

S1 Appendix
Supporting information

for

Characterization of two family AA9 LPMOs from *Aspergillus tamarii*
that are active on xyloglucan

Antonielle V. Monclaro^{1,2}, Dejan M. Petrović², Gabriel S. C. Alves³, Marcos M. C. Costa⁴,
Glauca E. O. Midorikawa³, Robert N. G. Miller³, Edivaldo X. F. Filho¹, Vincent G. H. Eijsink²,
Anikó Várnai^{2,*}

List of content

Supplementary tables

- S1 Table.** Sequences of putative AA9 LPMOs from *Aspergillus tamarii* BLU37.
- S2 Table.** Comparison of AA9 LPMOs found in the *A. tamarii* BLU37 transcriptome with putative AA9 LPMOs encoded in the genomes of *A. oryzae* RIB40 and *A. tamarii* CBS 117626, and identification of experimentally characterized LPMOs with the highest sequence identity.
- S3 Table.** Domain structure and predicted properties of the *A. tamarii* AA9 LPMOs.
- S4 Table.** Regioselectivity and substrate specificity of *AtAA9A-N*, *AtAA9B-N*, and experimentally characterized AA9 LPMOs that are active on xyloglucan.

Supplementary figures

- S1 Fig.** Consensus and frequency of amino acids in the putative C-terminal domains of (A) *AtAA9B* and (B) *AtAA9C*.
- S2 Fig.** Phylogenetic tree of the AA9 domains of *A. tamarii* AA9s with characterized AA9 LPMOs.
- S3 Fig.** SDS-PAGE of the purified recombinant *AtAA9A-N* and *AtAA9B-N*.
- S4 Fig.** MALDI-ToF spectra showing products generated upon incubation of *AtAA9A-N* and *AtAA9B-N* with PASC.
- S5 Fig.** Multiple sequence alignment of the catalytic domains of AA9 LPMOs for which activity on xyloglucan has been demonstrated.
- S6 Fig.** Structural superposition of (substitution-intolerant) *NcAA9C-N*, *LsAA9A*, and *CvAA9A* and (substitution-tolerant) *TaAA9A*, showing surface-exposed side chains that (potentially) take part in protein–substrate interactions.
- S7 Fig.** The predicted substrate-binding surface of (A) *AtAA9A-N* and (B) *AtAA9B-N*.
- S8 Fig.** Sideview of xyloglucan-active LPMOs which have (A) substitution-intolerant, (B) substitution-tolerant, or (C) unknown cleavage pattern.

Supplementary tables

S1 Table. Sequences of putative AA9 LPMOs from *Aspergillus tamarii* BLU37. The sequences were derived from earlier RNA-sequencing of the transcriptome of *Aspergillus tamarii* BLU37 [1]. Regions annotated as AA9 domains, using the dbCAN2 metaserver (<http://bcb.unl.edu/dbCAN2/>) and by structural alignment with so far characterized AA9 LPMOs using T-Coffee's Espresso tool (<http://tcoffee.crg.cat/apps/tcoffee/do:espresso>), are marked in blue; regions annotated as CBM1 domains are marked in green; putative C-terminal domains in *AtAA9B* and *AtAA9C* are marked in red and orange, respectively (for details, see S1 Fig). N-terminal signal peptides and non-annotated regions, including regions of low sequence complexity that may be flexible linkers, appear in black. Amino acids that represent deviations from the predicted AA9 sequences in *A. tamarii* CBS 117626 [2] are highlighted in grey. (Note that the C-terminal sequences of *AtAA9D* and *AtAA9F* are truncated compared to the corresponding proteins in *A. tamarii* CBS 117626.)

>AtAA9A

MKSSSTFGMLALAAAALVSAHTTVHAVWINDVDQEGEGNSQSGYIRSPPSNSPITDVTSKDMTCNVNNKAT
AKTLEVKAGDKITFEWHHDSRSESDDI IASSHNGPILVYMAPTEKGTAGNGWVKIAEDGYTDGTWAVETL
IKNRGKHSVTVDPVAAGEYLFRRPEI IALHEGNREGGAQFYMECVQVKVTSSGSKTLPEGVSI PGAYTATD
KGILFNIYDSFDSYPIPGPAVWDGASGSSSSSSSSASASAPAPTSAPAPSSFTTIAKQPATSSSTEAPS
TENTPSETTSTTSAIVSTTAVASTTAPATPSTTSAIASSAAPTNSVPQPSSNAGGAVKEWYQCGGLNYS
STQCEEGLTCKKWNPPYHQCVSA

>AtAA9B

MSIAKIAGVVLGSAALVAGHGYVSGAVVDGQYYSGYDMSYHYMSDPPKVIWSTDATDLGFVDGSSYADA
DI ICHKNAKNGAI SAEIAAGKQVELQWTDWPESHKGPVITYLANCNGDCATVDKTQLEFFKIDKGLISG
SDNTWASDNLISSNNSWTVTIPSSIAAGNYVMRHEI IALHSAGNKDGAQNYQPCLNFKVTGGGSDKPEGT
LGTALYKDTDPGILVNIYQTLSSYTI PGPALYSGSSSSGSSSSGSSSSAAPSATASASASATAAPVQTST
ATAYQTSTAVASVTVTGSAPAQTHVQATSSSAAASTPTASSGASSGSGSSSSSSSSDLTDYFNLSADEL
LNVIKQTLSWLVTDKI HARDISA

>AtAA9C

MFRSALFLLLAPLALSHTTFTTLYVDEVNQGDTGTCVRMNRDANTVITYPIEPLSSKDIACGKDGEKAVSRV
CPAKANSLLTFEFRAWADGAQPGSIDISHKGPICAVYMKKVDDATADNNAAGDGWFKIWHTGYDESTKWC
TEKLIDNNGFLSVRVPDIEQGYLVRTPELLALHAASDAPPDPQFYVNCAQIFVQGGGSAKPETVSI GEG
YYSLDSPGVKYNIEKPLQLPYP I PGPTVYESKGV EERSVCPAQKRTATAQNKGLKPAGCILQRDNWCGF
EVPDYSDENGCWA VCSSFPYQDFMVNTNSLLVIQEV LGSE

>AtAA9D

MKLSLLAIAAIAPFVSAHYFFDTLIIDGQESSPNQYVRSNTRAAKYNPTKWNTRDNMTPDMPDFRCNKG
AFTFAGQGTAEVKAGSKLALKLGVGATMKHPGALVYMSKAPSTAKTYQGDGDWFKIYEEGVCDKNKDL
KSDAWCSWDKDRVEFTIPADLPDGEYLIRPEHIGVRVHGHAHAGEAEFYR

>AtAA9E

MAMSKIMSLTGLLASASLVAGHGYVSGVAAAAYGGYLVDKYAYS DNP PETIGWSTTATDLGFVDGTGYQSP
DI ICHKDGPALSAEVAAGGEIELQWTEWPESHHG PVLNLYLAPCGGDCSAVDKTSLEFFKIEAKGLIDG
SSPPGHWATDDLISNNSWTVTIPASVQEGNYVLRHEI IGLHSAGQKDGAQNYQPQCINIKVTGGGAATPA
GTAGEALYKDTDPGILFDIYSDLSGGYPI PGPEV FSA

>AtAA9F

MRHVQSASLLTALLSATKVAAHGHVSNIVINGVYYEGFDINSFPYMGENAPTVAAWTTPNTGNGPLAPDD
YSSPDIICHQNATAGKGYVEVNAGDRISLQWTPWPESHGPPVDYLARCEPNCASVDKTSLEFFKIDGVG
IVDGSSVPGVWGDDQLIKNNNTWLVEIPKSIAPGYVLRHELIALHSAGTEGGAQNYPSCFNLKVNGDGT
DKPAGVVGTELYTPTGDGIIFNIIYQTVSSYPVPGPTLYTGAATGVTQATSAITSTGTALTVGAAATTPAS
GSGASSAAPSSSAAATPSSRLSSLCCCVPL

>AtAA9G

MKLNLASLCLFLASIAPLVSGHYVFSKLIIVDGKTTKDFEYIRENSNGYQPTLASEIVSNDFRCNKGSMESSA
AKTKVYTVAPGAEMGFQLAYGASMKHPGLQIYMSKAPGDVKAYDGSGDWFVKVYQEGVCNDISGGLKDTD
WCTWPKDTASFKIPENTPPGQYLVRVEHIGLHRGFSGNSEFYFTCAQIEVTGSGSGVPGPLVKIPGVYKP
EDPNIHFNIYHPVPTSVDLPGPSVWSSGGVSDSSSSISAPPVNAAAASSVTPTTLVTLTKTSSTPAATSS
AAPTSSAPSNGTIKKYYQCGGQGWGSGSCEAGTSCREWNTWYFQCV

S2 Table. Comparison of AA9 LPMOs found in the *A. tamarii* BLU37 transcriptome with putative AA9 LPMOs encoded in the genomes of *A. oryzae* RIB40 and *A. tamarii* CBS 117626, and identification of experimentally characterized LPMOs with the highest sequence identity. AA9 LPMOs that have been identified in the transcriptome of *A. tamarii* BLU37 during growth on sugar cane bagasse [1] are compared to AA9 LPMOs that have been identified in the genomes of *A. oryzae* RIB40 [3] and *A. tamarii* CBS 117626 [2]. The closest (partially) characterized LPMOs from *Aspergillus* species are provided in the footnotes.

Protein from <i>A. tamarii</i> BLU37 transcriptome	Protein ID of the corresponding protein in the <i>A. oryzae</i> RIB40 genome	Protein ID in the <i>A. tamarii</i> CBS 117626 (Asptam1) genome	Closest characterized relatives ^a	Upregulation after 36 h ^h	Upregulation after 48 h ^h
<i>AtAA9A</i>	1567	254541	<i>LsAA9A</i> [4]; 58% ^b	4.7	5.3
<i>AtAA9B</i>	2234	140265	<i>TaAA9A</i> [5, 6]; 71% ^c	0.78	2.1
<i>AtAA9C</i>	4102	258171	– ^d	1.4	1.1
<i>AtAA9D</i>	4194	312543	– ^{d,e}	7.0	6.7
<i>AtAA9E</i>	4749	218852	<i>TaAA9A</i> [5, 6]; 69% ^f	10.5	11.0
<i>AtAA9F</i>	5772	312044	<i>TrAA9A</i> [7-9]; 58% ^g	1.9	-1.0
<i>AtAA9G</i>	11276	288991	– ^c	5.5	6.1
– ⁱ	9997	273838 ⁱ		– ⁱ	– ⁱ
– ⁱ		303761 ⁱ		– ⁱ	– ⁱ

^a Enzymes with known crystal structures for the catalytic domain; the sequence identity for the catalytic domain is indicated after the enzyme name.

^b Of the reported LPMOs, *AtAA9A*-N shares 80% sequence identity with An1602, for which C4-oxidizing activity on cellulose has been demonstrated [10].

^c The closest experimentally characterized relatives of *AtAA9B* are *Aspte3*, *Aspfu3*, and *Chacr2* [11] with 80-84% sequence identity; the regioselectivity of these LPMOs has not yet been identified beyond doubt and data for xyloglucan are lacking.

^d No sequence identity >50% with an experimentally characterized LPMO.

^e Of the reported LPMOs, *AtAA9D*-N shares 84% sequence identity with An3046-N; the regioselectivity of this enzyme is not known; experiments with non-purified enzymes suggest activity on cellulose and xyloglucan [12].

^f The closest experimentally characterized relatives of *AtAA9E* are *Aspfu5* [11] and *AfAA9B* [13, 14] with 72-73% sequence identity; the regioselectivity of these LPMOs has not yet been identified and experiments with xyloglucan were not reported.

^g The closest experimentally characterized relatives of *AtAA9F* are *Aspte5*, *Aspfu2*, and *Chacr1* [11] with 71-73% sequence identity; the regioselectivity of these LPMOs has not been identified beyond doubt and data for xyloglucan are lacking.

^h transcript levels during growth on sugar cane bagasse. The numbers show log₂ fold change of differentially expressed genes when comparing transcript accumulation using steam-exploded bagasse or glucose as carbon source and are from Midorikawa et al. [1]. Significant upregulation is highlighted by the orange color.

ⁱ AA9 LPMOs present in the genome of *A. tamarii* CBS 117626 for which no corresponding AA9 have been found in the *A. tamarii* BLU37 transcriptome.

S3 Table. Domain structure and predicted properties of the *A. tamarii* AA9 LPMOs. Full-length and AA9 domains only. Domains, as marked in S1 Table, were annotated by HMMER analysis using the dbCAN2 metaserver (<http://bcb.unl.edu/dbCAN2/>) and by structural alignment with experimentally characterized AA9 LPMOs using T-Coffee's Espresso tool (<http://tcoffee.crg.cat/apps/tcoffee/do:espresso>). The other listed characteristics were calculated using Expasy's ProtParam tool (<https://web.expasy.org/protparam/>).

Protein	Domain structure	Full-length enzyme ^b			AA9 domain (amino acids)		
		Length (amino acids)	Molecular mass (kDa)	pI	Length (amino acids)	Molecular mass (kDa)	pI
<i>AtAA9A</i>	AA9-linker-CBM1	353	36.68	5.00	214	23.10	4.97
<i>AtAA9B</i>	AA9-linker-unknown1	354	36.32	4.68	225	24.10	4.66
<i>AtAA9C</i>	AA9-linker-unknown2 ^c	325	35.54	4.66	247	26.93	4.66
<i>AtAA9D</i>	AA9 ^a	173					
<i>AtAA9E</i>	AA9	226	23.65	4.41	226	23.65	4.41
<i>AtAA9F</i>	AA9-linker	290	29.96	4.69	229	24.53	4.60
<i>AtAA9G</i>	AA9-linker-CBM1	307	32.75	6.09	217	23.75	5.86

^a fragment only.

^b without the signal peptide.

^c potentially truncated at the C-terminus; see main text for details.

S4 Table. Regioselectivity and substrate specificity of AtAA9A-N, AtAA9B-N, and experimentally characterized AA9 LPMOs that are active on xyloglucan. LPMOs for which there is structural data are highlighted in yellow. Regarding the activity on cello-oligosaccharides, the degree of polymerization of the tested oligomers and activity levels against those are listed. Activity on other hemicellulosic substrates, whether it has been tested (reported/not reported) and found (+/-), are also given.

Enzyme	Domain	Regio-selectivity on cellulose	Cello-oligomers ^d	L3 loop	Cleavage type on xyloglucan	Other hemicelluloses	Reference
<i>NcAA9C</i>	AA9-CBM1	C4	+ / 5-6 (+) / 4	+	substitution-intolerant	reported; +	[15]
<i>FgAA9A</i>	AA9-[...] ^a	C1/C4	- / 3-6	-	substitution-tolerant	reported; -	[16]
<i>GtAA9A-2</i>	AA9-CBMx ^b	C1/C4	- / 5-6	-	substitution-tolerant	reported; -	[17]
<i>TaAA9A</i>	AA9 only	C1/C4	- / 5	-	substitution-tolerant	reported; -	[6]
MYCTH_79765	AA9 only	C4	+ / 5-6	+	substitution-intolerant	reported; +	[18]
<i>GtAA9B</i>	AA9 only	C1/C4	- / 5-6	-	substitution-tolerant	reported; -	[19]
<i>LsAA9A</i>	AA9 only	C4	+ / 4-6	+	substitution-intolerant	reported; +	[20]
<i>CvAA9A</i>	AA9 only	C4	+ / 4-6	+	substitution-intolerant	reported; +	[20]
<i>PaAA9H</i>	AA9-CBM1	C1/C4	+ / 5	+	substitution-tolerant	reported; +	[21, 22]
MYCTH_85556	AA9 only	C1/C4	- / 1-5	-	? ^c	reported; +	[23]
MYCTH_100518	AA9 only	C4	- / 1-5	+	? ^c	reported; +	[24]
<i>NcAA9A</i>	AA9-CBM1	C4	(+) / 5	+	substitution-intolerant; ((+)) ^d	reported; -	[25]
<i>NcAA9D</i>	AA9 only	C4	- / 5	+	substitution-intolerant; (+) ^d	reported; -	[25]
<i>McAA9A</i>	AA9-[...] ^a	C1/C4	+ / 6	-	substitution-tolerant	reported; +	[26]
<i>McAA9B</i>	AA9 only	C1/C4	- / 6	-	substitution-tolerant	reported; +	[26]
<i>McAA9F</i>	AA9 only	C1/C4	+ / 6	-	substitution-tolerant	reported; +	[26]
<i>McAA9H</i>	AA9 only	C1/C4	- / 6	-	substitution-tolerant ^e	reported; +	[26]
<i>GcAA9A</i>	AA9 only	C1/C4	not reported	-	? ^c	not reported	[27]
<i>GcAA9B</i>	AA9-[...] ^a	C1/C4	not reported	-	? ^c	not reported	[27]
<i>An3046</i>	AA9-[...] ^a	? ^c	not reported	-	? ^c	not reported	[12]
<i>TtAA9E</i>	AA9 only	C1	not reported	-	? ^c	not reported	[28]
<i>NcAA9M</i>	AA9 only	C1/C4	not reported	-	substitution-tolerant	not reported	[29]
<i>AtAA9A-N</i>	AA9-CBM1	C4	+ / 5-6	+	substitution-intolerant	reported; -	this study
<i>AtAA9B-N</i>	AA9-[...] ^a	C1/C4	- / 5-6	-	substitution-tolerant	reported; -	this study

^a unidentified C-terminal extension

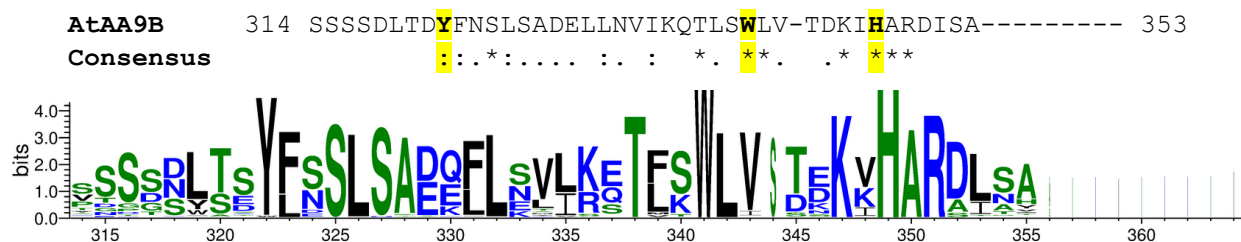
^b unclassified carbohydrate-binding module (CBM)

^c unclear; it cannot be determined from the available data

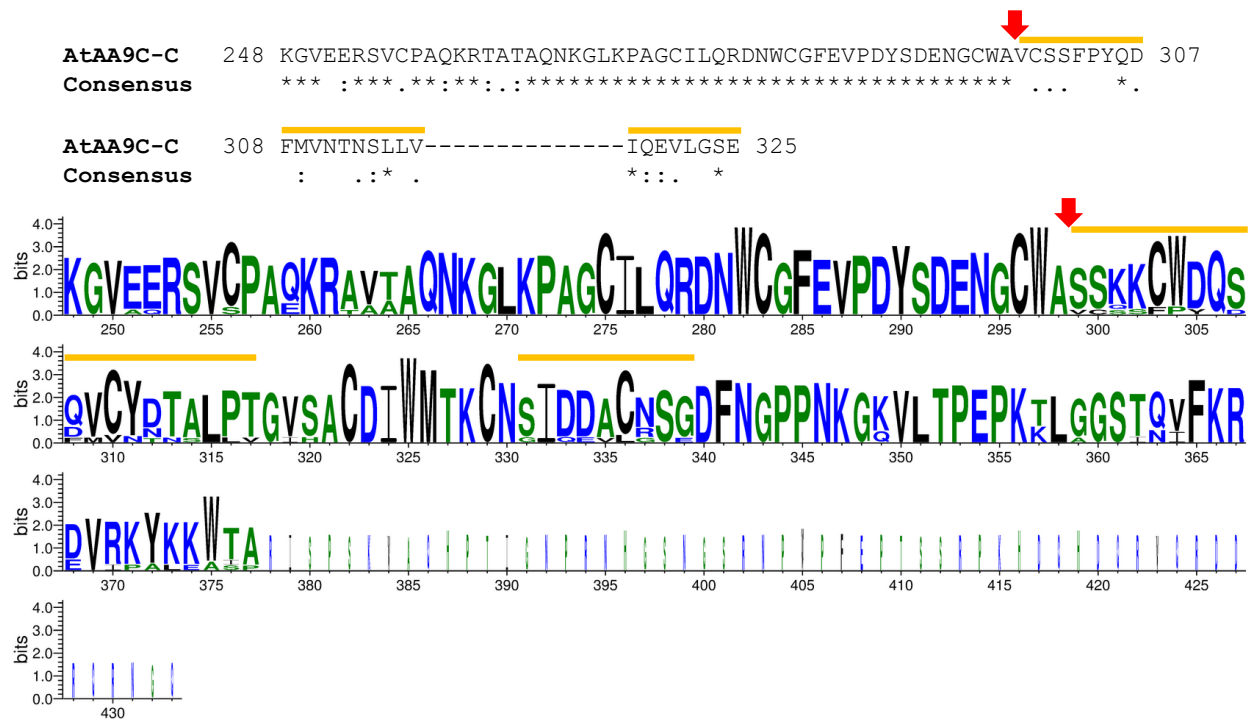
^d activity levels are indicated as: -, inactive; ((+)), hardly active; (+) little active; +, active

^e activity on xyloglucan was only found when co-incubating xyloglucan and cellulose

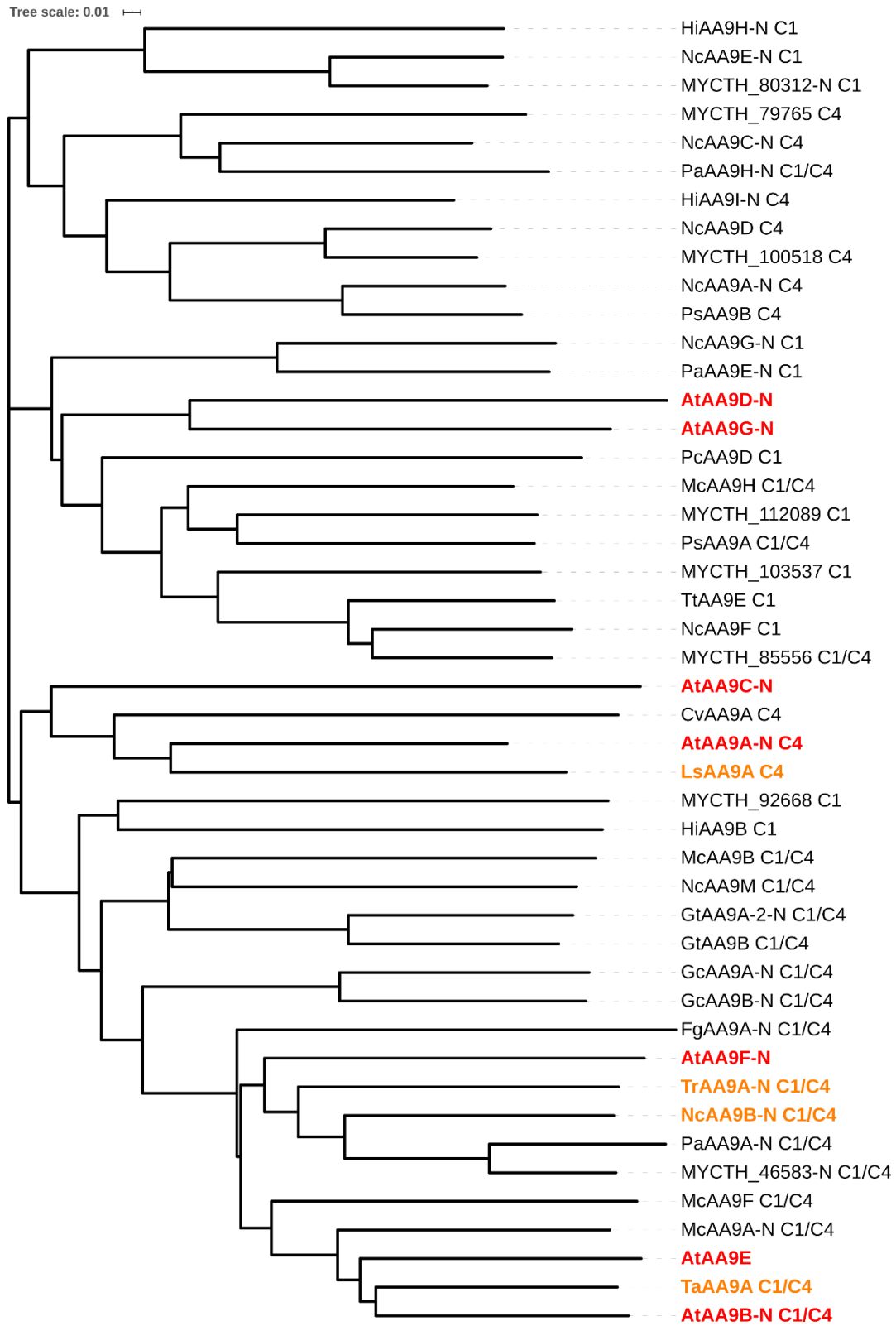
Supplementary figures



S1A Fig. Consensus and frequency of amino acids in the putative C-terminal domain of *AtAA9B*. Consensus was based on multiple sequence alignment of the putative C-terminal domain of *AtAA9B* (“unknown 1” in S3 Table) with similar domains in other proteins in the UniProt database; the frequency of amino acids was visualized using WebLogo 3 (<http://weblogo.threeplusone.com/create.cgi>). Altogether 98 proteins were found (with an E-value <0.0001) when blasting the C-terminus of *AtAA9B* (top of the figure) against the UniProt database. All these proteins were identified as AA9 LPMOs using the dbCAN2 metaserver (<http://bcb.unl.edu/dbCAN2/>). Notably, 96 of the 98 proteins were from *Aspergillus* or *Penicillium* species. The conserved aromatic residues Y322 (Y in 96 and W in 2 sequences), W341, and H348 are highlighted in yellow.

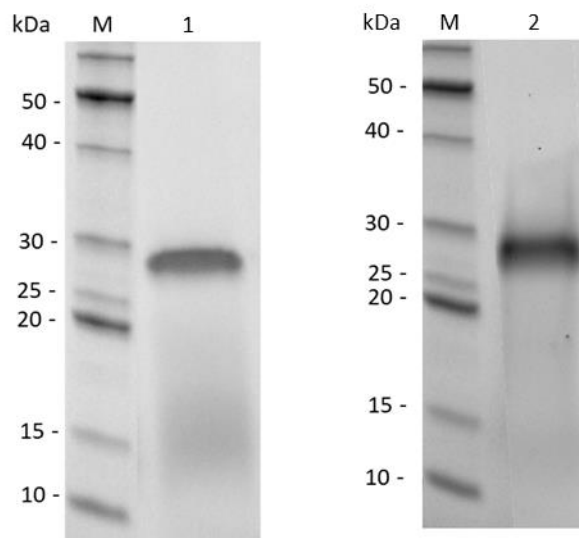


S1B Fig. Consensus and frequency of amino acids in the putative C-terminal domain of *AtAA9C*. Consensus was based on multiple sequence alignment of the putative C-terminal domain of *AtAA9C* (“unknown 2” in S3 Table) with similar domains in other proteins in the UniProt database; the frequency of amino acids was visualized using WebLogo 3 (<http://weblogo.threeplusone.com/create.cgi>). Altogether 136 proteins were found (with an E-value <0.0001) when blasting the C-terminus of *AtAA9C* (top of the figure) against the UniProt database. All these proteins were identified as AA9 LPMOs using the dbCAN2 metaserver (<http://bcb.unl.edu/dbCAN2/>), with the exception of one, which is an LPMO fragment (based on multiple sequence alignment). This figure is based on the alignment of the C-termini of *AtAA9C* and eight proteins (i.e. the proteins with >90% sequence identity) in the UniProt database. The C-terminal sequence of *AtAA9C* seems to be truncated compared to the sequences of these proteins (with >90% sequence identity); the LPMOs shown in the sequence logo are approximately 50 amino acids longer. The putative truncation point in *AtAA9C* is indicated by a red arrow and the sequence after the truncated point by orange bars above the sequence and the sequence logo.

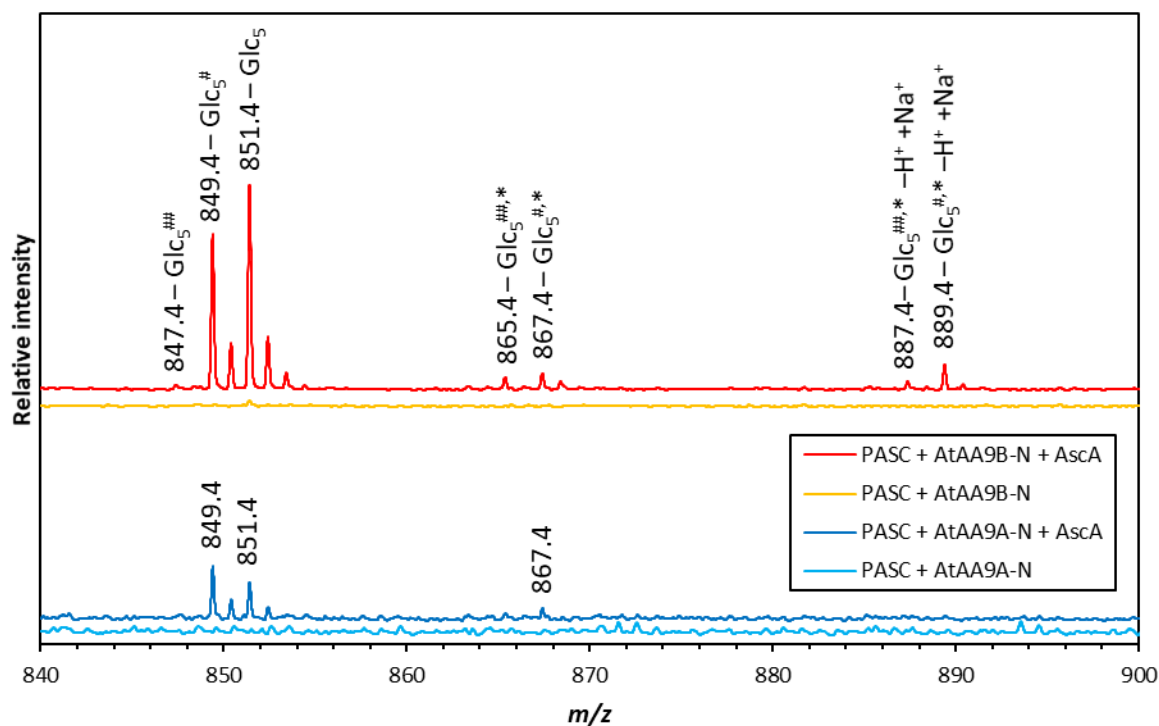


S2 Fig. See legend on the next page.

S2 Fig. Phylogenetic tree based on a multiple sequence alignment of the AA9 domains of *A. tamaritii* AA9s with characterized AA9 LPMOs. LPMOs where the reported data were unclear or insufficient to unambiguously identify regioselectivity on cellulose were omitted from the figure. An “-N” after the LPMO name indicates the presence of a C-terminal extension that was omitted from the comparison. *AtAA9*s are indicated in red and bold; the closest characterized relatives of the *AtAA9*s are indicated in orange and bold. Regioselectivity on cellulose (C1-, C4- and C1/C4-oxidizing) is given after the name of each LPMO.



S3 Fig. SDS-PAGE of the purified recombinant *AtAA9A-N* and *AtAA9B-N*. Lane M, Benchmark Protein Ladder; Lane 1, purified *AtAA9A-N*; Lane 2, purified *AtAA9B-N*.



S4 Fig. MALDI-ToF spectra showing products generated upon incubation of *AtAA9A-N* and *AtAA9B-N* with PASC. The spectra show the DP 5 cluster displaying a pattern typical for C4-oxidation for *AtAA9A-N*, while the pattern for *AtAA9B-N* is typical for C1/C4-oxidation. Single or double oxidation is denoted with # or ##; hydration is denoted with *. The peaks at 887.4 and 889.4 correspond to the Na⁺-salt of the aldonic acid, which is denoted as “-H⁺ +Na⁺”. When it comes to regioselectivity, the clearest indicators are: absence of sodium salts of aldonic acids in the spectrum for *AtAA9A-N*, which indicates the absence of C1-oxidation; presence of aldonic acids and double oxidized products in the spectrum for *AtAA9B-N*, which indicates that both C1- and C4-oxidation occur. Further evidence for the regioselectivity of the two LPMOs is provided by the HPAEC chromatograms shown in Fig. 1 of the main manuscript.

			130	140	150	160	170	180	
								
McAA9B	99	P-G--DCSEYEP-GTDAVWFKIAEDGKH----DDG-SWASDPLIN-DV-PYEFTIPEGLA							147
NcAA9M	91	P-D-TGCQDWTSPASDKVWFKIKEGGREG---TSN-VWAATPLMTAP-ANYEYAI P SC L K							143
GtAA9A-2-N	98	P-SNTDITSYSPTGSDVIWFKIDEAGYE-----NG-KWAATD I MSA Q NST W TV T IP K LA							150
GtAA9B	98	P-SSTDVTKYSPTGSDVIWFKIDEAGYE-----NG-KWAATD V LSA Q NST W TV T IP S SLA							150
FgAA9A-N	98	G-SAG-CAKVDK--TSLKFFKIAEAGMTS----GG-KFASDD L IA A G-NTWEVTV P TS I K							147
McAA9F	93	N-GD--CSSASA--GSLNFVKIAEKGLISGS-NPG-FWAAD E LI Q NG-NSWEV T IPAN L A							144
McAA9A-N	98	N-GD--CSSVDK--TSLKFFKISEAGLNDG S NA P G-QWASDD L IA N N-NSWT V TIP K S I A							150
AtAA9B-N	97	N-GD--CATVDK--TQLEFFKID E K L IS G --SDN-TWASD N LI S S N -NSWT V TIP S S I A							147
TaAA9A	98	N-GD--CSTVDK--TQLEFFKIA E S L IND N PP G -IWASD N LI A AN-NSWT V TIP T T I A							150
CvAA9A	91	TS-----PTMDG--TGPV W SKI H E E G D AST---K-SWAVD K LI A N K -GMW D FT L PS Q L K							138
AtAA9A-N	94	EK-----GT--AGNGW V KIA E D G YT--D---G-TWAV E T L IK N R-GK H SV T VP D -VA							135
LsAA9A	90	AS-----NG--QGN V W V K L F E D A YN V T N ---S-TWAV D R L IT A H-G Q H S V V VP H -VA							133
McAA9H	84	PDGQ-TAD S WD G --SGQ V W F KI F E Q GP Q ID P -S L G-TW P SD G L-----S Q V Q V T IP S SL P							133
NcAA9A-N	93	D-D---AL T D T G--IGG G W F KI Q E D G Y NN-----G-Q W G T ST V IT N G-G F Q Y ID I P A C I P							139
NcAA9D	96	D-N---A A T A S G --V G NG W F K I Q Q D G M D S ----S G -V W G T ER V ING K -GR H S I KI P E C I A							143
MYCTH_79765	95	D-N---A A T A S K --T G L K W F KI E W E D T F N P-----S T K T W G V D N L I N N N -G W V F N L P Q C I A							143
NcAA9C-N	95	D-N---A A T T G T --S G L K W F K V A E A G LS N -----G-K W AV D D L I A N N -G W S Y F D M P T C I A							141
PaAA9H-N	96	N-N---A A N A G T --S L Q W F K V A E Q GL N N-----G-V W AV D N M I S NG-G W H Y F D M P S C V A							142

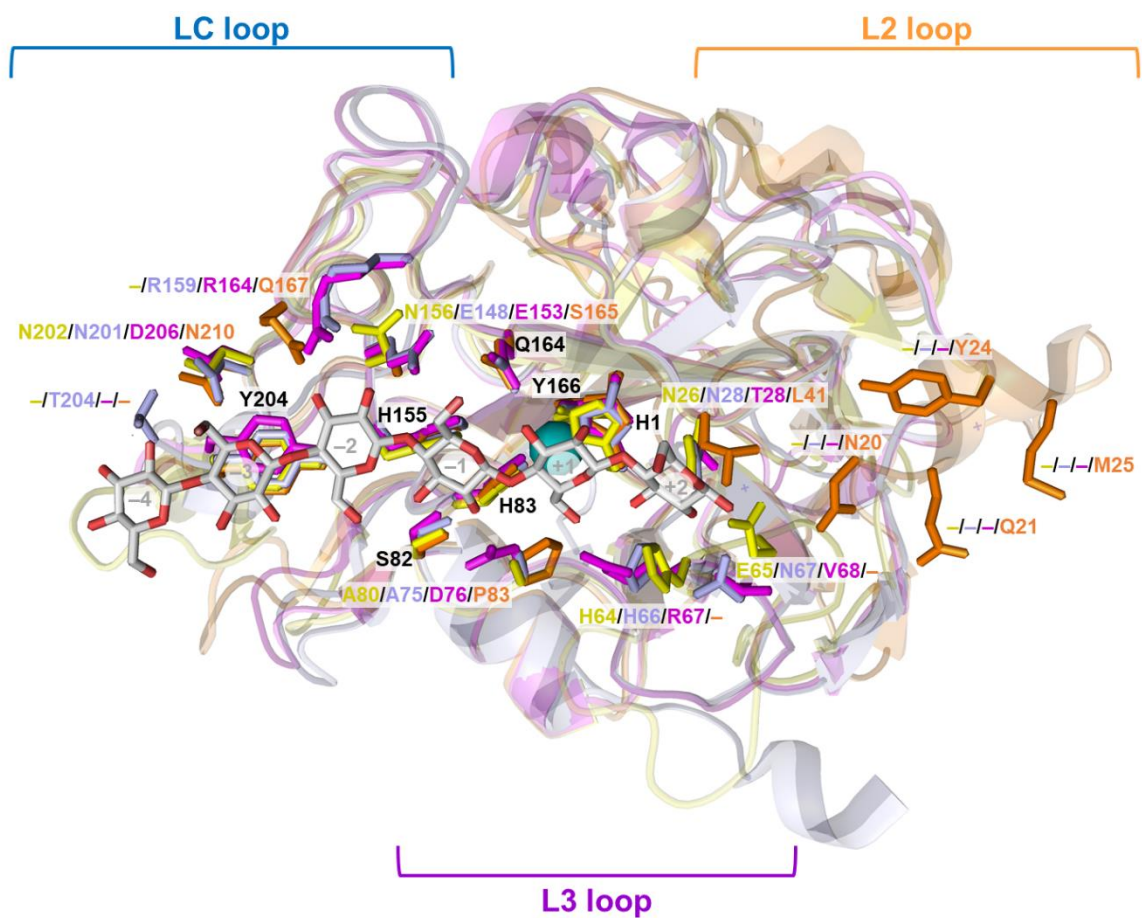
Consensus : *: : : : : : * :

			190	200	210	220	230	240	
								
McAA9B	148	PGNYIVRHE L W L A A W T -----Y P G A Q V Y P S C F Q V K V V G D G-T Q -Q T -N--L V A F P G							196
NcAA9M	144	PGYYLVRHE I I A L H A S Y-----Y P G A Q F Y P G C H Q L Q V T G S G-T K -T P S-S-G L V S F P G							193
GtAA9A-2-N	151	PGQYIVRHE I I A L H A E T-----Y P G A Q F Y P D C F Q V Q V T G P G-T E -T P T-S Q A L V S F P G							201
GtAA9B	151	PGQYIVRHE I I A L H A Q T-----Y P G A Q S Y P D C F Q I R V T G S G-N K -T P S-G S Y L V S F P G							201
FgAA9A-N	148	AGNYVLRHE I I A L H A G Q-----E N G A Q N Y P Q C F N L E V E S D G-T A -E P A----G V A G T S							195
McAA9F	145	PGKYVLRHE I I A L H A G N-----P N G A Q A Y P Q C I N L E V T G G G-S A -T P S----G Q P A T S							192
McAA9A-N	151	PGNYVLRHE I I A L H A G N-----Q N G A Q N Y P Q C F N L E I T S N G-S D -N P E----G V L G T E							198
AtAA9B-N	148	AGNYVMRHE I I A L H A G N-----K D G A Q N Y P Q C L N F K V T G G G-S D -K P E----G T L G T A							195
TaAA9A	151	PGNYVLRHE I I A L H A Q N-----Q D G A Q N Y P Q C I N L Q V T G G G-S D -N P A----G T L G T A							198
CvAA9A	139	PGKYMLR Q E I V A H H E S D A T F D K N P K R G A Q F Y P S C V Q V D V K G V G-G D -A V P D Q--A F D F N K							194
AtAA9A-N	136	AGEYLF R P E I I A L H E G N R-----E G G A Q F Y M E C V Q V K V T S S G-S K -T L P-E--G V S I P G							184
LsAA9A	134	PGDYLF R A E I I A L H E A D S L Y S Q N P I R G A Q F Y I S C A Q I T I N S S D -D S -T P L P A--G V P F P G							189
McAA9H	134	SGDYLLR V E Q I G L H A S S-----V N G A Q F Y L S C A Q L T V T G G G-N G -N P G-P--L V S F P G							182
NcAA9A-N	140	SGQYLLR A E M I A L H A S S-----T A G A Q L Y M E C A Q I N I V G G T G G T A L P S -T--T Y S I P G							190
NcAA9D	144	PGQYLLR A E M I A L H A S N-----Y P G A Q F Y M E C A Q L N V V G G T-G A -K T P-S--T V S F P G							192
MYCTH_79765	144	DGNYLLR V E V L A L H A S Y-----Q Q Q A Q F Y Q S C A Q I N V S G G G-S F -T P A-S--T V S F P G							192
NcAA9C-N	142	PGQYLMR A E L I A L H A G S-----Q A G A Q F Y I G C A Q I N V T G G G-S A -S P S-N--T V S F P G							190
PaAA9H-N	143	PGHYLMR V E L L A L H A S V-----R G A A Q F Y M E C A Q I E I T G S G-T N -T G S-N--F V S F P G							191

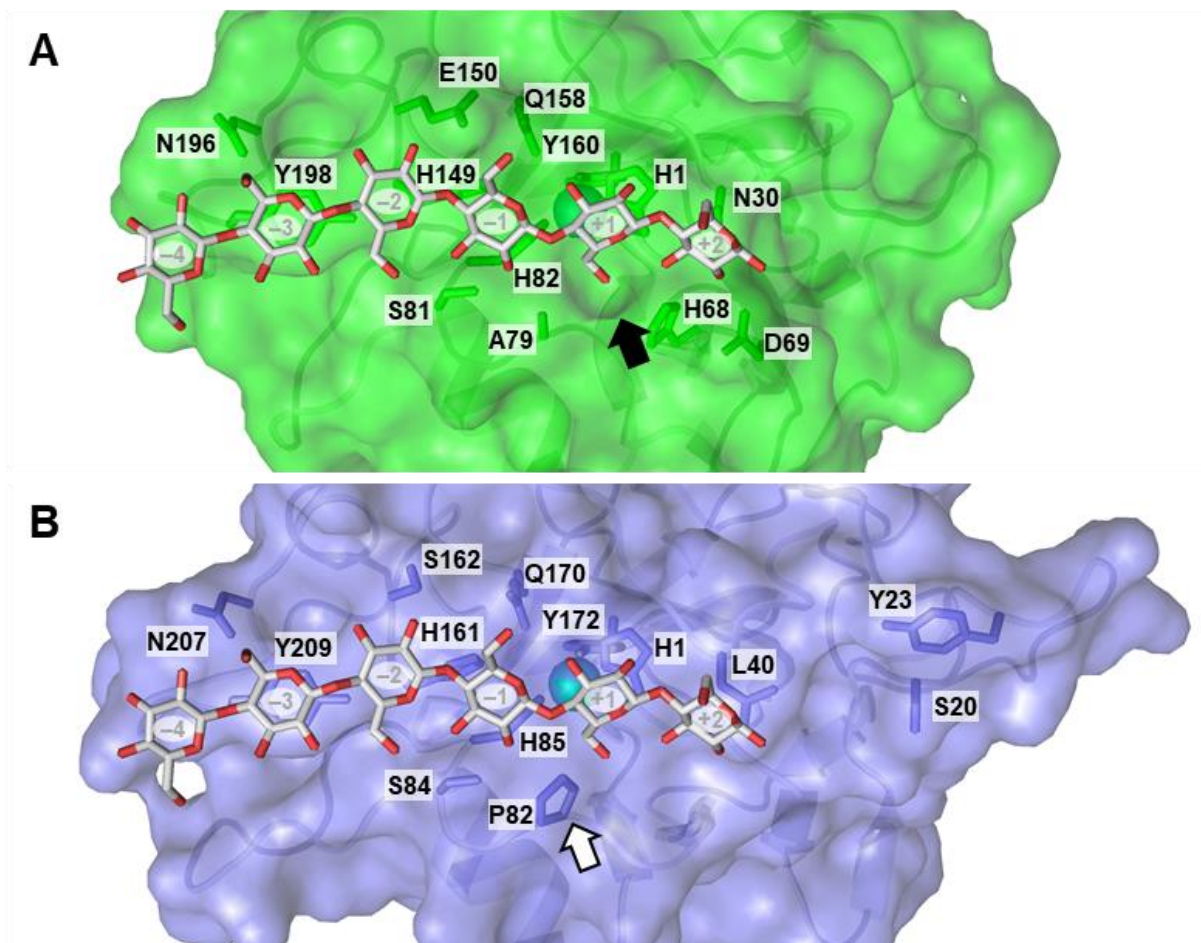
Consensus * * : . * * . * . ** * * : . : .

S5 Fig. (continued)

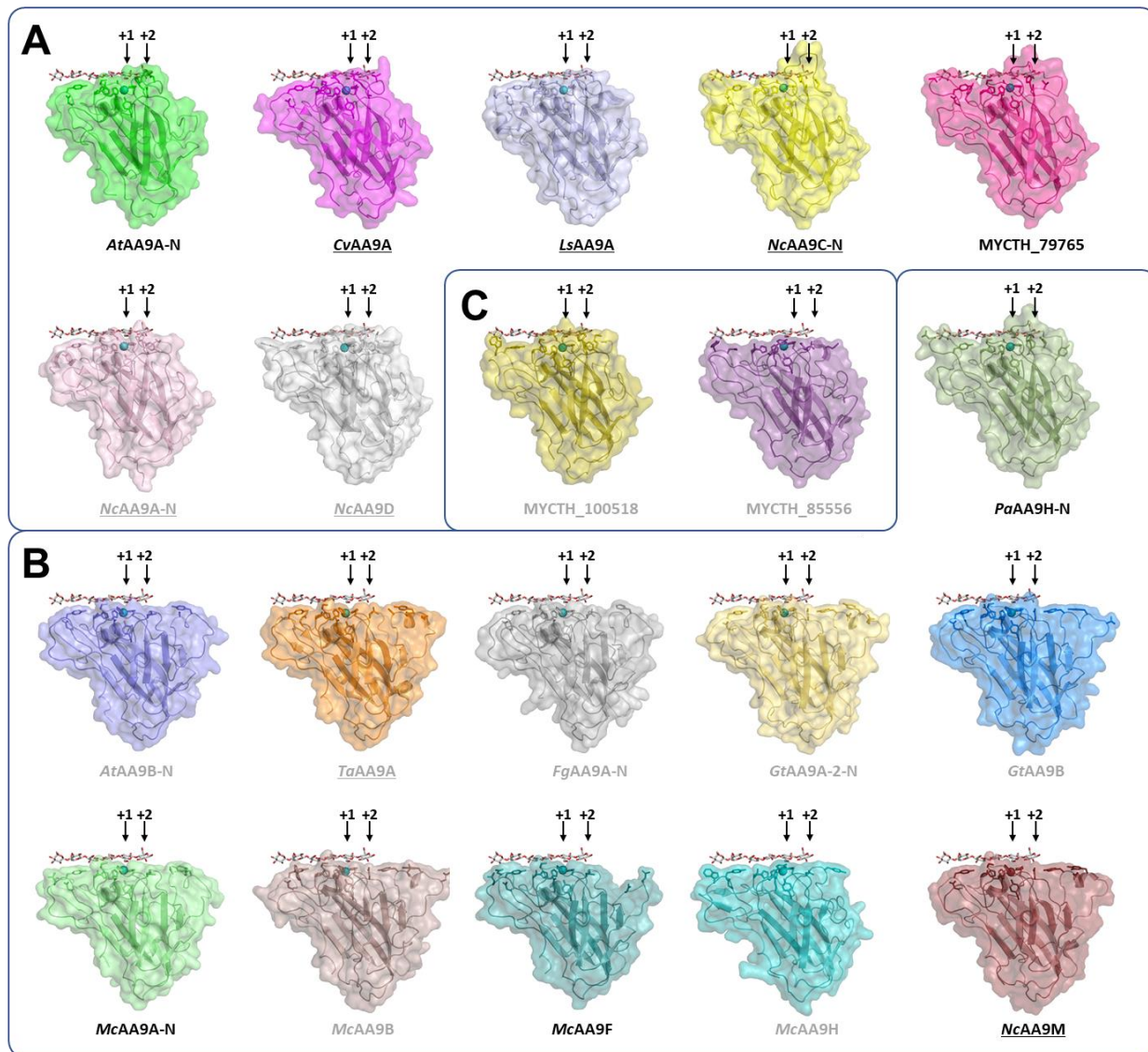
been reported; its product profile on PASC (short cello-oligosaccharides with DP 2-4) is similar to that of *NcAA9C*, indicating that activity on soluble cello-oligosaccharides is very likely [30].) The two histidines and the axial tyrosine that coordinate the active site copper are marked with blue triangles and are highlighted in blue. Additional residues potentially involved in substrate–protein interactions (as experimentally identified for *NcAA9C* [31] and *LsAA9A* [4, 20]) are highlighted in green and marked with green triangles. A few additional residues, which are discussed in the main text, because we speculate that they may affect xyloglucan binding or the ability to cleave water-soluble cello-oligosaccharides, appear in black frames.



S6 Fig. Structural superposition of (substitution-intolerant) *NcAA9C-N* (PDB: 4D7U; gold), *LsAA9A* (PDB: 5ACI; light purple) and *CvAA9A* (PDB: 5NLT; magenta) with (substitution-tolerant) *TaAA9A* (PDB: 2YET; orange), showing surface-exposed side chains that (potentially) take part in protein–substrate interactions. The L2, L3 and LC loops are marked using the color coding used in S4 Fig. Side chain labels are colored according to the corresponding structure; the labels of fully conserved residues appear in black with numbering referring to *NcAA9C*.



S7 Fig. The predicted substrate-binding surface of (A) *AtAA9A-N* and (B) *AtAA9B-N*. Structures were built with SWISS MODEL [32] based on the structures of *LsAA9A* (PDB ID, 5N05) and *TaAA9A* (PDB ID, 3ZUD), respectively. Analogously to Fig 7, panel **A** shows a cavity (black arrow) formed by the L3 loop of *AtAA9A*, and panel **B** shows the lack of cavity and a conserved surface-exposed proline (white arrow) in *AtAA9B*. The cellohexaose was superposed from the *LsAA9A*-cellohexaose (PDB ID, 5ACI) structure.



S8 Fig. Sideview of xyloglucan-active LPMOs which have (A) substitution-intolerant, (B) substitution-tolerant, or (C) unknown cleavage pattern. The names of LPMOs with known crystal structures (*CvAA9A*, PDB: 5NLT; *LsAA9A*, PDB: 5ACI; *NcAA9A-N*, PDB: 5FOH; *NcAA9C-N*, PDB: 4D7U; *NcAA9D*, PDB: 4EIR; *NcAA9M*, PDB: 4EIS; *TaAA9A*, PDB: 2YET) are underlined. The other structures shown are models based on PDB structures with ID 3ZUD (*AtAA9B-N*, *FgAA9A-N*), 4B5Q (*McAA9H*), 4D7U (*MYCTH_79765*, *PaAA9H-N*), 4EIR (*MYCTH_100518*), 4EIS (*GtAA9A-2-N*, *GtAA9B*, *McAA9B*), 4QI8 (*MYCTH_85556*), 5N05 (*AtAA9A-N*) and 6H1Z (*McAA9A-N*, *McAA9F*). The positioning of cellohexaose (Cell_6) was modeled based on the *LsAA9A*– Cell_6 structure (5ACI). Note that much of the variation between LPMOs occurs in the surface loops that make up the substrate-binding surface and that such loops are difficult to model accurately. Thus, the models cannot be used for detailed analysis of putative enzyme-substrate interactions.

The L3 loop appears as a protrusion of varying height behind the +1 and +2 subsites that are marked with an arrow and is present in all LPMOs in panel **A**, in *PaAA9H-N* in panel **B**, and in MYCTH_100518 in panel **C**. The LPMOs in panel **B** (except *PaAA9H-N*) and MYCTH_85556 in panel **C** have a more open and flat substrate-binding surface, which is extended towards the left (+ subsites) due to a longer L2 region (except in the case of *McAA9H* and *PaAA9H*). The side chains of (putative) substrate-binding residues including the His-brace, surface-exposed aromatic residues and substrate-binding residues identified by NMR [31] and crystallography [20] are shown (see S4 Fig for more details). The names of LPMOs that are active on cello-oligosaccharides appear in black, whereas the names of LPMOs that are not active on cello-oligosaccharides appear in grey.

References

1. Midorikawa GEO, Correa CL, Noronha EF, Filho EXF, Togawa RC, Costa M, et al. Analysis of the transcriptome in *Aspergillus tamaritii* during enzymatic degradation of sugarcane bagasse. *Front Bioeng Biotechnol.* 2018;6:123.
2. Kjaerbølling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, Kildgaard S, et al. A comparative genomics study of 23 *Aspergillus* species from section *Flavi*. *Nat Commun.* 2020;11(1):1106.
3. Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, et al. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature.* 2005;438(7071):1157-61.
4. Frandsen KE, Simmons TJ, Dupree P, Poulsen JC, Hemsworth GR, Ciano L, et al. The molecular basis of polysaccharide cleavage by lytic polysaccharide monoxygenases. *Nat Chem Biol.* 2016;12(4):298-303.
5. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JC, Johansen KS, et al. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA.* 2011;108(37):15079-84.
6. Petrović DM, Bissaro B, Chylenski P, Skaugen M, Sørli M, Jensen MS, et al. Methylation of the N-terminal histidine protects a lytic polysaccharide monoxygenase from auto-oxidative inactivation. *Protein Sci.* 2018;27(9):1636-50.
7. Karlsson J, Saloheimo M, Siika-aho M, Tenkanen M, Penttila M, Tjerneld F. Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei*. *Eur J Biochem.* 2001;268(24):6498-507.
8. Pierce BC, Agger JW, Wichmann J, Meyer AS. Oxidative cleavage and hydrolytic boosting of cellulose in soybean spent flakes by *Trichoderma reesei* Cel61A lytic polysaccharide monoxygenase. *Enzyme Microb Technol.* 2017;98:58-66.
9. Hansson H, Karkehabadi S, Mikkelsen N, Douglas NR, Kim S, Lam A, et al. High-resolution structure of a lytic polysaccharide monoxygenase from *Hypocrea jecorina* reveals a predicted linker as an integral part of the catalytic domain. *J Biol Chem.* 2017;292(46):19099-109.
10. Jagadeeswaran G, Gainey L, Mort AJ. An AA9-LPMO containing a CBM1 domain in *Aspergillus nidulans* is active on cellulose and cleaves cello-oligosaccharides. *AMB Express.* 2018;8(1):171.
11. Pierce BC, Agger JW, Zhang Z, Wichmann J, Meyer AS. A comparative study on the activity of fungal lytic polysaccharide monoxygenases for the depolymerization of cellulose in soybean spent flakes. *Carbohydr Res.* 2017;449:85-94.
12. Jagadeeswaran G, Gainey L, Prade R, Mort AJ. A family of AA9 lytic polysaccharide monoxygenases in *Aspergillus nidulans* is differentially regulated by multiple substrates and at least one is active on cellulose and xyloglucan. *Appl Microbiol Biotechnol.* 2016;100(10):4535-47.
13. de Gouvêa PF, Gerolamo LE, Bernardi AV, Pereira LMS, Uyemura SA, Dinamarco TM. Lytic polysaccharide monoxygenase from *Aspergillus fumigatus* can improve enzymatic cocktail activity during sugarcane bagasse hydrolysis. *Protein Pept Lett.* 2019;26(5):377-85.
14. Lo Leggio L, Weihe CD, Poulsen JN, Sweeney M, Rasmussen F, Lin J, et al. Structure of a lytic polysaccharide monoxygenase from *Aspergillus fumigatus* and an engineered thermostable variant. *Carbohydr Res.* 2018;469:55-9.

15. Agger JW, Isaksen T, Várnai A, Vidal-Melgosa S, Willats WGT, Ludwig R, et al. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc Natl Acad Sci USA*. 2014;111(17):6287-92.
16. Nekiunaite L, Petrović DM, Westereng B, Vaaje-Kolstad G, Hachem MA, Várnai A, et al. FgLPMO9A from *Fusarium graminearum* cleaves xyloglucan independently of the backbone substitution pattern. *FEBS Lett*. 2016;590(19):3346-56.
17. Kojima Y, Várnai A, Ishida T, Sunagawa N, Petrović DM, Igarashi K, et al. Characterization of an LPMO from the brown-rot fungus *Gloeophyllum trabeum* with broad xyloglucan specificity, and its action on cellulose-xyloglucan complexes. *Appl Environ Microbiol*. 2016;82(22):6557-72.
18. Kadowaki MAS, Várnai A, Jameson JK, AE TL, Costa-Filho AJ, Kumagai PS, et al. Functional characterization of a lytic polysaccharide monooxygenase from the thermophilic fungus *Myceliophthora thermophila*. *PLoS ONE*. 2018;13(8):e0202148.
19. Hegnar OA, Petrović DM, Bissaro B, Alfredsen G, Várnai A, Eijsink VGH. pH-dependent relationship between catalytic activity and hydrogen peroxide production shown via characterization of a lytic polysaccharide monooxygenase from *Gloeophyllum trabeum*. *Appl Environ Microbiol*. 2019;85(5):pii: e02612-18.
20. Simmons TJ, Frandsen KEH, Ciano L, Tryfona T, Lenfant N, Poulsen JC, et al. Structural and electronic determinants of lytic polysaccharide monooxygenase reactivity on polysaccharide substrates. *Nat Commun*. 2017;8(1):1064.
21. Bennati-Granier C, Garajova S, Champion C, Grisel S, Haon M, Zhou S, et al. Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. *Biotechnol Biofuels*. 2015;8:90.
22. Fanuel M, Garajova S, Ropartz D, McGregor N, Brumer H, Rogniaux H, et al. The *Podospora anserina* lytic polysaccharide monooxygenase PaLPMO9H catalyzes oxidative cleavage of diverse plant cell wall matrix glycans. *Biotechnol Biofuels*. 2017;10:63.
23. Frommhagen M, Sforza S, Westphal AH, Visser J, Hinz SW, Koetsier MJ, et al. Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnol Biofuels*. 2015;8:101.
24. Frommhagen M, Koetsier MJ, Westphal AH, Visser J, Hinz SW, Vincken JP, et al. Lytic polysaccharide monooxygenases from *Myceliophthora thermophila* C1 differ in substrate preference and reducing agent specificity. *Biotechnol Biofuels*. 2016;9(1):186.
25. Petrović DM, Várnai A, Dimarogona M, Mathiesen G, Sandgren M, Westereng B, et al. Comparison of three seemingly similar lytic polysaccharide monooxygenases from *Neurospora crassa* suggests different roles in plant biomass degradation. *J Biol Chem*. 2019;294(41):15068-81.
26. Hüttner S, Várnai A, Petrović DM, Bach CX, Kim Anh DT, Thanh VN, et al. Specific xylan activity revealed for AA9 lytic polysaccharide monooxygenases of the thermophilic fungus *Malbranchea cinnamomea* by functional characterization. *Appl Environ Microbiol*. 2019;85(23):pii: e01408-19.
27. Ladevèze S, Haon M, Villares A, Cathala B, Grisel S, Herpoël-Gimbert I, et al. The yeast *Geotrichum candidum* encodes functional lytic polysaccharide monooxygenases. *Biotechnol Biofuels*. 2017;10(1):215.
28. Cannella D, Mollers KB, Frigaard NU, Jensen PE, Bjerrum MJ, Johansen KS, et al. Light-driven oxidation of polysaccharides by photosynthetic pigments and a metalloenzyme. *Nat Commun*. 2016;7:11134.

29. Sun P, Laurent CVFP, Scheiblbrandner S, Frommhagen M, Kouzounis D, Sanders MG, et al. Configuration of active site segments in lytic polysaccharide monooxygenases steers oxidative xyloglucan degradation. *Biotechnol Biofuels*. 2020;13:95.
30. Vu VV, Beeson WT, Phillips CM, Cate JHD, Marletta MA. Determinants of regioselective hydroxylation in the fungal polysaccharide monooxygenases. *J Am Chem Soc*. 2014;136(2):562-5.
31. Courtade G, Wimmer R, Røhr ÅK, Preims M, Felice AK, Dimarogona M, et al. Interactions of a fungal lytic polysaccharide monooxygenase with beta-glucan substrates and cellobiose dehydrogenase. *Proc Natl Acad Sci USA*. 2016;113(21):5922-7.
32. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-303.