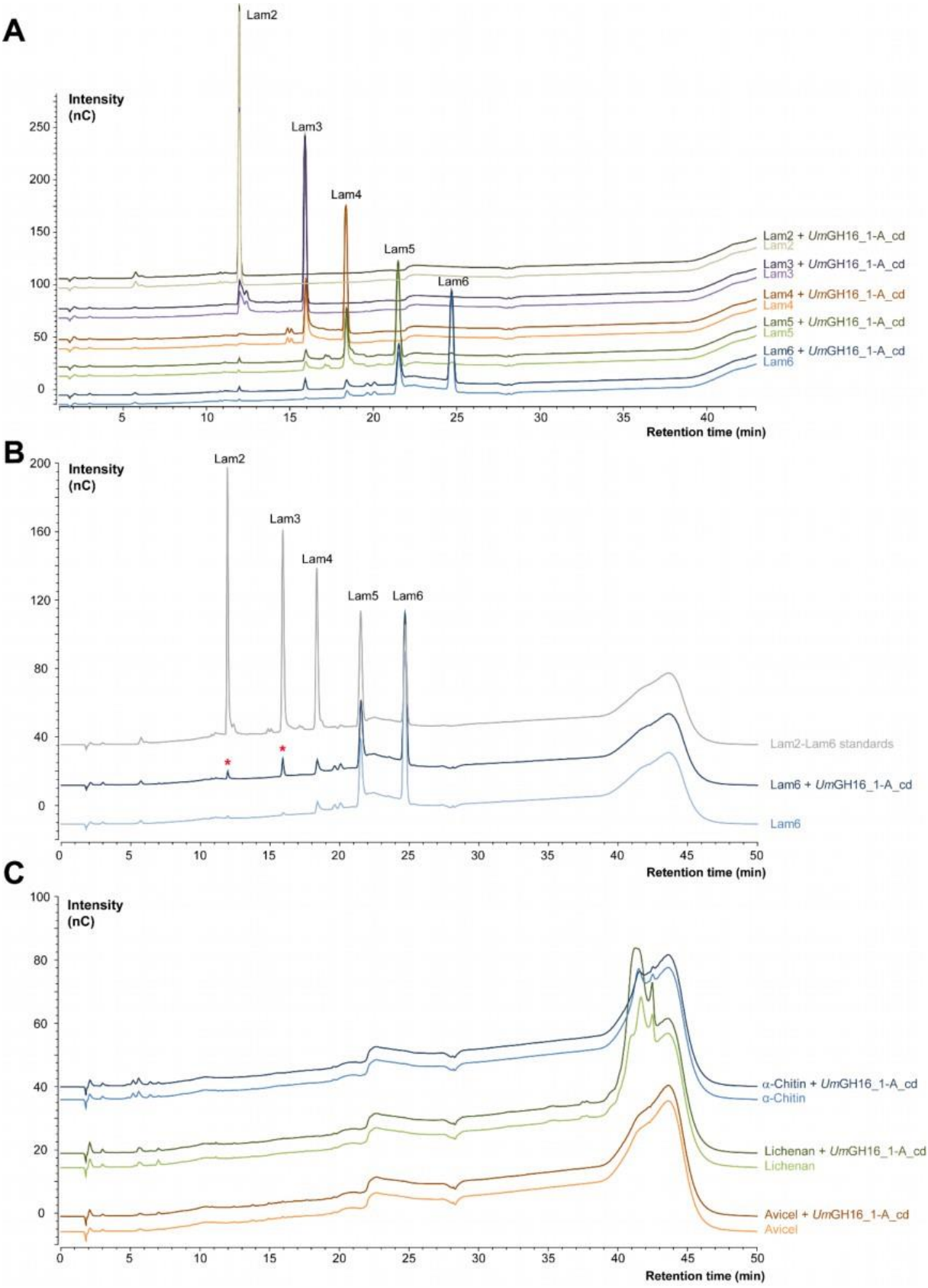


Fig. S8. Control reactions of *Um*GH16_1-A on β-1,4 glucans, mixed β-1,3/1,4 glucan and β-1,3 glucooligosaccharides. The graphs show HPAEC-PAD chromatograms of reaction

products released from **(A-B)** laminari-oligosaccharides (DP2-DP6; 1 mM each) and **(C)** Avicel, α-chitin or Lichenan (10 mg.mL^{-1} final concentration) by *Um*GH16_1-A_cd (10 nM). Panel B shows a zoom-in view of chromatograms displayed in panel A for reactions on Lam6 only (Lam2 to Lam 5 were not recognized as substrates by *Um*GH16_1-A_cd). On Lam6, very small amounts of products Lam2 and Lam 3 (at 12 and 16 min, red stars) were detected. In panel C, Avicel and α-chitin are linear polymers of β-1,4-linked D-glucose and *N*-acetyl-glucosamine units, respectively. Lichenan is a mixed β-1,3/1,4 glucan. All reactions were incubated during 4 h, in citrate phosphate buffer (50 mM, pH 5.5), in a thermomixer (30 °C, 1,000 rpm). All experiments were carried put in triplicate. However, for the sake of clarity, only one replicate is shown.

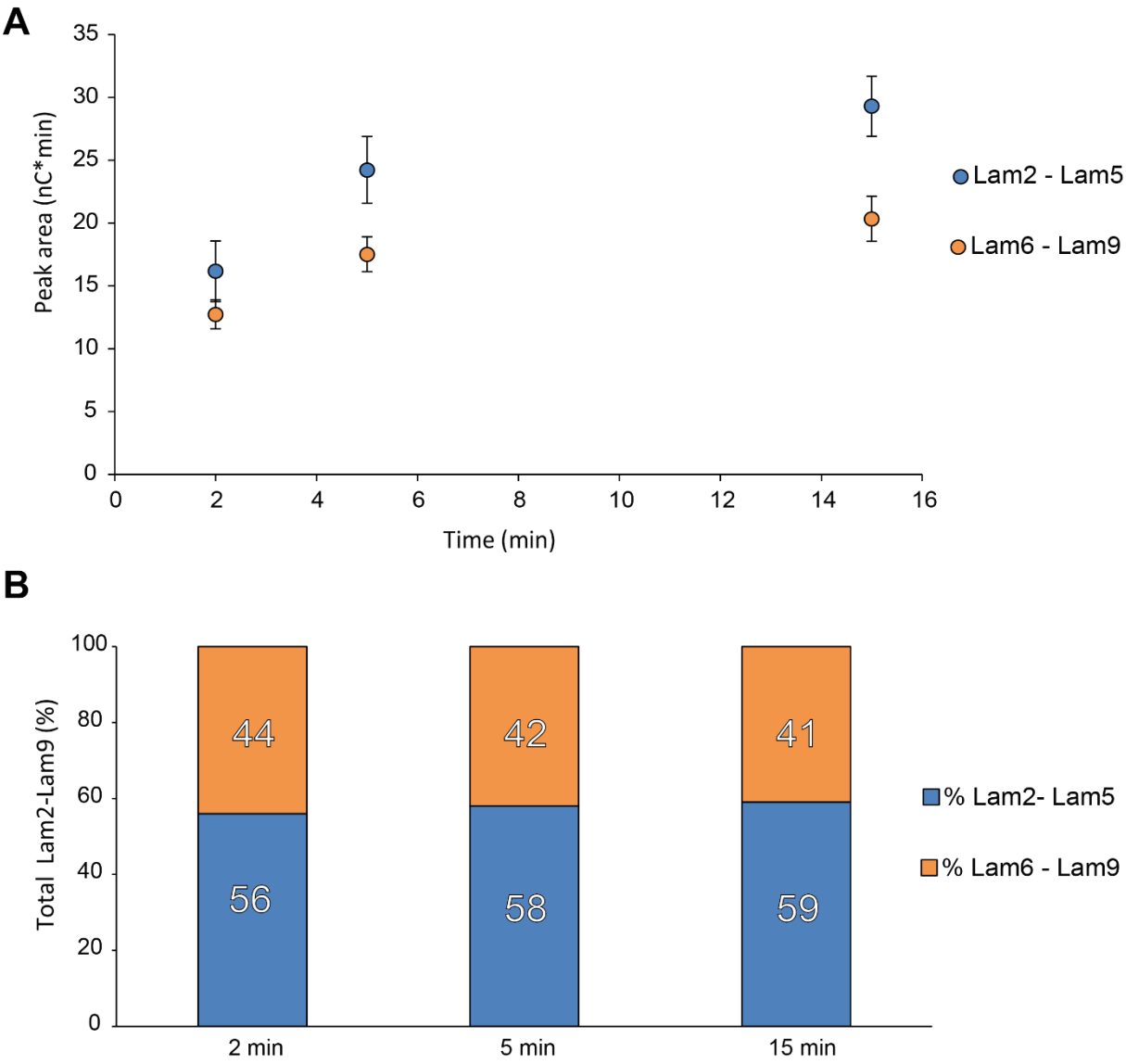


Fig. S9. Time-course release of short (Lam2-Lam5) and long (Lam6-Lam9) oligosaccharides from Laminarin by *Um*GH16_1-A_cd. The amount of oligosaccharides is expressed as **(A)** the absolute sum of peak areas and **(B)** the proportion of short and long oligosaccharides at each time point (the sum of short and long oligosaccharides is equal to 100% of released oligosaccharides at each time point). The reactions contained laminarin (10 $mg.mL^{-1}$), *Um*GH16_1-A_cd (10 nM) in citrate phosphate buffer (50 mM, pH 5.5) and incubated in a thermomixer (30 °C, 1,000 rpm). Data are presented as average values (n = 3, independent biological replicates) and error bars show s.d.

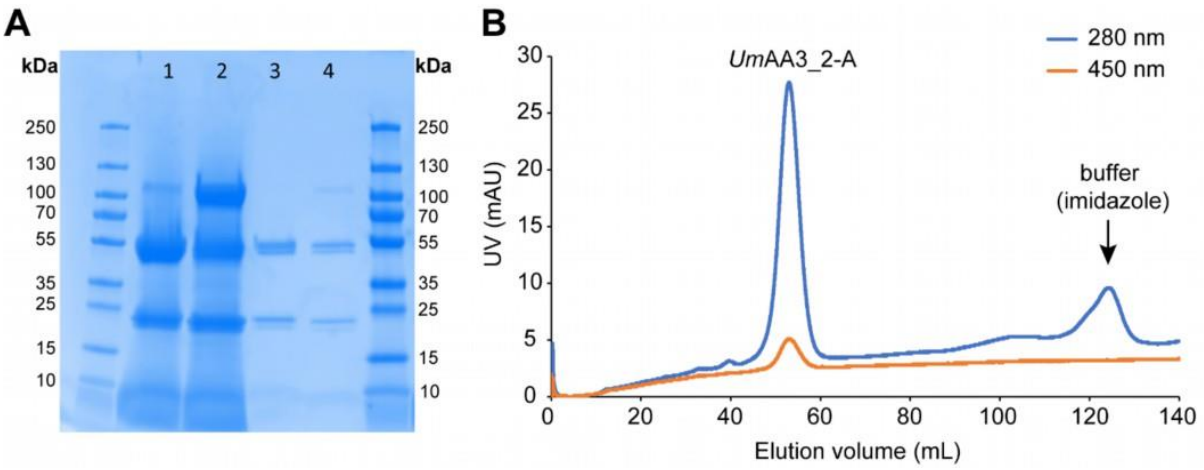


Fig. S10. Purity and molecular weight analyses of *Um*AA3_2-A. (A) SDS-PAGE and **(B)** SEC analyses**.** The enzyme was expressed in *P. pastoris* and purified by IMAC followed by SEC. In panel A, the following samples were loaded on the SDS-PAGE gel: lane 1, IMAC-purified *Um*AA3_2-A; lane 2, heat-treated, IMAC-purified *Um*AA3_2-A; lane 3, SEC-purified *Um*AA3_2-A; lane 4, heat-treated, SEC-purified *Um*AA3_2-A. In panel B, using a calibrated SEC column, we determined an experimental MW of *Um*AA3_2-A of 48.2 kDa (average of n = 2 independent experiments) (theoretical MW = 64.9 kDa).

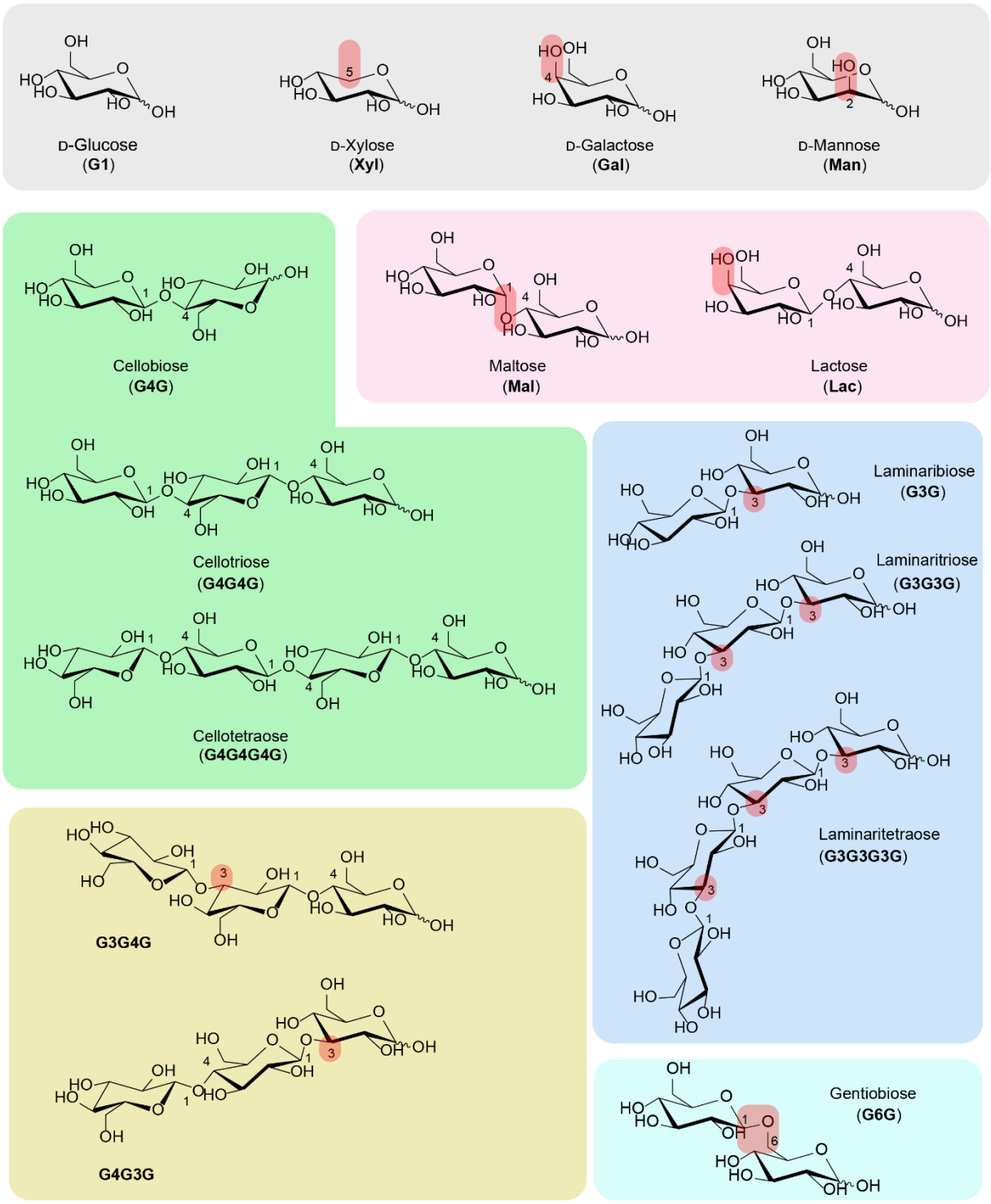


Fig. S11. Chemical structures of oligosaccharides tested in *Um*AA3_2-A substrate specificity screening.

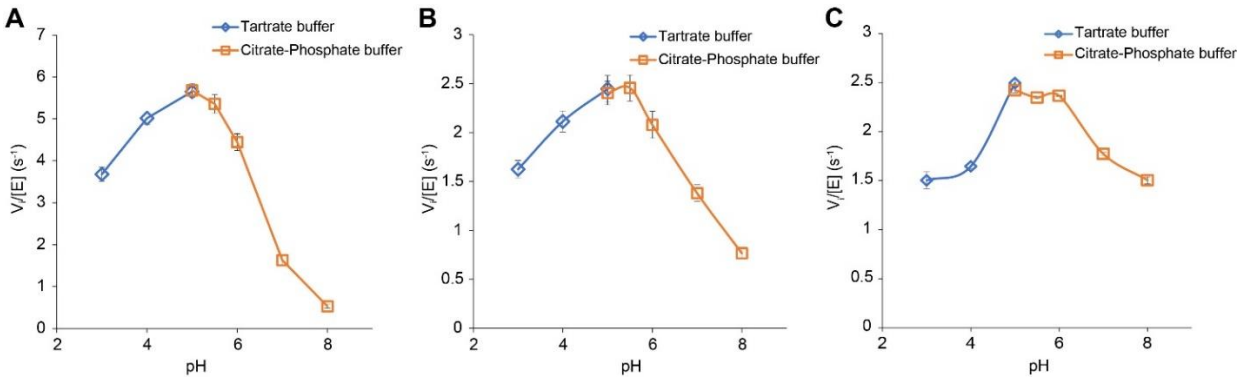


Fig. S12. pH activity profile of *Um*AA3_2-A on (A) glucose and (B) G3G and (C) G6G. The graph shows the reduction rate of DCIP (400 µM) by *Um*AA3_2-A (110 nM), in the presence of glucose (500 mM), G3G (5 mM) or G6G (5 mM) at different pH values (50 mM of tartrate or citrate-phosphate buffer). All reactions were carried out at 30 °C. Data points show average values and error bars show s.d. (n = 3 independent biological replicates).

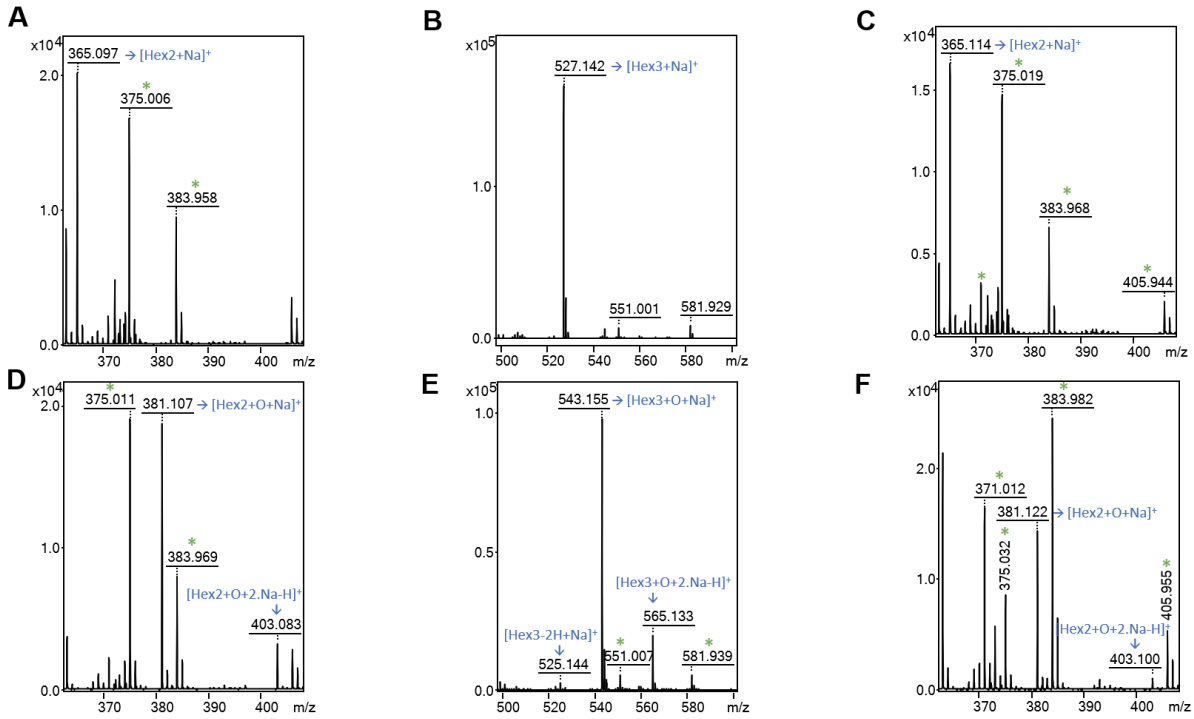


Fig. S13. MALDI-ToF MS analysis of G3G, G3G3G and G6G oxidation by *Um*AA3_2-A. Spectra of G3G **(A-D),** G3G3G **(B-E)** and G6G **(C-F)** before **(A-C)** and after **(D-F)** treatment by *Um*AA3_2-A. The spectra show the detection of native oligosaccharides [M+Na] with *m/z* = 365 (for G3G and G6G) or 527 g.mol^{-1} (for G3G3G). Upon addition of *Um*AA3_2-A oxidized species emerge: simple sodium adducts of oxidized form [M+16+Na] (*m/z* = 381 for G3G and G6G, or 543 for G3G3G) and double sodium adducts of oxidized form [M+16-H+2Na] (*m/z* =403 for G3G and G6G, or 581 for G3G3G), which suggest the formation of aldonic acids (see **Fig. S14** for validation).

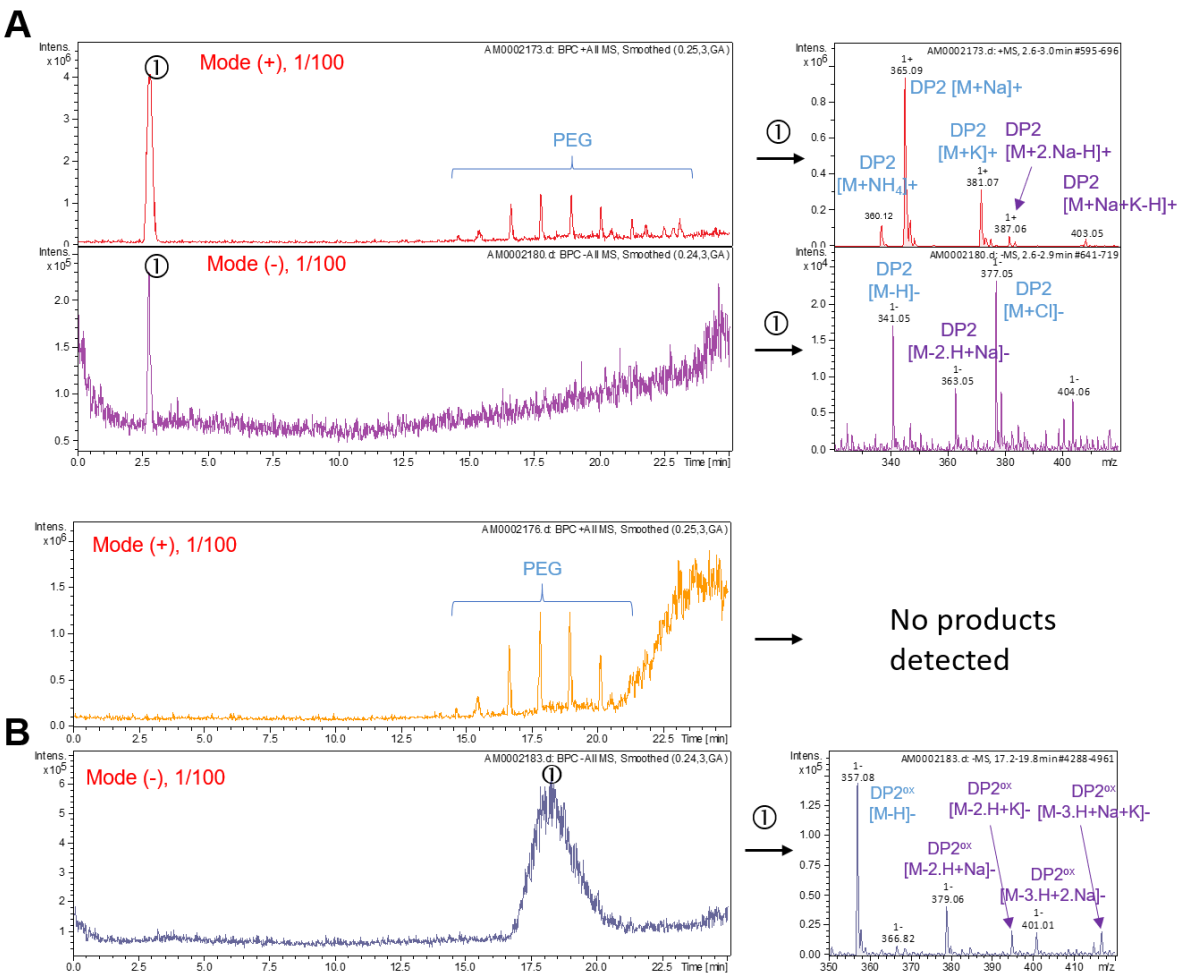


Fig. S14. UPLC-MS analysis of G3G before (A) and after (B) oxidation by *Um*AA3_2-A. Each panel shows the UPLC chromatogram (left) and MS spectra (right) using either positive (upper graph) or negative (lower graph) ionization mode.

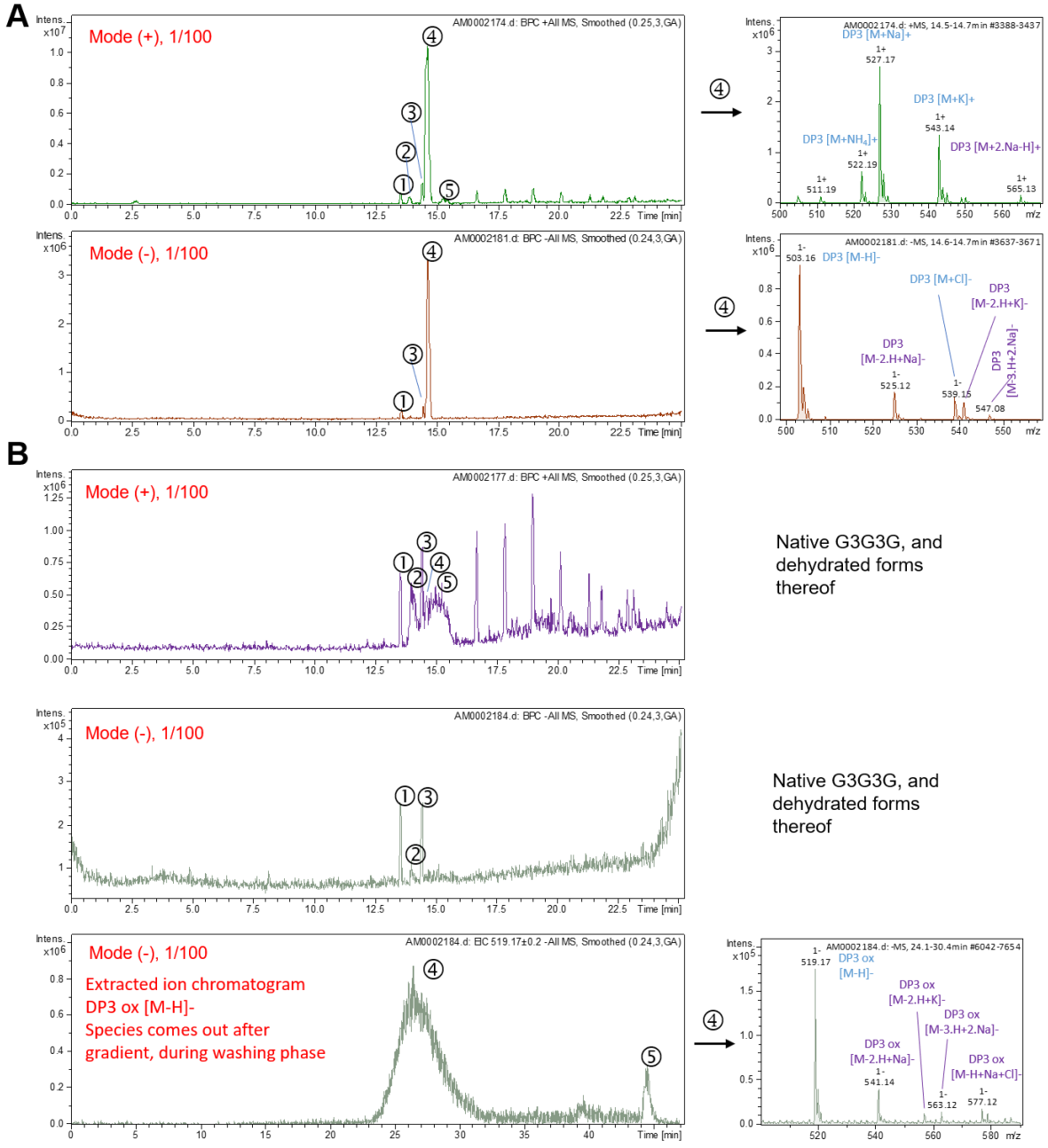


Fig. S15. UPLC-MS analysis of G3G3G before (A) and after (B) oxidation by *Um*AA3_2-A. Each panel shows the UPLC chromatogram (left) and MS spectra (right) using either positive (upper graph) or negative (lower graph) ionization mode.

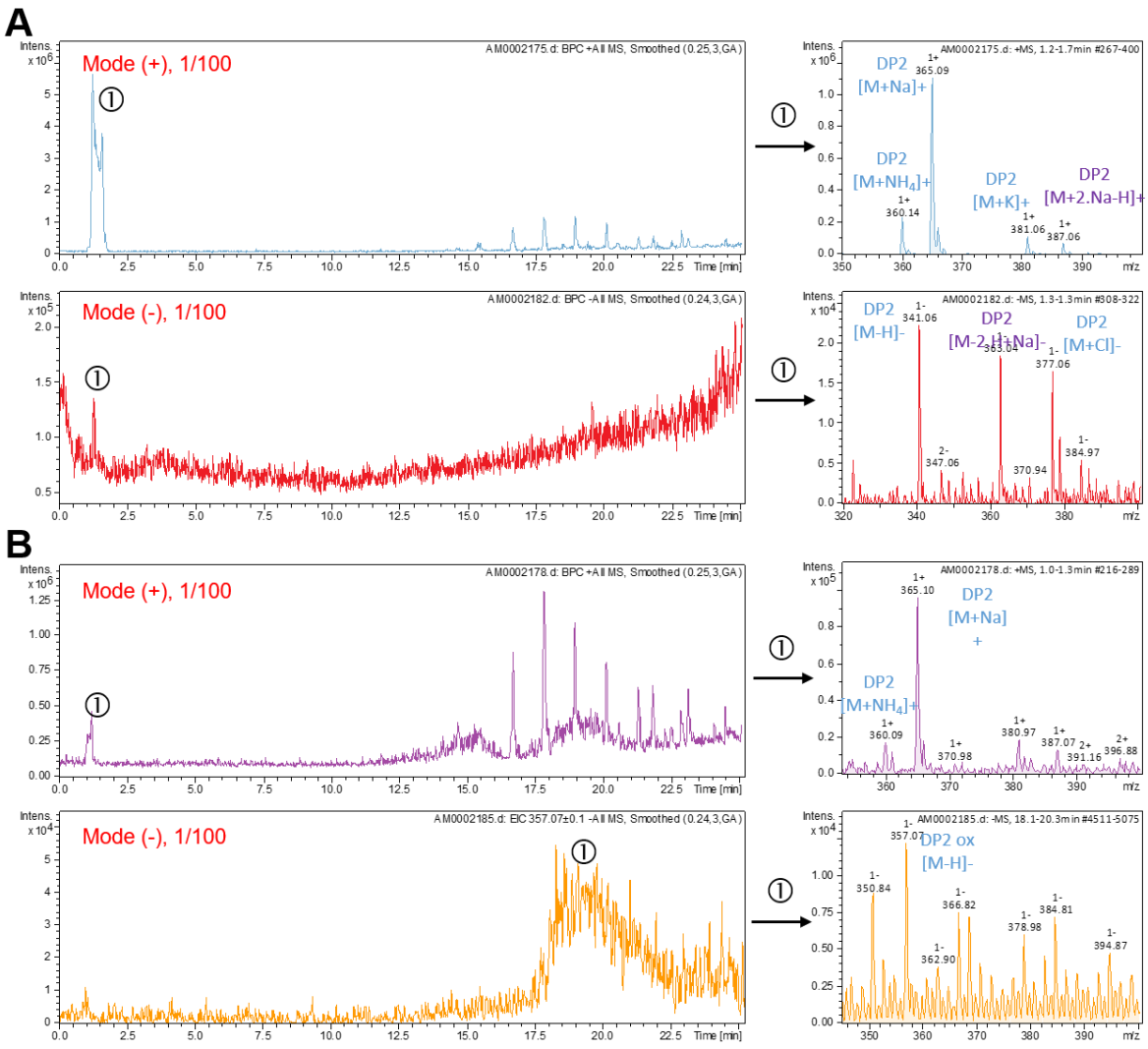


Fig. S16. UPLC-MS analysis of G6G before (A) and after (B) oxidation by *Um*AA3_2-A.

Each panel shows the UPLC chromatogram (left) and MS spectra (right) using either positive (upper graph) or negative (lower graph) ionization mode.

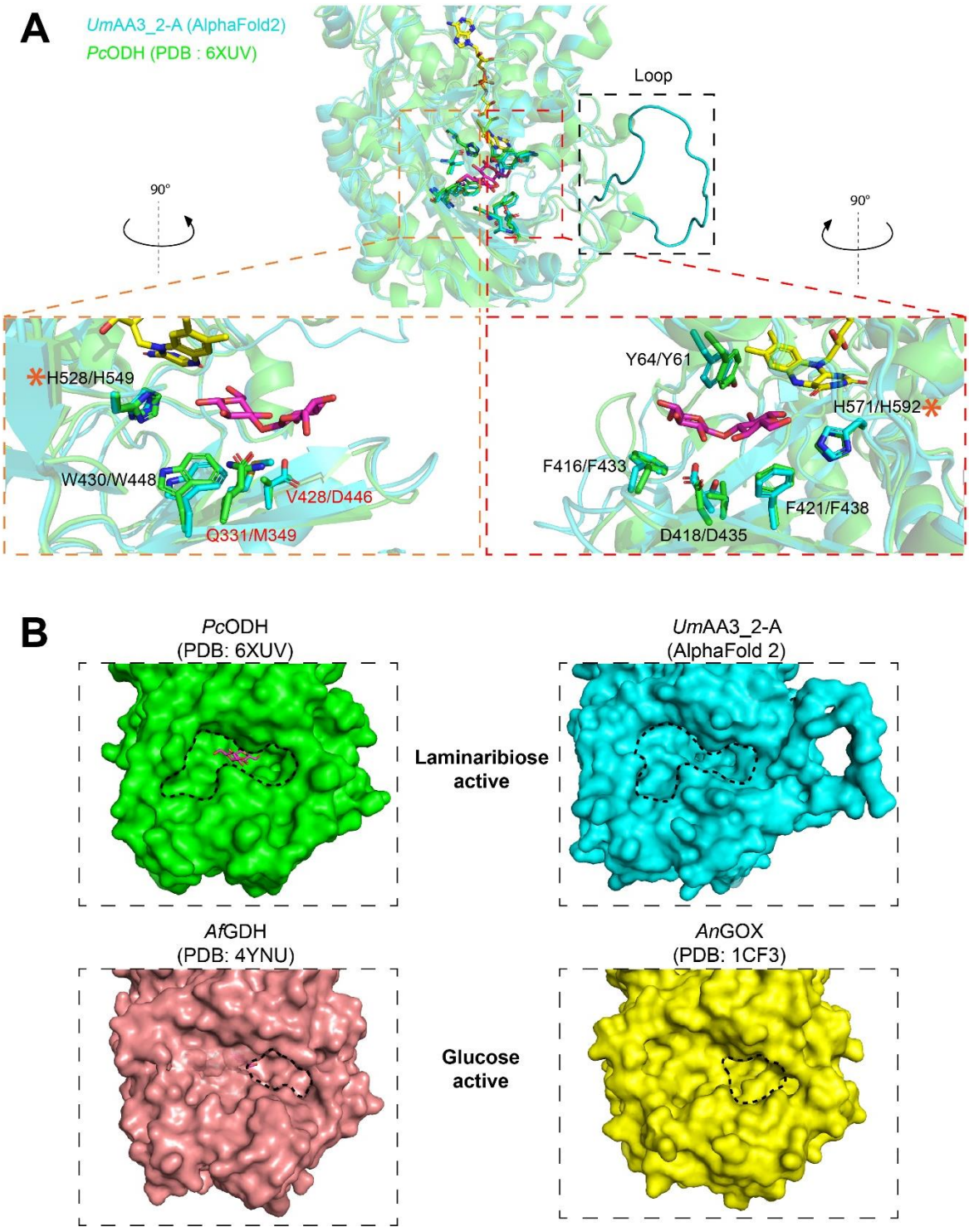


Fig. S17. Structural comparison of *Pc*ODH (PDB id 6XUV), and *Um*AA3_2-A (model). **(A)** The structure of *Pc*ODH (in green) and *Um*AA3_2-A (in blue) were superimposed in Pymol and shown as cartoon. A global view facing the entrance active site is displayed on top of the figure. The additional loop of *Um*AA3_2-A is squared in black and the active site of both enzymes are subdivided in two parts: “left side” (squared in orange) and “right side” (squared in red). Zooms on these two parts are shown on the lower part of the figure, where the keys

amino acids for laminaribiose binding in *Pc*ODH and their structural equivalents in *Um*AA3_2-A are displayed as stick and annotated according to amino acid numbering in *Pc*ODH/*Um*AA3_2-A. These amino acids are annotated in black when conserved in both proteins or in red when substituted. The two catalytic histidines are labelled with an orange star. The Laminaribiose (colored in purple) and the Flavin adenine dinucleotide (FAD) co-factor (colored in yellow) are shown in sticks. **(B)** Comparison of the active site of laminaribiose active enzymes (*Pc*ODH and *Um*AA3_2-A) with glucose active enzymes (*Af*GDH and *An*GOX). The proteins structures are shown as surface and approximative perimeter of the active site entrance is delimited by a black dotted line.

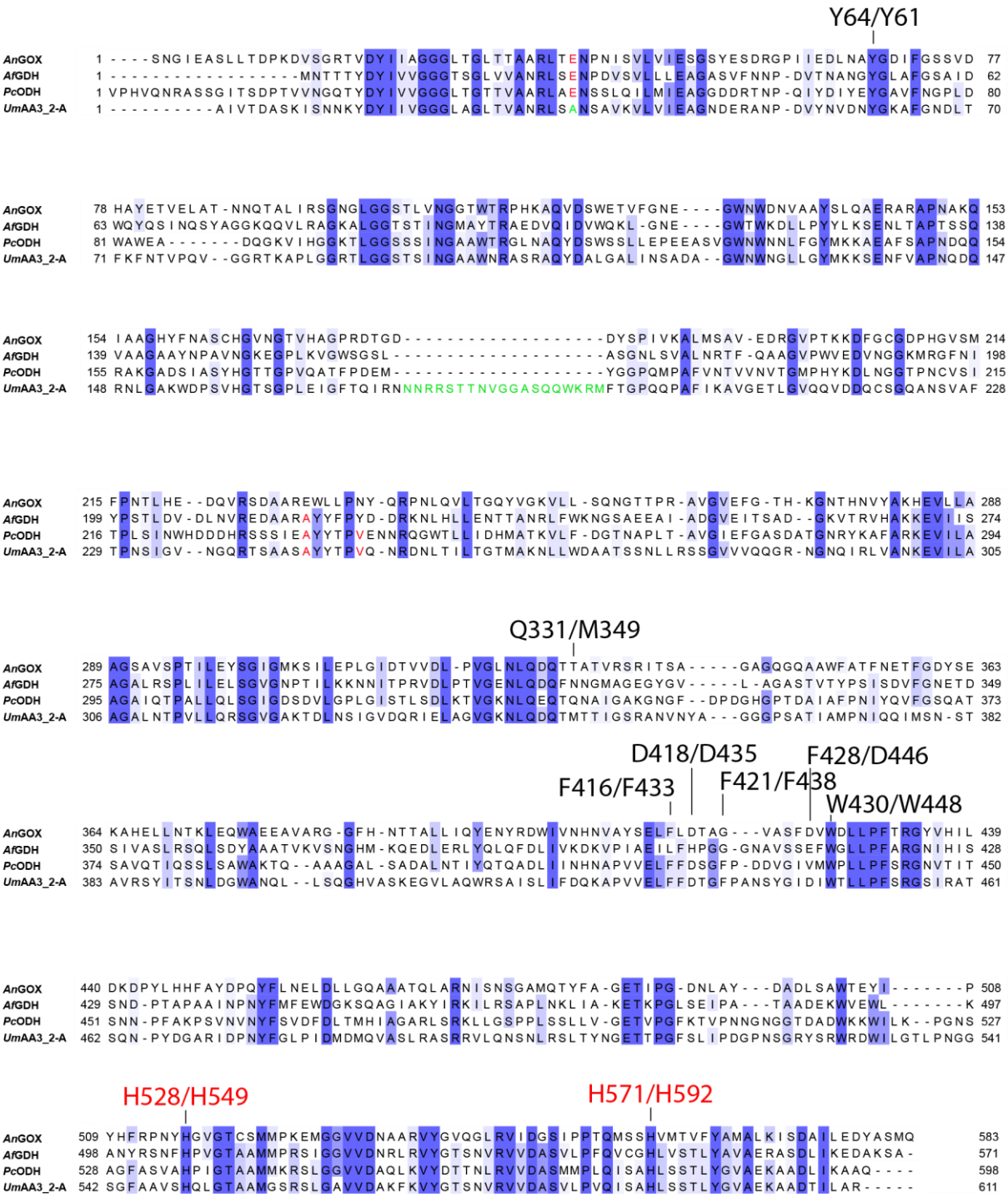


Fig. S18. Multiple sequence alignment (MSA) of *An*GOX, *Af*GDH, *Pc*ODH and *Um*GH16_1-A. The MSA shows the presence of an extra loop in *Um*AA3_2-A (printed in green) and key residues involved in substrate binding (annotated according to amino acid numbering in *Pc*ODH/*Um*AA3_2-A; catalytic histidines are printed in red).

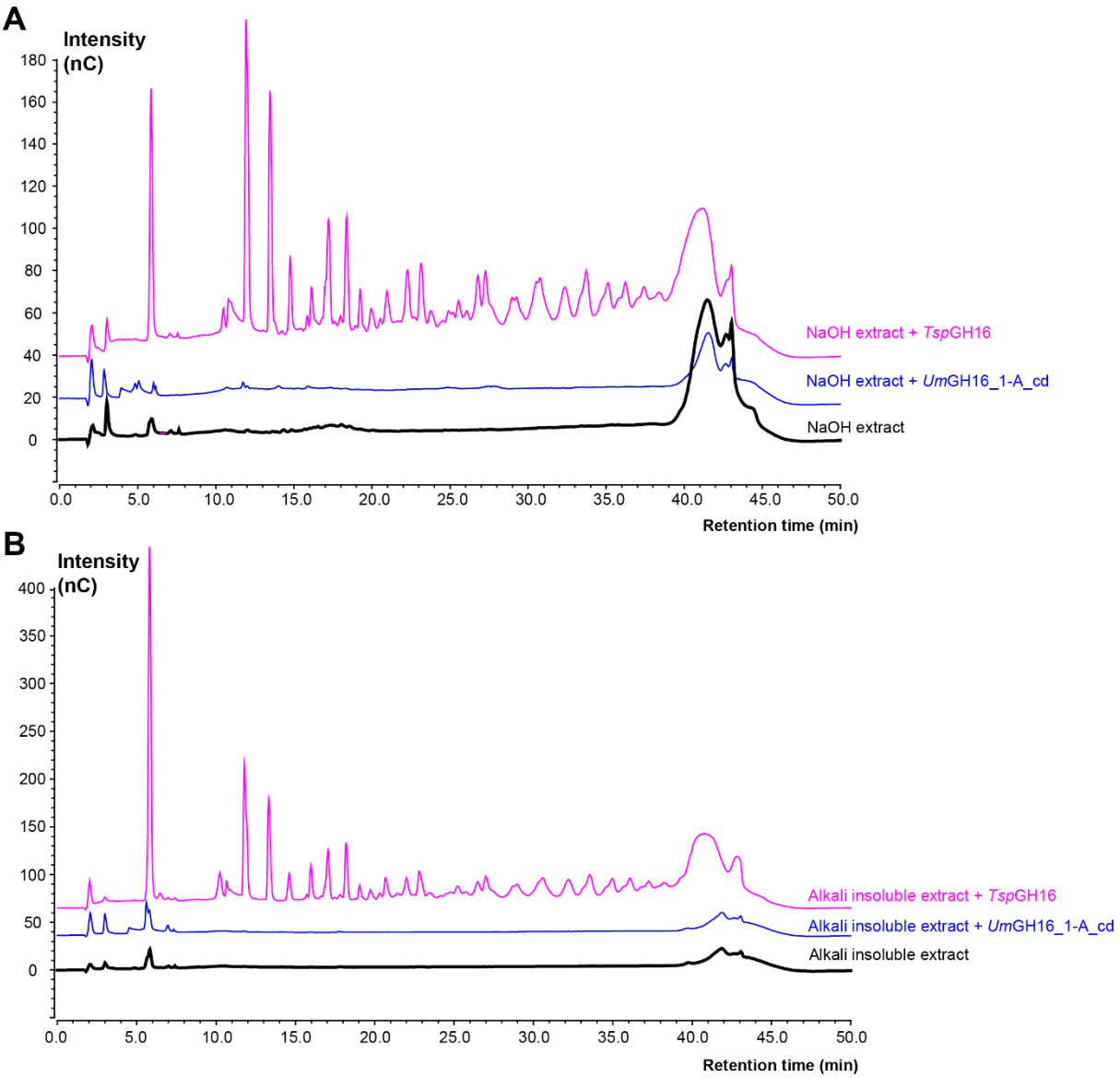


Fig. S19. Activity of *Um*GH16_1-A_cd and *Tsp*GH16_3 (from Megazyme) on polysaccharides extracts from *U. maydis*. The graphs show HPAEC-PAD chromatograms of reaction products released from **(A)** NaOH-extracted polysaccharides (5 mg.mL^{-1} final concentration), and **(B)** from alkali insoluble material (approx. 1-5 mg.mL^{-1}), by the commercial *Tsp*GH16_3 (100 nM) and *Um*GH16_1-A_cd (100 nM). All reactions were incubated during 16 h, in citrate phosphate buffer (50 mM, pH 5.5), in a thermomixer (30 °C, 1,000 rpm). For the sake of clarity, only one chromatogram for each reaction condition is shown (each experiment was carried out at least in triplicate).

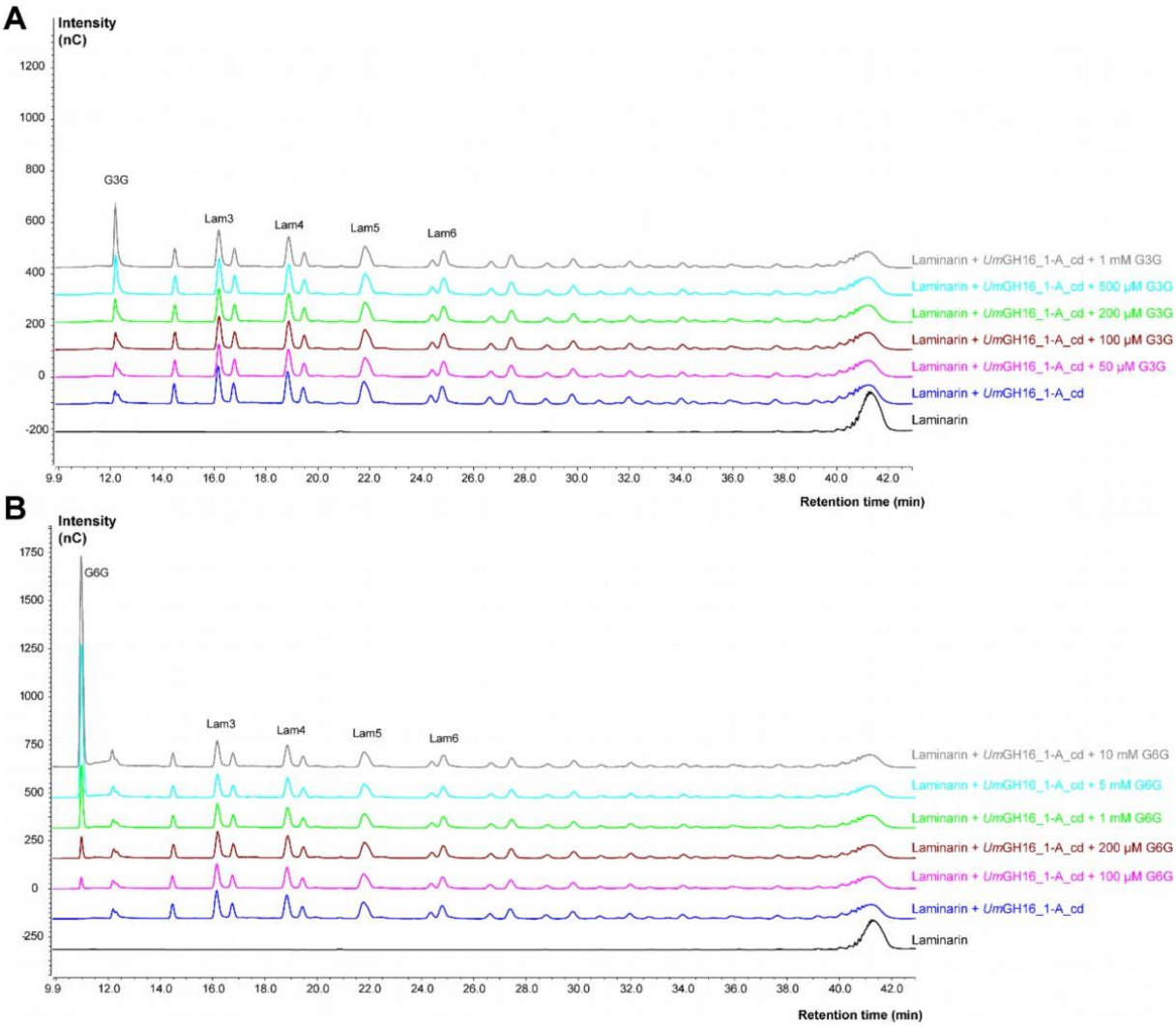


Fig. S20. *Um*GH16_1-A is not inhibited by G3G or G6G. The graphs show HPAEC-PAD chromatograms of reaction products released from laminarin (10 mg.mL^{-1}) by *Um*GH16_1-A_cd (10 nM) in the presence of various concentrations of **(A)** G3G (0-1 mM) or **(B)** G6G (0-10 mM). All reactions were incubated during 4 h, in citrate phosphate buffer (50 mM, pH 5.5), in a thermomixer (30 °C, 1,000 rpm), (n = 1). In the negative control reaction, Laminarin in the absence of enzyme was incubated in the same conditions as other reactions. Abbreviations: G3G, laminaribiose (also called Lam2); G6G, gentiobiose; Lam3 to Lam6, β-1,3-glucooligosaccharides of DP3 to 6.

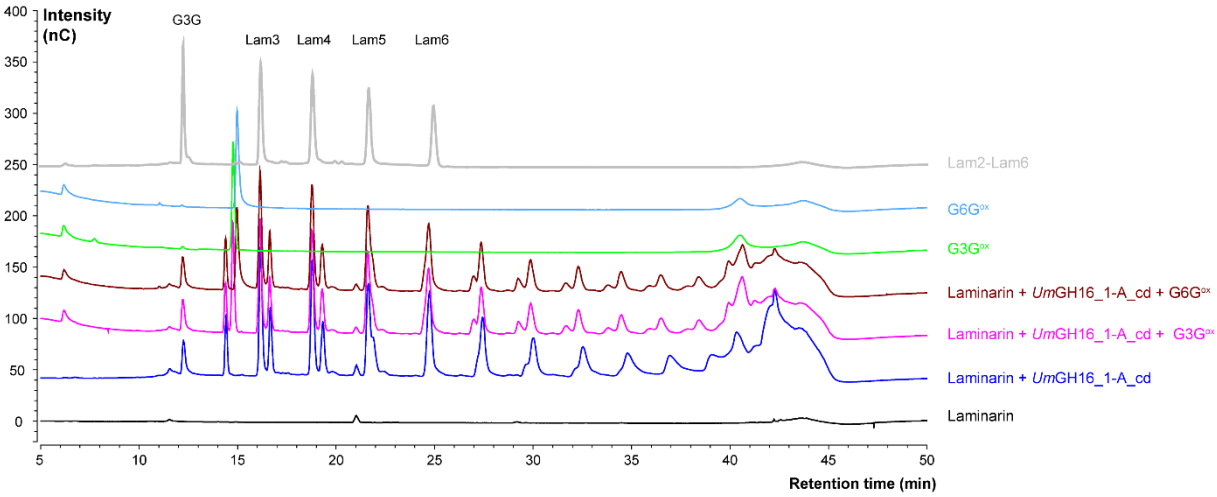


Fig. S21. *Um*GH16_1-A is not inhibited by oxidized G3G or G6G. The graphs show HPAEC-PAD chromatograms of reaction products released from laminarin (10 mg.mL^{-1}) by *Um*GH16_1-A_cd (10 nM) in the presence of oxidized G3G (G3G^{ox}; 1 mM) or oxidized G6G (G6G^{ox}, 1 mM). All reactions were incubated during 4 h, in citrate phosphate buffer (50 mM, pH 5.5), in a thermomixer (30 °C, 1,000 rpm), (n = 1). See the experimental section for the preparation of G3G^{ox} and G6G^{ox}. Abbreviations: G3G, laminaribiose (also called Lam2); G6G, gentiobiose; Lam3 to Lam6, β-1,3-glucooligosaccharides of DP3 to 6.

Table. S1. The 21 CAZymes present in the TOP50 proteins secreted by *U. maydis* after re-analysis in 2022[a].

RANK in TOP 50[b]	JGI ProtId	Annotation in 2012	Re-Annotation in 2022	Score	Predicted Target (phylogeny-based)[c]	Biochemically characterized?
1	jgi|Ustma2_2|11640|	GH27-CBM35	GH27-CBM35	586	PCW	NO
2	jgi|Ustma2_2|11831|	GH62	GH62	556	PCW	NO
3	jgi|Ustma2_2|10689|	GH10	GH10	429	PCW	NO
5	jgi|Ustma2_2|7673|	GH51	GH51	281	PCW	NO
6	jgi|Ustma2_2|10518|	CDH_2	AA3_2	233	ND	NO
7	jgi|Ustma2_2|11351|	CDH_2	AA3_2	188	ND	NO
8	jgi|Ustma2_2|13127|	FAD-Oxidase	AA7	180	FCW	NO
9	jgi|Ustma2_2|9924|	CE4	CE4	119	FCW	Rizzi et al.
11	jgi|Ustma2_2|13337|	UNK	GH135	90	FCW	NO
12	jgi|Ustma2_2|12578|	GH5	GH5_16	88	PCW	NO
16	jgi|Ustma2_2|7458|	CE4	CE4	58	FCW	Rizzi et al.
17	jgi|Ustma2_2|13257|	GH3	GH3	56	PCW	NO
24	jgi|Ustma2_2|9413|	GH37	GH37	32	PCW	NO
27	jgi|Ustma2_2|9206|	GH26	GH26	27	PCW	NO
29	jgi|Ustma2_2|8673|	CE4	CE4	26	FCW	Rizzi et al.
35	jgi|Ustma2_2|9331|	GH16	GH16	22	FCW ?	This Study
37	jgi|Ustma2_2|8391|	EXPN	EXPN	20	PCW	NO
40	jgi|Ustma2_2|9202|	GH45	GH45	19	PCW	NO
41	jgi|Ustma2_2|12699|	GH5	GH5_9	19	FCW	NO
47	jgi|Ustma2_2|7038|	GH5	GH5_9	17	FCW	NO
49	jgi|Ustma2_2|10841|	CDH_2	AA3_2	16	FCW ?	This Study

[a] Initial secretomic data (*U. maydis* secretome harvested after 7 days of growth on maize bran) were published by (4).
[b] Only CAZymes are shown in this Table.
[c] Abbreviations: FCW, Fungal Cell Wall ; PCW, Plant Cell Wall ; ND, Not determined; UNK, Unknown.

Table S2. Kinetic parameters of *Um*AA3_2-A and *Pc*ODH[a,b]

	Ustilago maydis								***Pycnoporus cinnabarinus***			
	UmAA3_2-A (This Study)				***Um*GDHIII (Wijayanti et al. (2021))[c]**				***Pc*ODH (Cerruti et al. (2021))**			
	M-M model			initial slope	M-M model			initial slope	M-M model			initial slope
	k_{cat}	K_M	k_{cat}/K_M	k_{cat}/K_M	k_{cat}	K_M	k_{cat}/K_M	k_{cat}/K_M	k_{cat}	K_M	k_{cat}/K_M	k_{cat}/K_M
	(s^{-1})	(mM)	($s^{-1}.M^{-1}$)	($s^{-1}.M^{-1}$)	(s^{-1})	(mM)	($s^{-1}.M^{-1}$)	($s^{-1}.M^{-1}$)	(s^{-1})	(mM)	($s^{-1}.M^{-1}$)	($s^{-1}.M^{-1}$)
Glc	9.2	454	20.2	18.0	0.2	12.5	18.4	-	50	755	67	47
	± 0.6	*± 52*	*± 3.7*	*± 2*	*±0.009*	*±0.4*		-	*± 3*	*±110*	± 10	± 1
G3G	21.5	36	600.00	636.00	1.0	250	4.0	-	71	77	917	777
	± 1.4	*± 4*	*± 105*	*± 76*	*±0.21*	*±20*		-	*± 4*	*± 10*	*± 129*	*± 21*
G6G	48.2	86	560	697.00	6.3	51	122.3	-	-	-	-	-
	± 4.5	± 12	± 130	*± 68*	*±0.3*	*±4*		-	-	-	-	-

[a]Kinetic parameters were calculated via non-linear regression fitting to the Michaelis-Menten equation. Catalytic efficiencies (k_{cat}/K_M) were also calculated by measurement of the slope of the linear phase (low [S] << K_M) of the Michaelis-Menten plot.

[b]*Um*AA3_2-A and *Um*GDHIII are the same enzymes. *Pc*ODH was previously known as *Pc*GDH.

[c]Wijayanti et al. used benzoquinone as electron acceptor, when dichlorophenolindophenol (DCIP) was used in the present study and by Cerruti et al.

2. Full abbreviations list

Enzymes/Microorganisms

AA: Auxiliary activities
AR: AmplexRed®
CE: Carbohydrate Esterases
GH: Glycoside hydrolases
HRP: Horseradish peroxidase
Um : *Ustilago maydis*
Tsp : *Trichoderma spp.*

Substrates/products

DCIP: 2,6-Dichlorophenolindophenol
DMSO: dimethyl sulfoxide
G3G: Laminaribiose
G6G: Gentiobiose
FCW: Fungal cell wall
PCW: Plant cell wall
YPD: Yeast Extract–Peptone–Dextrose

Methods

HPAEC-PAD: high-performance anion-exchange chromatography coupled with pulsed amperometric detection
UPLC-MS: Ultra-performance liquid chromatography- mass spectrometry
UPLC-ESI-MS: Ultra-performance liquid chromatography coupled to electrospray ionization and mass spectrometry
UHPLC-ESI-IT: Ultra High-Performance Liquid Chromatography -Electrospray -Ion trap
LC-MS: Liquid chromatography–mass spectrometry
MALDI-ToF: Matrix-assisted laser desorption/ionization-Time-of-flight
GC-MS: Gas chromatography–mass spectrometry
SEC: Size-exclusion chromatography
MSA: Multiple sequences alignment
MW: Molecular weight
SDS-PAGE: Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

3. Supplementary references list

1. Lanver D, Müller AN, Happel P, Schweizer G, Haas FB, Franitza M, Pellegrin C, Reissmann S, Altmüller J, Rensing SA, Kahmann R. 2018. The biotrophic development of *Ustilago maydis* studied by RNA-seq analysis. *Plant Cell* 30:300–323.
2. Vasur J, Kawai R, Andersson E, Igarashi K, Sandgren M, Samejima M, Ståhlberg J. 2009. X-ray crystal structures of *Phanerochaete chrysosporium* Laminarinase 16A in complex with products from lichenin and laminarin hydrolysis. *FEBS J* 276:3858–3869.
3. Song J, Tan H, Perry AJ, Akutsu T, Webb GI. 2012. PROSPER: An Integrated Feature-Based Tool for Predicting Protease Substrate Cleavage Sites. *PLoS One* 7:50300.
4. Couturier M, Navarro D, Olivé C, Chevret D, Haon M, Favel A, Lesage-Meessen L, Henrissat B, Coutinho PM, Berrin JG. 2012. Post-genomic analyses of fungal lignocellulosic biomass degradation reveal the unexpected potential of the plant pathogen *Ustilago maydis*. *BMC Genomics* 13:57.