## S1 Appendix

## Supporting information

for

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

Antonielle V. Monclaro<sup>1,2</sup>, Dejan M. Petrović<sup>2</sup>, Gabriel S. C. Alves<sup>3</sup>, Marcos M. C. Costa<sup>4</sup>, Glaucia E. O. Midorikawa<sup>3</sup>, Robert N. G. Miller<sup>3</sup>, Edivaldo X. F. Filho<sup>1</sup>, Vincent G. H. Eijsink<sup>2</sup>, Anikó Várnai<sup>2,\*</sup>

## 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 *At*AA9A-N, *At*AA9B-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) *At*AA9B and (B) *At*AA9C.
- 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 *At*AA9A-N and *At*AA9B-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) *Nc*AA9C-N, *Ls*AA9A, and *Cv*AA9A and (substitution-tolerant) *Ta*AA9A, showing surface-exposed side chains that (potentially) take part in protein–substrate interactions.
- S7 Fig. The predicted substrate-binding surface of (A) *At*AA9A-N and (B) *At*AA9B-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 annotated AA9 domains, using the [1]. Regions as dbCAN2 metaserver (http://bcb.unl.edu/dbCAN2/) and by structural alignment with so far characterized AA9 LPMOs using T-Coffee's Expresso tool (http://tcoffee.crg.cat/apps/tcoffee/do:expresso), 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). Nterminal 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 Cterminal sequences of AtA9D and AtA9F are truncated compared to the corresponding proteins in A. tamarii CBS 117626.)

#### >Ataa9a

MKSSTFGMLALAAAAKLVSAHTTVHAVWINDVDQGEGNSQSGYIRSPPSNSPITDVTSKDMTCNVNNKAT AKTLEVKAGDKITFEWHHDSRSESDDIIASSHNGPILVYMAPTEKGTAGNGWVKIAEDGYTDGTWAVETL IKNRGKHSVTVPDVAAGEYLFRPEIIALHEGNREGGAQFYMECVQVKVTSSGSKTLPEGVSIPGAYTATD KGILFNIYDSFDSYPIPGPAVWDGASGSSSSSSSASASAPAPTSAAPAPSSFTTIAKQPATSSSTEAPS TENTPSETTSTTSAIVSTTAVASTTAPATPSTTSAIASSAAPTNSVPQPSSNAGGAVKEWYQCGGLNYSG STQCEEGLTCKKWNPYYHQCVSA

#### >AtAA9B

MSIAKIAGVVLGSAALVAGHGYVSGAVVDGQYYSGYDMSYHYMSDPPKVIGWSTDATDLGFVDGSSYADA DIICHKNAKNGAISAEIAAGKQVELQWTDWPESHKGPVITYLANCNGDCATVDKTQLEFFKIDEKGLISG SDNTWASDNLISSNNSWTVTIPSSIAAGNYVMRHEIIALHSAGNKDGAQNYPQCLNFKVTGGGSDKPEGT LGTALYKDTDPGILVNIYQTLSSYTIPGPALYSGSSSGSSSGSSGSSGSSSAAPSATASASASATAAPVQTST ATAYQTSTAVASVTVTGSAPAQTHVQATSSSAAASTPTASSGASSGSSSSSSSSSDLTDYFNSLSADEL LNVIKQTLSWLVTDKIHARDISA

#### >AtAA9C

MFRSALFLLLAPLALSHTTFTTLYVDEVNQGDGTCVRMNRDANTVTYPIEPLSSKDIACGKDGEKAVSRV CPAKANSLLTFEFRAWADGAQPGSIDISHKGPCAVYMKKVDDATADNNAAGDGWFKIWHTGYDESTEKWC TEKLIDNNGFLSVRVPSDIEQGYYLVRTELLALHAASDAPPDPQFYVNCAQIFVQGGGSAKPETVSIGEG YYSLDSPGVKYNIYEKPLQLPYPIPGPTVYESKGVEERSVCPAQKRTATAQNKGLKPAGCILQRDNWCGF EVPDYSDENGCWAVCSSFPYQDFMVNTNSLLVIQEVLGSE

#### >AtAA9D

MKLSLLAIAAIAPFVSAHYFFDTLIIDGQESSPNQYVRSNTRAAKYNPTKWVNTRDNMTPDMPDFRCNKG AFTFAGQTGTAEVKAGSKLALKLGVGATMKHPGPALVYMSKAPSTAKTYQGDGDWFKIYEEGVCDKNKDL KSDAWCSWDKDRVEFTIPADLPDGEYLIRPEHIGVRVHGAHAGEAEFYYR

#### >AtAA9E

MAMSKIMSLTGLLASASLVAGHGYVSGVAAAYGGYLVDKYAYSDNPPETIGWSTTATDLGFVDGTGYQSP DIICHKDGKPGALSAEVAAGGEIELQWTEWPESHHGPVLNYLAPCGGDCSAVDKTSLEFFKIEAKGLIDG SSPPGHWATDDLISNNNSWTVTIPASVQEGNYVLRHEIIGLHSAGQKDGAQNYPQCINIKVTGGGAATPA GTAGEALYKDTDPGILFDIYSDLSGGYPIPGPEVFSA

#### >AtAA9F

MRHVQSASLLTALLSATKVAAHGHVSNIVINGVYYEGFDINSFPYMGENAPTVAAWTTPNTGNGPLAPDD YSSPDIICHQNATAGKGYVEVNAGDRISLQWTPWPESHHGPVVDYLARCEPNCASVDKTSLEFFKIDGVG IVDGSSVPGVWGDDQLIKNNNTWLVEIPKSIAPGYYVLRHELIALHSAGTEGGAQNYPSCFNLKVNGDGT DKPAGVVGTELYTPTGDGIIFNIYQTVSSYPVPGPTLYTGAATGVTQATSAITSTGTALTVGAAATTPAS GSGASSSAAPSSSAAATPSSRLSSLCCCVPL

#### >AtAA9G

MKLNLASLCFLASIAPLVSGHYVFSKLIVDGKTTKDFEYIRENSNGYQPTLASEIVSNDFRCNKGSMESA AKTKVYTVAPGAEMGFQLAYGASMKHPGPLQIYMSKAPGDVKAYDGSGDWFKVYQEGVCNDISGGLKDTD WCTWGKDTASFKIPENTPPGQYLVRVEHIGLHRGFSGNSEFYFTCAQIEVTGSGSGVPGPLVKIPGVYKP EDPNIHFNIYHPVPTSYDLPGPSVWSGGVSDSSSSISAPPVNNAAAASSVTPTTLVTLSKTSSTPAATSS AAPTSSAPSNGTIKKYYQCGGQGWTGSGSCEAGTSCREWNTWYFQCV 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 A. tamarii BLU37 transcriptome	Protein ID of the corresponding protein in the A. oryzae RIB40 genome	Protein ID in the A. tamarii CBS 117626 (Asptam1) genome	Closest characterized relatives <sup>a</sup>	Upregu- lation after 36 h <sup>h</sup>	Upregu- lation after 48 h <sup>h</sup>
AtAA9A	1567	254541	<i>Ls</i> AA9A [4]; 58% <sup>b</sup>	4.7	5.3
AtAA9B	2234	140265	<i>Ta</i> AA9A [5, 6]; 71% °	0.78	2.1
AtAA9C	4102	258171	_ <sup>d</sup>	1.4	1.1
AtAA9D	4194	312543	_ d,e	7.0	6.7
AtAA9E	4749	218852	<i>Ta</i> AA9A [5, 6]; 69% <sup>f</sup>	10.5	11.0
AtAA9F	5772	312044	<i>Tr</i> AA9A [7-9]; 58% <sup>g</sup>	1.9	-1.0
AtAA9G	11276	288991	_ c	5.5	6.1
- <sup>i</sup>	9997	273838 <sup>i</sup>		_ i	_ i
- <sup>i</sup>		303761 <sup>i</sup>		_ i	_ i

<sup>a</sup> Enzymes with known crystal structures for the catalytic domain; the sequence identity for the catalytic domain is indicated after the enzyme name.

<sup>b</sup> Of the reported LPMOs, *At*AA9A-N shares 80% sequence identity with An1602, for which C4-oxidizing activity on cellulose has been demonstrated [10].

<sup>c</sup> The closest experimentally characterized relatives of *At*AA9B 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.

<sup>d</sup> No sequence identity >50% with an experimentally characterized LPMO.

<sup>e</sup> Of the reported LPMOs, *At*AA9D-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].

<sup>f</sup> The closest experimentally characterized relatives of *At*AA9E are Aspfu5 [11] and *Af*AA9B [13, 14] with 72-73% sequence identity; the regioselectivity of these LPMOs has not yet been identified and experiments with xyloglucan were not reported.

<sup>g</sup> The closest experimentally characterized relatives of *At*AA9F 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.

<sup>h</sup> 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.

<sup>i</sup> 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.** Fulllength and AA9 domains only. Domains, as marked in S1 Table, were annotated by HMMER analysis using the dbCAN2 metaserver (<u>http://bcb.unl.edu/dbCAN2/</u>) and by structural alignment with experimentally characterized AA9 LPMOs using T-Coffee's Expresso tool (<u>http://tcoffee.crg.cat/apps/tcoffee/do:expresso</u>). The other listed characteristics were calculated using Expasy's ProtParam tool (<u>https://web.expasy.org/protparam/</u>).

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

<sup>a</sup> fragment only.

<sup>b</sup> without the signal peptide.

<sup>c</sup> 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 <sup>d</sup>	L3 loop	Cleavage type on xyloglucan	Other hemi- celluloses	Reference
NcAA9C	AA9–CBM1	C4	+/5-6 (+)/4	+	substitution- intolerant	reported; +	[15]
FgAA9A	AA9–[] <sup>a</sup>	C1/C4	-/3-6	-	substitution- tolerant	reported; -	[16]
GtAA9A-2	AA9–CBMx <sup>b</sup>	C1/C4	- / 5-6	-	substitution- tolerant	reported; -	[17]
TaAA9A	AA9 only	C1/C4	-/5	-	substitution- tolerant	reported; -	[6]
MYCTH_79765	AA9 only	C4	+ / 5-6	+	substitution- intolerant	reported; +	[18]
GtAA9B	AA9 only	C1/C4	-/5-6	-	substitution- tolerant	reported; -	[19]
LsAA9A	AA9 only	C4	+ / 4-6	+	substitution- intolerant	reported; +	[20]
CvAA9A	AA9 only	C4	+/4-6	+	substitution- intolerant	reported; +	[20]
РаАА9Н	AA9–CBM1	C1/ <u>C4</u>	+/5	+	substitution- tolerant	reported; +	[21, 22]
MYCTH_85556	AA9 only	<u>C1</u> /C4	-/1-5	-	?°	reported; +	[23]
MYCTH_100518	AA9 only	C4	-/1-5	+	?°	reported; +	[24]
NcAA9A	AA9–CBM1	C4	(+) / 5	+	substitution- intolerant; ((+)) <sup>d</sup>	reported; -	[25]
NcAA9D	AA9 only	C4	-/5	+	substitution- intolerant; (+) <sup>d</sup>	reported; -	[25]
McAA9A	AA9–[] <sup>a</sup>	C1/ <u>C4</u>	+/6	-	substitution- tolerant	reported; +	[26]
McAA9B	AA9 only	C1/ <u>C4</u>	-/6	-	substitution- tolerant	reported; +	[26]
McAA9F	AA9 only	C1/ <u>C4</u>	+/6	-	substitution- tolerant	reported; +	[26]
МсАА9Н	AA9 only	<u>C1</u> /C4	-/6	-	substitution- tolerant <sup>e</sup>	reported; +	[26]
GcAA9A	AA9 only	<u>C1</u> /C4	not reported	_	?°	not reported	[27]
GcAA9B	AA9–[] <sup>a</sup>	<u>C1</u> /C4	not reported		?c	not reported	[27]
An3046	AA9–[] <sup>a</sup>	?°	not reported		?c	not reported	[12]
TtAA9E	AA9 only	C1	not reported	-	?c	not reported	[28]
NcAA9M	AA9 only	C1/C4	not reported	-	substitution- tolerant	not reported	[29]
AtAA9A-N	AA9–CBM1	C4	+ / 5-6	+	substitution- intolerant	reported;-	this study
AtAA9B-N	$A\overline{A9}-[\dots]^a$	C1/C4	-/5-6	-	substitution- tolerant	reported;-	this study

<sup>a</sup> unidentified C-terminal extension

<sup>b</sup> unclassified carbohydrate-binding module (CBM)

<sup>c</sup> unclear; it cannot be determined from the available data

<sup>d</sup> activity levels are indicated as: –, inactive; ((+)), hardly active; (+) little active; +, active

<sup>e</sup> 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 amino the frequency acids was visualized using database: of WebLogo 3 (http://weblogo.threeplusone.com/create.cgi). Altogether 136 proteins were found (with an Evalue <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 Cterminal 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.

```
Tree scale: 0.01 🛏
```



**S2 Fig.** See legend on the next page.

**S2 Fig. Phylogenetic tree based on a multiple sequence alignment of the AA9 domains of** *A. tamarii* **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. *At*AA9s are indicated in red and bold; the closest characterized relatives of the *At*AA9s 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** *At***A49A-N and** *At***A49B-N.** Lane M, Benchmark Protein Ladder; Lane 1, purified *At*A49A-N; Lane 2, purified *At*A49B-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 <sup>#</sup> or <sup>##</sup>; hydration is denoted with <sup>\*</sup>. The peaks at 887.4 and 889.4 correspond to the Na<sup>+</sup>-salt of the aldonic acid, which is denoted as " $-H^+$  +Na<sup>+</sup>". 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.

		▼			L2 loop	•			
		10	20	30	40		50	60	
						.			
McAA9B	1	<b>H</b> GYVSKAILDGKEYT	<b>GYLPYEDPY</b>	HNPP-PERIFI	RKIAG	- <mark>N</mark> GPI-	EDLTS	IDLQC	52
NcAA9M	1	<mark>H</mark> GFVDNATIGGQFY-	QPYQDPY	MGSP-PDRIS	RKIPG	- <mark>N</mark> GPV-	EDVTSI	LAIQC	49
Gtaa9a-2-n	1	<mark>H</mark> GYVDQVTIGGQVYT	GYQPYQDPY	E-SPVPQRIE	RAIPG	-NGPV-	EDLTLI	LDIQC	52
GtAA9B	1	HGYVDTLNVGGTQYT	GYLPYNDPY	T-TPAPQRIE	RPIPG	-NGPV-	-TALTT	IDVQC	52
FgAA9A-N	1	HGHVESITVGGTEYE	GLNPGAAAF	E-NPRKELAAN	FATNTD	- <mark>N</mark> GFVI	PSAFGD	ADIIC	56
McAA9F	1	HGYVSSIQADGQTYP	GADPH	NPNPESPG	QAENTD	- GFVI	PSAFST	PAIAC	51
McAA9A-N	1	HGYVSGIVVDGAYHG	GYIVDKYPY	MPNP-PDVVG	STTATD	- GFVI	PDAFGDI	PDIIC	56
Ataa9B-N	1	HGYVSGAVVDGQYYS	GYDMS-YHY	MSDP-PKVIG	STDATD	-LGFVI	GSSYAD	ADIIC	55
TaAA9A	1	HGFVQNIVIDGKNYG	GYLVNQYPY	MSNP-PEVIA	STTATD	- <mark>L</mark> GFVI	GTGYQTI	PDIIC	56
CvAA9A	1	HTRMFSVWVNGVDQG	DGQN	vy11	RTPPN	-TDPI-	-KDLASI	PALAC	41
Ataa9a-N	1	HTTVHAVWINDVDQG	EGN-SQS	GYII	RSPPS	- <mark>N</mark> SPI-	-TDVTSI	KDMTC	43
LsAA9A	1	HTLVWGVWVNGVDQG	DGRN	IYII	RSPPN	- <mark>NNPV</mark> -	-KNLTSI	PDMTC	41
МсАА9Н	1	HYTFPSLIANGVVTG	EW	EYVI	RQTENHY:	SNAPV-	-TDVSSI	EAIRC	42
NcAA9A-N	1	HTIFVQLEADGTTYP	vs	YGII	RTPSY	-DGPI-	-TDVTSI	NDLAC	39
NcAA9D	1	HTIFSSLEVNGVNOG	LG	EGVI	RVPTY	- <mark>N</mark> GPI-	EDVTS	ASIAC	39
МҮСТН 79765	1	HAIFQKVSVNGADQG	SL	TGLI	RAPNN	- <mark>NNPV</mark> -	-QNVNS	DMIC	39
NcAA9C-N	1	HTIFQKVSVNGADQG	QL	KGII	RAPAN	- <mark>NNPV</mark> -	-TDVMS	SDIIC	39
PaAA9H-N	1	HSIFOKVSVNGVDOG	0L	KGVI	RAPYS	-NFPI-	-ENVNHI	PDFAC	39
0		т 						-	
consensus		· · · ·				•		• *	
		I.2 1000		<b>VV</b> T.	3 1000		,		
	-	70	80	90	100		10	120	
		, , , , , , , , , , , , , , , , , , , ,	00	90	TOO	_		120	
McAA9B			1	1 1		1	1 1	1	
NCAA9M	53	GGWONSGSAPAPI.TA	EPVTPGTVO					MARC	98
	53 50	GGWQNSGSAPAPLTA	EPVTPGTVQ	 KLQWT TLRWT			KGPIIT	MARC	98 90
GtAA9A-2-N	53 50 53	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA	EPVTPGTVQ SAAAGSTV-	 KLQWT TLRWT		TWEDS IWEDS TWESS	KGPIIT	MARC MARC	98 90 97
GtAA9A-2-N GtAA9B	53 50 53 53	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI-	 KLQWT TLRWT AFHWT		TWPDS IWPDS IWPSS IWPSSI	 KGPIIT VGPVIT VGPVIT	 YMARC YMARC YMGKV	98 90 97
GtAA9A-2-N GtAA9B FgAA9A-N	53 50 53 53 57	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI-	 KLQWT TLRWT AFHWT AFHWT TIKWD		TWIDI IWIDI IWIDI IWISI IWISI IWISI	KGPIIT VGPVIT VGPVIT VGPVIT	YMARC YMARC YMARC YMGKV YLGKV YLASC	98 90 97 97
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F	53 50 53 53 57 52	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVOAGSTI-	 KLQWT TLRWT AFHWT TIKWD KLTWN		TW DEF IW DEF TW SEF TW ESF	KGPIIT VGPVIT VGPVIT VGPVIT KGPVID	XMARC XMARC XMGKV XLGKV XLGKV XLASC	98 90 97 97 97 97
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N	53 50 53 53 57 52 52 57	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI-	 KLQWT TLRWT AFHWT TIKWD KLTWN ELOWN			KGPIIT VGPVIT VGPVIT KGPVIT KGPVID HGPVLD	MARC MARC MGKV LGKV LGKV LASC LASC LANC	98 90 97 97 97 92 97
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N	53 50 53 53 57 52 57 57 56	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI- EIAAGKOV-	 KLQWT TLRWT AFHWT TIKWD KLTWN ELQWN			KGPIIT VGPVIT VGPVIT VGPVIT KGPVID HGPVID	MARC MARC MGKV LGKV LGKV LASC LANC	98 90 97 97 97 92 97 96
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A	53 50 53 57 52 57 56 57	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI- EIAAGKQV- PVSPGGTV-	 KLQWT TLRWT AFHWT TIKWD KLTWN ELQWN ELQWT			KGPIIT VGPVIT VGPVIT VGPVIT KGPVID HGPVID KGPVIT	MARC MARC MGKV LGKV LASC LASC LANC LANC LANC	98 90 97 97 97 92 97 96 97
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A	53 50 53 57 52 57 56 57 42	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPOFV	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI- EIAAGKQV- PVSPGGTV- SASAGDKL-	 KLQWT TLRWT AFHWT TIKWD KLTWN ELQWT ELQWT TFEWYRVK		IWDD II IWDD II IWDS II IWDS II IWE II IWE II IWE II IN EI IN EI	KGPIIT VGPVIT VGPVIT VGPVIT KGPVID HGPVID KGPVIT KGPVIT	MARC MARC MGKV LGKV LASC LASC LANC LANC LANC MAAF	98 90 97 97 97 92 97 96 97 90
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N	53 50 53 57 52 57 56 57 42 44	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARPGAIHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI- EIAAGKQV- PVSPGGTV- SASAGDKL- EVKAGDKI-	 KLQWT TLRWT AFHWT TIKWD KLTWN ELQWT ELQWT TFEWYRVK TFEWHHDSRS	RGDD	IW D I	KGPIIT VGPVIT VGPVIT KGPVIT KGPVID HGPVID KGPVIT KGPVIT SGPITT	MARC MARC MARC MGKV LGKV LASC LASC LASC LASC LASC MAAF	98 90 97 97 97 92 97 96 97 90 90
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A	53 50 53 57 52 57 56 57 42 44 42	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARPGAIHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL NVDNRVVPKSV	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI- EIAAGKQV- PVSPGGTV- SASAGDKL- EVKAGDKI- PVNAGDTL-	 KLQWT TIRWT AFHWT KLTWN ELQWN ELQWT ELQWT TFEWYRVK TFEWHHD SRS- TFEWHHD SRS-	RGDD	IWDD I IWDD I IWSD I IWSS I IWSS I IWE I IWE I IWE I IN E I I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I SSI I I I SSI I I I I I I I I I I I I I I I I I I I	KGPIIT VGPVIT VGPVIT KGPVID HGPVLD KGPVID KGPVIT KGPVIT SGPITT NGPILV	YMARC     YMARC     YMARC     YMGKV     YLGKV     YLASC     YLASC     YLAPC     YLANC     YLANC     YLAPC     VIAAF     YMAPT     YIAPA	98 90 97 97 92 97 96 97 90 90 93 89
GtaA9A-2-N       GtaA9B       FgAA9A-N       McAA9F       McAA9A-N       AtaA9B-N       TaAA9A       CvAA9A       AtaA9A-N       LsAA9A       McAA9H	53 50 53 57 52 57 56 57 42 44 42 43	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARPGAIHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL NVDNRVAF	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI- EIAAGKQV- PVSPGGTV- SASAGDKL- EVKAGDKI- PVNAGDTL- SVAAGSTV-	 KLQWT AFHWT AFHWT KLTWN ELQWN ELQWT ELQWT TFEWYRVK TFEWHHD SRS- TFEWY HN T	RGDD	IWD I IWD I IWD I IWS I IWS I IWE I IWE I IWE I IN E I IN SI I I SI I	KGPIIT VGPVIT VGPVIT KGPVID HGPVLD KGPVID KGPVIT KGPVIT SGPITT NGPILV HGPIAV	MARC MARC MGKV LGKV LASC LASC LANC LANC LANC LANC MAAF MAAF MARV	98 90 97 97 97 92 97 96 97 96 97 90 93 89
GtaA9A-2-N       GtaA9B       FgAA9A-N       McAA9F       McAA9A-N       AtaA9B-N       TaAA9A       CvAA9A       AtaA9A-N       LsAA9A       McAA9H       NcAA9A-N	53 50 53 57 52 57 56 57 42 44 42 43	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL NVDNRVVPKSV YENPGRPAAKTL NGGPN-PTTPSDKTT	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- KVNAGATI- EIAAGKQV- PVSPGGTV- SASAGDKL- EVKAGDKI- PVNAGDTL- SVAAGSTV- TVNAGSTV-	 KLQWT AFHWT AFHWT KLTWN ELQWN ELQWT ELQWT TFEWYRVK TFEWHHDSRS TFEWHHDSRS TFEWHHDSRS-	RGDD RGDD 	IWD I IWD I IWD I IWD SI IWS I IWE I IWE I IWE I I I I SI I V M D SI I I SI I I A SI I I A SI	KGPIIT VGPVIT VGPVIT KGPVID KGPVID KGPVID KGPVID KGPVIT SGPITT NGPILV KGPILV KGPTLA	YMARC     YMARC     YMARC     YMGKV     YLGKV     YLASC     YLAPC     YLANC     YLANC     YLANC     YLANC     YLAPC     YLAPC     YIAPT     YIAPA     YMARV     YLKKY	98 90 97 97 92 97 96 97 90 93 89 83 92
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A McAA9H NcAA9A-N NcAA9D	53 50 53 57 52 57 56 57 42 44 42 43 40 40	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL NVDNRVVPKSV YENPGRPAAKTL NGGPN-PTTPSDKII NGSPN-TVASTSKVI	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- EIAAGKQV- PVSPGGTV- SASAGDKL- EVKAGDKI- PVNAGDTL- SVAAGSTV- TVNAGSTV-	 KLQWT AFHWT AFHWT KLTWN ELQWN ELQWT ELQWT TFEWYRVK TFEWHHD SRS- TFEWHHD SRS- TFEWHHD SRS- TFEWYHNT GFTVS KAIWRHTLTS-	RGDD ESDD RDDD RDDD 	IW D I IW D I IW S I IW S I IW E IW E IW E IW E I I I I I I I I I I I I I I I I I I I	KGPIIT VGPVIT VGPVIT KGPVID KGPVID KGPVID KGPVID KGPVIT SGPITT NGPILV KGPTLA KGPTLA	YMARC     YMARC     YMARC     YMGKV     YLGKV     YLASC     YLANC     YLKKV     YLKKV	98 90 97 97 92 97 96 97 96 97 90 93 89 83 92
GtaA9A-2-N       GtaA9B       FgAA9A-N       McAA9F       McAA9A-N       AtaA9B-N       TaAA9A       CvAA9A       AtaA9A-N       LsAA9A       McAA9A-N       LsAA9A       NcAA9A-N       NcAA9A       McAA9A       McAA9A       McAA9A       McAA9A       McAA9A       McAA9A       NcAA9A       McAA9A       NcAA9A       MyCTH       79765	53 50 53 57 52 57 56 57 42 44 42 42 43 40 40	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL NVDNRVVPKSV YENPGRPAAKTL NGGPN-PTTPSDKII NGSPN-TVASTSKVI GOSGSTSNTIT	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- KVNAGATI- EIAAGKQV- PVSPGGTV- SASAGDKL- EVKAGDKI- PVNAGDTL- SVAAGSTV- TVNAGSTV- TVQAGTNV- EVKAGDRI-	 KLQWT AFHWT AFHWT KLTWN ELQWN ELQWT ELQWT ELQWT TFEWYRVK TFEWHHD SRS- TFEWHHD SRS- TFEWYHNT GFTVS KAIWRHTLTS- TAIWRYMLST- GAWYOHU IGG2	RGDD 	IW   D     IW   D     IW   D     IW   S     IW   S     IW   E     II   D     SI   -     VMDASI     VMDASI     PIAK	KGPIIT VGPVIT VGPVIT KGPVID KGPVID KGPVID KGPVID KGPVIT SGPITT NGPILV KGPILA KGPTLA KGPVA	YMARC     YMARC     YMARC     YMARC     YMARC     YLGKV     YLANC     YLANC <t< th=""><th>98 90 97 97 92 97 96 97 90 93 83 89 83 92 95 94</th></t<>	98 90 97 97 92 97 96 97 90 93 83 89 83 92 95 94
GtaA9A-2-N       GtaA9B       FgAA9A-N       McAA9F       McAA9A-N       AtaA9B-N       TaAA9A       CvAA9A       AtaA9A-N       LsAA9A       McAA9H       NcAA9A-N       LsAA9A       McAA9A-N       LsAA9A       McAA9A-N       LsAA9A       McAA9H       NcAA9D       MYCTH_79765       NcAA9C-N	53 50 53 57 52 57 52 57 56 57 42 44 42 43 40 40 40	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL NVDNRVVPKSV YENPGRPAAKTL NGGPN-PTTPSDKII NGSPN-TVASTSKVI GQSGSTSNTII NAVTMKDSNVI	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- KVNAGATI- EIAAGKQV- PVSPGGTV- SASAGDKL- EVKAGDKI- PVNAGDTL- SVAAGSTV- TVNAGSTV- TVQAGTNV- EVKAGDRI- TVPAGAKV-	I     I       KLQWT     I       TIRWT     I       AFHWT     I       GELQWT     I       FEWYRVK     I       TFEWYRVK     I       GFTVS     I       GFTVS     I       GAWYQHY     IGG3       GHFWGHE     IGG3	RGDD ESDD RDDD RDDD RDDD 	IW D F IW D F IW S I IW S I IW E IW E IW E IW E IW E I I I I S I I I S I I I S I I I S I I I S I I I S I I I S I I I S I I I S I I I S I I I S I I I I S I	KGPUIT KGPVIT KGPVIT KGPVID KGPVID KGPVID KGPVID KGPVIT KGPVIT KGPILV KGPILV KGPTLA KGPTLA KGPVMA	YMARC     YMARC     YMARC     YMARC     YMARC     YLGKV     YLANC     YLANC <t< th=""><th>98 90 97 97 97 92 97 96 97 90 93 89 83 92 95 94 94</th></t<>	98 90 97 97 97 92 97 96 97 90 93 89 83 92 95 94 94
GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A McAA9H NcAA9A-N NcAA9D MYCTH_79765 NcAA9C-N PaAA9H-N	53 50 53 53 57 52 57 56 57 42 44 42 43 40 40 40 40 40	GGWQNSGSAPAPLTA NADSAPAKLHA NGSGGSGTKPAALIA NGENGGGSSPAPLVA HRGAENAVKSA HKNARAPPAHA HRDGAPGAIHA HKNAKNGAISA HRGAKPGALTA NVKGGEPVPQFV NVNNKATAKTL NVDNRVVPKSV YENPGRPAAKTL NGGPN-PTTPSDKII NGSPN-TVASTSKVI GQSGSTSNTII NAVTMKDSNVL	EPVTPGTVQ SAAAGSTV- SAAAGDEI- TIAAGGKI- KVKAGEKI- TVQAGSTI- EVAAGKU- PVSPGGTV- SASAGDKL- EVKAGDKI- PVNAGDTL- SVAAGSTV- TVNAGSTV- TVQAGTNV- EVKAGDRI- TVPAGAKV-	Image: Non-State State		IW   D     IW   D     IW   D     IW   S     IW   S     IW   S     IW   S     IW   E     IW   E     IW   E     IW   E     II   P     II   S     II <th>KGPVIT KGPVIT KGPVIT KGPVID KGPVID KGPVID KGPVIT KGPVIT KGPVIT KGPIL KGPIL KGPIL KGPTLA KGPTLA KGPIM</th> <th>YMARC     YMARC     YMARC     YMARC     YMARC     YLARC     YLARC     YLANC     YLANC     YLANC     YLAPC     VIAAF     YMAPT     YIAPA     YMARV     YLKKV     YLAKV     YLAKV     YLAKV</th> <th>98 90 97 97 97 92 97 96 97 90 93 89 83 92 95 94 94</th>	KGPVIT KGPVIT KGPVIT KGPVID KGPVID KGPVID KGPVIT KGPVIT KGPVIT KGPIL KGPIL KGPIL KGPTLA KGPTLA KGPIM	YMARC     YMARC     YMARC     YMARC     YMARC     YLARC     YLARC     YLANC     YLANC     YLANC     YLAPC     VIAAF     YMAPT     YIAPA     YMARV     YLKKV     YLAKV     YLAKV     YLAKV	98 90 97 97 97 92 97 96 97 90 93 89 83 92 95 94 94

Consensus

\* \*\* ::

S5 Fig. (continued)

. :

					LS 1	оор				
		130	140	150	)	160	1	70	180	)
			.	.     .				.		
McAA9B	99	P-GDCSEYEP-	GTDAVWFKIA	EDGKH	-DDG-	SWASDPLI	N-DV-I	PYEFTIE	PEGLA	147
NcAA9M	91	P-D-TGCQDWTPS.	ASDKVWFKIK	EGGREG	-TSN-V	WAATPLM	ITAP-AI	NYEYAII	SCLK	143
GtAA9A-2-N	98	P-SNTDITSYSPT	GSDVIWFKID	EAGYE	NG-1	<b>XWAATDIM</b>	ISAQNS	TWTVTIE	KALA	150
Gtaa9b	98	P-SSTDVTKYSPT	GSDVIWFKID	EGGYS	NG-1	KWAATDVI	SAQNS	TWTVTIE	SSLA	150
FgAA9A-N	98	G-SAG-CAKVDK-	-TSLKFFKIA	EAGMTS	GG-1	<b>KFASDDLI</b>	AAG-N	TWEVTVE	TSIK	147
McAA9F	93	N-GDCSSASA-	-GSLNFVKIA	EKGLISGS	-NPG-1	FWAADELI	QNG-N	SWEVTIE	PANLA	144
McAA9A-N	98	N-GDCSSVDK-	-TSLKFFKIS	EAGLNDGS	NAPG-9	WASDDLI	ANN-N	SWTVTIE	<b>KSIA</b>	150
Ataa9B-N	97	N-GDCATVDK-	-TQLEFFKID	EKGLISG-	-SDN-	<b>FWASDNLI</b>	SSN-N	SWTVTIE	SSIA	147
TaAA9A	98	N-GDCSTVDK-	-TQLEFFKIA	ESGLINDD	NPPG-:	IWASDNLI	AAN-N	SWTVTIE	AITTS	150
CvAA9A	91	TSPTMDG-	-TGPVWSKIH	EEGYDAST	K-8	SWAVDKLI	ANK-G	MWDFTLE	SQLK	138
Ataa9a-n	94	EKGT-	-AGNGWVKIA	EDGYTD	)G-!	<b>FWAVETLI</b>	KNR-GI	KHSVTVE	PD-VA	135
LsAA9A	90	ASNG-	-QGNVWVKLF	EDAYNVTN	IS-!	<b>WAVDRLI</b>	TAH-G	QHSVVVE	PH-VA	133
МсАА9Н	84	PDGQ-TADSWDG-	-SGQVWFKIF	EQGPQIDP	-SGL-	TWPSDGL-	S	QVQVTIE	SSLP	133
NcAA9A-N	93	D-DALTDTG-	-IGGGWFKIQ	EDGYNN	G-9	QWGTSTVI	TNG-G	FQYIDIE	ACIP	139
NcAA9D	96	D-NAATASG-	-VGNGWFKIQ	QDGMDS	SG-1	VWGTERVI	NGK-GI	RHSIKI	PECIA	143
MYCTH_79765	95	D-NAATASK-	-TGLKWFKIW	EDTFNP	STK	<b>FWGVDNL</b> I	NNN-G	WVYFNLE	QCIA	143
NcAA9C-N	95	D-NAATTGT-	-SGLKWFKVA	EAGLSN	G-1	KWAVDDLI	ANN-G	WSYFDME	TCIA	141
PaAA9H-N	96	N-NAANAGT-	-SGLQWFKVA	EQGLNN	G-1	WAVDNMI	SNG-G	WHYFDME	SCVA	142
Consensus			• *•					• •	· ·	
								•	•	
		,	••	•	,		L	C loop		
		190	200	<b>2</b> 10		220	<b>L</b> (	<mark>C loop</mark> 30	24(	)
		190	200	210	, , , , ,   ,	220	23	<mark>C loop</mark> 30	240	)
Мсаа9в	148	190 	200 .	₹ 210 .  .	PSCFO	220	23	<mark>C loop</mark> 30   . <b>T-NLV</b>	24( 	) <b>196</b>
McAA9B NcAA9M	148 144	190    PGNYIVRHELWAL PGYYLVRHEIIAL	200 .   AWT	210 .  . -YPGAQVY -YPGAOFY	PSCFQ	220   VKVVGDG- LOVTGSG-	22 22 1 TQ-QP TK-TP	C loop 30   . T-NLV S-S-GLV	24(   /AFPG /SFPG	) 196 193
McAA9B NcAA9M GtAA9A-2-N	148 144 151	190    PGNYIVRHELWAL PGYYLVRHEIIAL PGOYIVRHEIIAL	200 .   AWT AYS	210 .  . -YPGAQVY -YPGAQFY -YPGAOFY	PSCFQ	220   VKVVGDG- LQVTGSG- VOVTGPG-	23   TQ-QP TK-TP	C loop 30   . T-NLV S-S-GLV T-SOALV	24(   /AFPG /SFPG	) 196 193 201
McAA9B NcAA9M GtAA9A-2-N GtAA9B	148 144 151 151	190    PGNYIVRHELWAL PGYYLVRHEIIAL PGOYIVRHEIIAL PGOYIVRHEIIAL	200 .   AWT AYS AET AOT	210 .  . -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQFY	PSCFQ PGCHQI PDCFQ PDCFQ	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG-	L( 2: -TQ-QP' -TK-TP! -TE-TP! -NK-TP!	C loop 30   . S-S-GLV T-SQALV S-GSYLV	24(   /AFPG /SFPG /SFPG	196 193 201 201
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N	148 144 151 151 148	190   PGNYIVRHELWAL PGYYLVRHEIIAL PGQYIVRHEIIAL AGNYVLRHEIIAL	200 .   AWT AYS AET AQT AGO	210 .  . -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -ENGAONY	PSCFQ PGCHQI PDCFQ PDCFQ POCFQ	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG-	L( 2: -TQ-QP! -TK-TP! -TE-TP! -NK-TP! -TA-EP!	C loop 30   . T-NLV S-S-GLV T-SQALV S-GSYLV AGV	24( //AFPG /SFPG /SFPG /SFPG /AGTS	196 193 201 201 195
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F	148 144 151 151 148 145	190   PGNYIVRHELWAL PGYYLVRHEIIAL PGQYIVRHEIIAL AGNYVLRHEIIAL PGKYVLRHEIIAL	200 .   AWT AYS AET AQT AGQ AGQ	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -ENGAQNY -PNGAQAY	PSCFQ PGCHQI PDCFQ PDCFQ PQCFNI PQCINI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG-	L( 2: -TQ-QP' -TK-TP' -TK-TP' -NK-TP' -TA-EP' -SA-TP'	C loop 30   . T-NLV S-S-GLV T-SQALV S-GSYLV AGV SGQ	24( /AFPG /SFPG /SFPG /SFPG /AGTS /PATS	196 193 201 201 195 192
MCAA9B NCAA9M GtAA9A-2-N GtAA9B FgAA9A-N MCAA9F MCAA9A-N	148 144 151 151 148 145 151	190   PGNYIVRHELWAL PGYYLVRHEIIAL PGQYIVRHEIIAL AGNYVLRHEIIAL PGKYVLRHEIIAL PGNYVLRHEIIAL	200 .   AWT AYS AET AQT AGQ AGN AGN	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -ENGAQNY -PNGAQAY -QNGAQNY	PSCFQ PGCHQI PDCFQ PDCFQ PDCFQ PQCFNI PQCFNI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG-	L( 2: -TQ-QP -TK-TP -TK-TP -TE-TP -NK-TP -SA-TP -SA-TP	C       loop         30	24( /AFPG /SFPG /SFPG /SFPG /AGTS /PATS /LGTE	196 193 201 201 195 192 198
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N	148 144 151 151 148 145 151 148	190   PGNYIVRHELWAL PGYYLVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGKYVLRHEIIAL PGNYVLRHEIIAL AGNYVMRHEIIAL	200 .   AWT AYS AET AQT AQQ AQQ AQN AQN AQN	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -ENGAQNY -PNGAQAY -QNGAQNY -KDGAQNY	PSCFQ PGCHQI PDCFQ PDCFQ PQCFNI PQCFNI PQCFNI PQCLNI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG-	L( 2: 	C       loop         30	240 //AFPG //SFPG //SFPG //AGTS //AGTS //LGTE //LGTE	196 193 201 201 195 192 198 195
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A	148 144 151 151 148 145 151 148 151	190   PGNYIVRHELWAL PGYYLVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGKYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL	200 .   AWT AYS AZT AQT AGQ AGN AGN AQN	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -KDGAQNY -QDGAQNY	PSCFQ PGCHQI PDCFQ PDCFQ PDCFQ PQCFNI PQCFNI PQCLNI PQCLNI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG- LQVTGGG-	L( 2: 	C       loop         30	240 //AFPG //SFPG //SFPG //SFPG //AGTS //AGTS //LGTE //LGTA //LGTA	196 193 201 201 195 192 198 195 198
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A	148 144 151 151 148 145 151 148 151 139	190   PGNYIVRHELWAL PGYYLVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGKYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL	200 .     AWT AYS AZT AQT AQQ AQQ AQN AQN HE SDATFDKN	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY PKRGAQFY	PSCFQ PGCHQI PDCFQ PDCFQ PQCFNI PQCFNI PQCFNI PQCLNI PQCLNI PQCLNI PSCVQ	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG- LQVTGGG- VDVKGVG-	L( 2: 	C loop 30   . T-NLV S-S-GLV T-SQALV S-GSYLV AGV EGV AGT AGT PDQAE	24( 7AFPG 7SFPG 7SFPG 7SFPG 7AGTS 2PATS 7LGTE TLGTA TLGTA	196 193 201 201 195 192 198 195 198 194
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A	148 144 151 151 148 145 151 148 151 139 136	190   PGNYIVRHEIWAL PGYYLVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGKYMLRQEIVAH AGEYLFRPEIIAL	200 .     AWT AYS AZT AQT AQQ	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY -QDGAQNY -EGGAQFY -EGGAQFY	PSCFQ PGCFQ PDCFQ PDCFQ PQCFN PQCFN PQCFN PQCINI PQCINI PSCVQ MECVQ	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG- LQVTGGG- VDVKGVG- VVVSSG-	L( 2: 	C loop 30   . T-NLV S-S-GLV T-SQALV S-GSYLV AGV EGV AGT PDQAE P-EGV	24( 7AFPG 7SFPG 7SFPG 7SFPG 7AGTS 2PATS 7LGTE TLGTA TLGTA TLGTA 7SIPG	196 193 201 201 195 192 198 195 198 194 184
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A	148 144 151 148 145 151 148 151 139 136 134	190   PGNYIVRHELWAL PGYYLVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGXYMLRQEIVAH AGEYLFRPEIIAL PGDYLFRAEIIAL	200 .     AWT AYS AZT AQT AQT AQQ	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY PKRGAQFY -EGGAQFY PIRGAOFY	PSCFQ PGCFQ PDCFQ PDCFQ PQCFN PQCFN PQCIN PQCIN PQCIN PSCVQ MECVQ ISCAO	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG- VDVKGVG- VVVTSSG- ITINSSD-	L( 2: 	C loop 30   . T-N-LV S-S-GLV S-GSYLV AGV EGV EGV PDQ-AE P-EGV LPAGV	24( AFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7DFNK 7SIPG 7FPG 7FPG	196 193 201 201 195 192 198 195 198 194 184 189
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A McAA9H	148 144 151 148 145 151 148 151 139 136 134	190   PGNYIVRHEIWAL PGYYLVRHEITAL PGQYIVRHEITAL PGQYIVRHEITAL AGNYVLRHEITAL PGNYVLRHEITAL PGNYVLRHEITAL PGNYVLRHEITAL PGNYVLRHEITAL PGYYLFREITAL PGDYLFRAEITAL SGDYLLRVEQIGL	200 	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY -QDGAQNY -EGGAQFY -EGGAQFY -EGGAQFY -VNGAQFY	PSCFQ PGCHQI PDCFQ PDCFQ PQCFNI PQCFNI PQCLNI PQCLNI PQCLNI PSCVQ MECVQ ISCAQ LSCAQI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG- VDVKGVG- VTGGG- VDVKGVG- VTSSG- ITINSSD- LTVTGGG-	L( 2) TQ-QP TK-TP TK-TP SA-TP SD-NP	C loop 30   . S-S-GLV S-S-GLV S-GSYLV AGV EGV EGV AGV PDQAE PDQ-AE P-EGV LPAGV G-PLV	24(   /AFPG /SFPG /SFPG /AGTS /LGTE /LGTA	196 193 201 201 195 192 198 195 198 194 184 189 182
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A McAA9H NcAA9A-N	148 144 151 148 145 151 148 151 139 136 134 134	190   PGNYIVRHEIWAL PGYYLVRHEITAL PGQYIVRHEITAL PGQYIVRHEITAL PGQYIVRHEITAL PGNYVLRHEITAL PGNYVLRHEITAL PGNYVLRHEITAL PGNYVLRHEITAL PGYYLRAEITAL SGDYLLRVEQIGL SGOYLLRAEMIAL	200 .     AWT AYS AQT AQT AQQ AQQ AQQ AQQ AQQ AQQ AQQ AQQ HESDATFDKN HEGNR HEADSLYSQN ASS	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY -QDGAQNY -EGGAQFY -EGGAQFY PIRGAQFY -VNGAQFY -TAGAQLY	PSCFQ PGCHQI PDCFQ PDCFQ PQCFNI PQCFNI PQCINI PQCINI PSCVQ MECVQ ISCAQI LSCAQI MECAQI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG- VDVKGVG- VTSSG- ITINSSD- LTVTGGG- INIVGGTG	L( 2) TQ-QP TK-TP TK-TP SA-TP SA-TP SD-NP	C       loop         30	24(   /AFPG /SFPG /SFPG /AGTS PATS /LGTE /LGTA	196 193 201 201 195 192 198 195 198 194 184 184 189 182 190
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A McAA9H NcAA9A-N NcAA9D	148 144 151 151 148 145 151 148 151 139 136 134 134 134 140 144	190   PGNYIVRHEIWAL PGYYLVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGYYLRAEIIAL SGDYLLRVEQIGL SGQYLLRAEMIAL PGQYLLRAEMIAL	200 .     AWT AYS AQT AQN HESDATFDKN HEGNR HEADSLYSQN ASS HAASS HAASS	210 -YPGAQYY -YPGAQFY -YPGAQFY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY -QDGAQNY -EGGAQFY -EGGAQFY -EGGAQFY -TAGAQLY -YPGAQFY	PSCFQ PGCHQI PDCFQ PDCFQ PDCFQ PQCFNI PQCINI PQCINI PQCINI PSCVQ MECVQ ISCAQI ISCAQI MECAQI MECAQI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEVTGGG- VDVKGVG- VTVGGG- UVVKGVG- ITINSSD- LTVTGGG- INIVGGTC	L( 2) TC-DP TK-TP TK-TP NK-TP SA-TP SA-TP SD-NP SD-NP SD-NP SD-NP SD-NP SD-NP SD-NP SD-NP SD-NP SD-NP SD-NP SC-NP	C       loop         30           .         F-N-LW         S-S-GLW         S-S-GLW         S-GSYLW         S-T-TY         P-S-TW	24( 7AFPG 7SFPG 7SFPG 7SFPG 7SFPG 7AGTS 2PATS 7LGTE 7LGTA 7LGTA 7LGTA 7LGTA 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG	196 193 201 195 192 198 195 198 194 184 184 189 182 190 192
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LSAA9A McAA9H NcAA9A-N NcAA9D MYCTH 79765	148 144 151 151 148 145 151 148 151 139 136 134 134 134 140 144	190   PGNYIVRHEIWAL PGYYLVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGQYIVRHEIIAL PGKYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGCYLFREIIAL PGDYLFRAEIIAL SGDYLLRVEQIGL SGQYLLRAEMIAL PGQYLLRAEMIAL DGNYLLRVEVLAL	200 .     AWT AYS AET AQT	210 -YPGAQYY -YPGAQFY -YPGAQFY -YPGAQSY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY -QDGAQNY PKRGAQFY -EGGAQFY PIRGAQFY -TAGAQLY -YPGAQFY -QGQAOFY	PSCFQ PGCHQI PDCFQ PDCFQ PDCFQ PQCFNI PQCINI PQCINI PQCINI PSCVQ MECVQ ISCAQI ISCAQI MECAQI QSCAQI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEVTGGG- VDVKGVG- VTGGG- VDVKGVG- VKVTSSG- ITINSSD- LTVTGGG- INIVGGT- INIVGGT-	L( 2) TC-TP: TK-TP: TK-TP: TA-EP/ SA-TP: SD-NP! SD-NP! SD-NP! GD-AV! SK-TL: OS-TP: NG-NP( STALP: GA-KT) SF-TP!	C       loop         30           .         F-N-LW         S-S-GLW         S-S-GLW         S-GSYLW         S-GSYLW         S-GSYLW         S-GSYLW         S-GSYLW         S-GSYLW         S-GSYLW         S-GSYLW         P-S-GW         PDQ-AB         PDQ-AB         P-E-GW         S-T-TY         P-S-TW         A-S-TW	24( 7AFPG 7SFPG 7SFPG 7SFPG 7SFPG 7AGTS 2PATS 7LGTE 7LGTA 7LGTA 7LGTA 7LGTA 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG	196 193 201 195 192 198 195 198 194 184 184 189 182 190 192 192
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9A-N LsAA9A McAA9A-N LsAA9A McAA9H NcAA9A-N NcAA9D MYCTH_79765 NcAA9C-N	148 144 151 151 148 145 151 148 151 139 136 134 134 134 140 144	190   PGNYIVRHEIWAL PGYYLVRHEIIAL PGQYIVRHEIIAL AGNYVLRHEIIAL PGKYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL SGDYLLRAEIIAL SGDYLLRVEQIGL SGQYLLRAEMIAL PGQYLLRAEMIAL PGQYLLRAEMIAL	200 .     AWT AYS AET AQT AQT AQT AQN AQN AQN AQN HESDATFDKN HEGNR HEADSLYSQN ASS HAASS HAASS HAASS HAASS HAASS	210 -YPGAQYY -YPGAQFY -YPGAQFY -YPGAQSY -YPGAQSY -PNGAQAY -PNGAQAY -QNGAQNY -QDGAQNY -QDGAQNY PKRGAQFY -EGGAQFY -EGGAQFY -TAGAQLY -YPGAQFY -QQQAQFY -QAGAQFY	PSCFQ PGCHQI PDCFQ PDCFQ PDCFQ PQCFNI PQCINI PQCINI PQCINI PQCINI PSCVQ MECVQ ISCAQI ISCAQI MECAQI QSCAQI IGCAQI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- LEITSNG- FKVTGGG- VDVKGVG- VDVKGVG- INIVGGG- INIVGGT- INIVGGG- INVSGGG-	L( 2) TT - TP TK - TP TK - TP TA - EP SA - TP SA - TP SD - KP SD - KP ST - TP ST - TP	C       loop         30	24(   /AFPG /SFPG /SFPG /ZSFPG /LGTE /LGTA /LGTA /LGTA /DFNK /SIPG /SFPG /SFPG /SFPG /SFPG /SFPG	196 193 201 195 192 198 195 198 194 184 184 189 182 190 192 192
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9F McAA9A-N AtAA9B-N TaAA9A CvAA9A AtAA9A-N LsAA9A McAA9H NcAA9A-N NcAA9D MYCTH_79765 NcAA9C-N PaAA9H-N	148 144 151 151 148 145 151 148 151 139 136 134 134 140 144 144 142 143	190   PGNYIVRHEIWAL PGYYLVRHEIIAL PGQYIVRHEIIAL AGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL PGNYVLRHEIIAL SGDYLLRAEIIAL SGDYLLRVEQIGL SGQYLLRAEMIAL PGQYLRAEMIAL PGQYLRAEMIAL PGQYLRAEMIAL PGQYLRAEMIAL	200 .   AWT AYS AZT AZT AZT AZT AZT AZT AZT AZT AZT AZT AZT AZT HEADSLYSQN ASS HAASS HSAYS HNAGS ASV	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQFY -PNGAQAY -PNGAQAY -QNGAQNY -QNGAQNY -QDGAQNY -QGGAQFY PIRGAQFY -TAGAQLY -YPGAQFY -QGQAQFY -QGQAQFY -QGQAQFY -RGAAQFY	PSCFQ PGCHQI PDCFQ PDCFQ PQCFNI PQCFNI PQCINI PQCINI PQCINI PQCINI PSCVQ MECVQ ISCAQI MECAQI MECAQI IGCAQI MECAQI	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVESDG- LEVTGGG- KVTGGG- VDVKGVG- VKVTSSG- INIVGGTG LNVVGGT- INVSGGG- INVTGGG- INVTGGG- INVTGGG-	L( 2: 	C       loop         30	24( AFPG 7SFPG 7SFPG 7SFPG 7AGTS PATS 7LGTE TLGTA TLGTA TLGTA TLGTA TLGTA TSFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG	196 193 201 201 195 192 198 195 198 194 184 189 182 190 192 192 190 191
McAA9B NcAA9M GtAA9A-2-N GtAA9B FgAA9A-N McAA9F McAA9A-N AtAA9B-N TaAA9A CVAA9A AtAA9A-N LSAA9A McAA9H NcAA9A-N NcAA9A-N NcAA9D MYCTH_79765 NcAA9C-N PaAA9H-N Consensus	148 144 151 151 148 145 151 148 151 139 136 134 134 140 144 144 142 143	190 	200 	210 -YPGAQVY -YPGAQFY -YPGAQFY -YPGAQFY -PNGAQAY -PNGAQAY -QNGAQNY -QNGAQNY -QGAQNY -EGGAQFY -EGGAQFY -TAGAQLY -YPGAQFY -YPGAQFY -QGQAQFY -QAGAQFY -RGAAQFY -K* *	PSCFQ PGCHQI PDCFQ PDCFQ PQCFNI PQCFNI PQCINI PQCINI PQCINI PQCINI PSCVQ MECVQ ISCAQ ISCAQ ISCAQ ISCAQ IGCAQ IGCAQ	220 VKVVGDG- LQVTGSG- VQVTGPG- IRVTGSG- LEVESDG- LEVTGGG- KVTGGG- VDVKGVG- VVVTSSG- ITINSD- LTVTGGG- INIVGGTG INIVGGG- INVVGGG- INVTGGG-	L( 2: 	C       loop         30         .         30         .         F-NLV       S-S-GLV         S-S-GLV       S-GSYLV         AGV       S-GSYLV         SGV       S-GSYLV         SGV       S-GSYLV         SGV       S-TTV         S-NTV       S-NFV	24( AFPG 7SFPG 7SFPG 7SFPG 7AGTS 2PATS 7LGTE 7LGTA 7LGTA 7LGTA 7LGTA 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG 7SFPG	196 193 201 201 195 192 198 195 198 194 184 184 184 189 192 190 191

S5 Fig. (continued)

				LC loop					
		250		260		270	280	290	300
		.	••	.	.				
McAA9B	197	EYTPDTPGVVY	DI	QN	NEE	-YPIPG	PPVWTPA		227
NcAA9M	194	AYKSTDPGVTY	DA	QA	A-1	-YTIPG	PAVFTC		222
GtAA9A-2-N	202	GYTPTTPGITF	V	SG	-SI-I	SYPIPG	PPVWTS		232
GtAA9B	202	AYTATTPGIAF	v	TN	F-1	SYPIPG	PAVWTGN		232
FgAA9A-N	196	LYTASEKGIVF	DL	(NN	PT-	SYPIPG	PKMNIA		225
McAA9F	193	FYSPNDPGILF	NL	QS	FD-	SYPIPG	PAVWSG		222
McAA9A-N	199	LYKADDPGILF	NI	QP	MD-	SYPIPG	PALYTG		228
Ataa9B-N	196	LYKDTDPGILV	NI	QT	LS-	SYTIPG	PALYSG		225
TaAA9A	199	LYHDTDPGILI	NI	QK	LS-	SYIIPG	PPLYTG		228
CvAA9A	195	GYKYSDPGIAF	DM.	TD	FD-	SYPIPG	PPVWDAQDEG	CCFIDGVDTTS	<b>VKEVV</b> 244
Ataa9a-n	185	AYTATDKGILF	NI	DS	FD-	SYPIPG	PAVWDG		214
LsAA9A	190	AYTDSTPGIQF	NI	TTP	AT-	SYVAPP	PSVWSGALGG	SIAQVGDAS	<b>SL</b> 234
МсАА9Н	183	AYSPTDPGLLI	NI	WPI	PT-	SYELPG	PPVWRG		213
NcAA9A-N	191	IYKATDPGLLV	NI	(SM	SPSS-	TYTIPG	PAKFTC		222
NcAA9D	193	AYSGSDPGVKI	SI	WPP	<b>VT</b> -	SYTVPG	PSVFTC		223
MYCTH_79765	193	AYSASDPGILI	NI	GATGQPDNN	GQ-	PYTAPG	PAPISC		229
NcAA9C-N	191	AYSASDPGILI	NI	GGSGKTDNG	GK-	PYQIPG	PALFTC		227
PaAA9H-N	192	AYTADHPGILV	SI	DLQGRPTNG	GR-	PYTIPG	PAPLTC		228

			310
			•••
McAA9B			
NcAA9M			-
GtAA9A-2-N			
GtAA9B			
FgAA9A-N			
McAA9F			-
McAA9A-N			-
AtAA9B-N			
TaAA9A			
CvAA9A	245	KQIICVI	<b>K</b> 252
Ataa9a-n			-
LsAA9A	235		<b>E</b> 235
МсАА9Н			
NcAA9A-N			
NcAA9D			
MYCTH_79765			-
NcAA9C-N			-
PaAA9H-N			_

\*:

Consensus

Consensus

**S5 Fig. Multiple sequence alignment of the catalytic domains of AA9 LPMOs for which activity on xyloglucan has been demonstrated.** The alignment was generated using T-Coffee's Expresso tool (<u>http://tcoffee.crg.cat/apps/tcoffee/do:expresso</u>) and reflects a structure-based alignment of crystal structures and three-dimensional models. Substitution-intolerant xyloglucan-active AA9s appear in red, substitution-tolerant xyloglucan-active AA9s appear in blue. Additional xyloglucan-active LPMOs with unknown cleavage specificity, which are not shown in this alignment, include MYCTH\_85556 [23] and MYCTH\_100518 [24] from *Myceliophthora thermophila*, *Tt*AA9E from *Thielavia terrestris* [28] and *Gc*AA9A and *Gc*AA9B from *Geotrichum candidum* [27]. The sequences of LPMOs that are not able to cleave soluble cello-oligosaccharides appear on a grey background. (Note that for *Nc*AA9M, activity on cello-oligosaccharides has not

been reported; its product profile on PASC (short cello-oligosaccharides with DP 2-4) is similar to that of *Nc*AA9C, 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 *Nc*AA9C [31] and *Ls*AA9A [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) *Nc*AA9C-N (PDB: 4D7U; gold), *Ls*AA9A (PDB: 5ACI; light purple) and *Cv*AA9A (PDB: 5NLT; magenta) with (substitution-tolerant) *Ta*AA9A (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 *Nc*AA9C.







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<sub>6</sub>) 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 *Pa*AA9H-N in panel **B**, and in MYCTH\_100518 in panel **C**. The LPMOs in panel **B** (except *Pa*AA9H-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 *Mc*AA9H and *Pa*AA9H). 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 tamarii* 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 monooxygenases. 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, Sørlie M, Jensen MS, et al. Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase 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 monooxygenase. Enzyme Microb Technol. 2017;98:58-66.

9. Hansson H, Karkehabadi S, Mikkelsen N, Douglas NR, Kim S, Lam A, et al. Highresolution structure of a lytic polysaccharide monooxygenase 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 monooxygenases 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 monooxygenases 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 monooxygenase 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 monooxygenase 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. *Fg*LPMO9A from *Fusarium graminearum* cleaves xyloglucan independently of the backbone substitution pattern. FEBS Lett. 2016;590(19):3346-56.

17. Kojima Y, Varnai 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. pHdependent 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 *Pa*LPMO9H 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.