

Arabinogalactan-proteins of *Zostera marina* L. contain unique glycan structures and provide insight into adaption processes to saline environments

Lukas Pfeifer¹, Thomas Shafee², Kim L. Johnson^{2,3}, Antony Bacic^{2,3} and Birgit Classen^{1*}

Affiliations:

¹ Pharmaceutical Institute, Department of Pharmaceutical Biology, Christian-Albrechts-University of Kiel, Gutenbergstr. 76, 24118 Kiel, Germany

² La Trobe Institute for Agriculture & Food, Department of Animal, Plant and Soil Sciences, La Trobe University, Melbourne, Victoria 3086, Australia

³ Sino-Australia Plant Cell Wall Research Centre, School of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, China

Corresponding author:

Birgit Classen, Pharmaceutical Institute, Department of Pharmaceutical Biology, Christian-Albrechts-University of Kiel, Gutenbergstr. 76, 24118 Kiel, Germany

Phone: +49-431-8801130

e-mail: bclassen@pharmazie.uni-kiel.de

Contents

Supplementary Tables	3
Supplementary Table S1. Tissue-specific neutral monosaccharide composition of HMFs	3
Supplementary Table S2. Tissue-specific neutral monosaccharide composition of AGPs	3
Supplementary Table S3. Mass and volume of whole plant AGPs	3
Supplementary Figures	4
Supplementary Figure S1. Mass spectrum of the uronic acid reduced sample.	4
Supplementary Figure S2. Comparison of inferred <i>Z. marina</i> and <i>A. sativa</i> glycans.	5
Supplementary Figure S3. Sequences with diagnostic motifs highlighted.	6
Supplementary Figure S4. Domain families of all chimeric AGPs.	8
Supplementary Figure S5. Most common chimeric AGP domain architectures.	9
Supplementary Figure S6. Phylogeny of GT31 family members.	10
Supplementary Figure S7. Phylogeny of GT14 family members.	11
Supplementary Figure S8. Phylogeny of arabinosyltransferases families.	12
Supplementary Data Files	13
Supplementary Data S1. Classical HRGPs from MAAB (signalP filtered)	13
Supplementary Data S2. Chimeras by clan (signalP filtered)	13
Supplementary Data S3. Enzyme trimmed alignments and trees	13

Supplementary Tables

Supplementary Table S1. Tissue-specific neutral monosaccharide composition of HMFs

Neutral monosaccharide composition of high molecular weight fractions (HMF) from *Z. marina* in % (mol mol⁻¹).

Neutral monosaccharide	<i>Z. marina</i> whole plant (n=3)	<i>Z. marina</i> leaves (n=1)	<i>Z. marina</i> rhizome (n=1)	<i>Z. marina</i> root (n=1)
Gal	31.0 (± 0.8)	33.6	29.6	30.9
Ara	29.5 (± 1.7)	28.9	30.7	18.6
Rha	7.1 (± 0.1)	2.3	8.5	13.0
Man	6.8 (± 0.5)	12.0	4.9	6.5
Glc	12.5 (± 0.6)	14.2	17.1	13.5
Xyl	11.7 (± 0.8)	9.0	9.2	15.8
Fuc	1.4 (± 0.8)	-	-	1.7

Supplementary Table S2. Tissue-specific neutral monosaccharide composition of AGPs

Neutral monosaccharide composition of AGPs from the different organs of *Z. marina* in % (mol mol⁻¹).

Neutral monosaccharide	<i>Z. marina</i> whole plant (n=3)	<i>Z. marina</i> leaves (n=3)	<i>Z. marina</i> rhizome (n=3)	<i>Z. marina</i> root (n=3)
Gal	43.0 (± 0.4)	49.2 (± 0.8)	41.1 (± 4.2)	48.6 (± 0.9)
Ara	41.1 (± 0.5)	38.8 (± 1.0)	38.9 (± 2.1)	32.9 (± 0.7)
Rha	6.1 (± 0.0)	2.3 (± 0.1)	8.8 (± 1.1)	9.0 (± 1.0)
Man	4.6 (± 0.3)	3.8 (± 0.5)	3.6 (± 0.3)	2.1 (± 0.8)
Glc	3.0 (± 0.7)	4.7 (± 1.0)	7.1 (± 1.9)	6.4 (± 0.9)
Xyl	2.2 (± 0.5)	1.2 (± 0.1)	0.5 (± 0.3)	1.0 (± 0.4)

Supplementary Table S3. Mass and volume of whole plant AGPs

Determination of absolute molecular weights and hydrodynamic volumes of different whole plant AGPs from *Z. marina* (kDa)

Sample	Absolute molecular mass (main peak*)	Hydrodynamic volume (main peak*)
<i>Z. marina</i>	240.0 (± 0.3 %)	47.0
<i>Z. marina</i> UR	165.2 (± 0.2 %)	40.7
<i>Z. marina</i> Ox	164.1 (± 0.4 %)	12.8

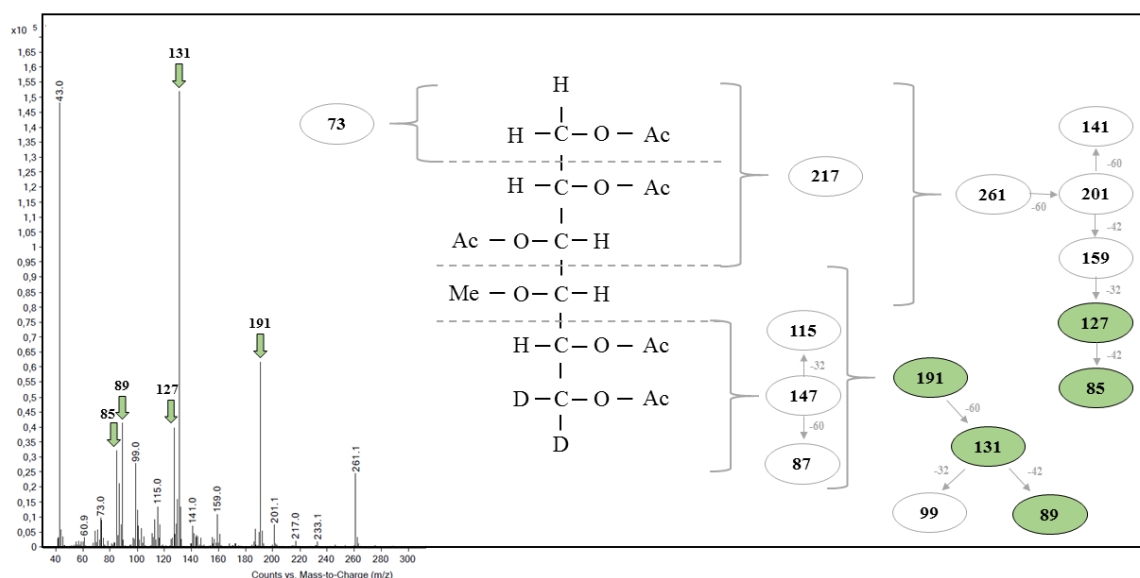
*main peak has a minimum mass recovery of 30% (m m⁻¹)

UR, uronic acid reduced; Ox, oxalic acid hydrolysed.

Supplementary Figures

Supplementary Figure S1. Mass spectrum of the uronic acid reduced sample.

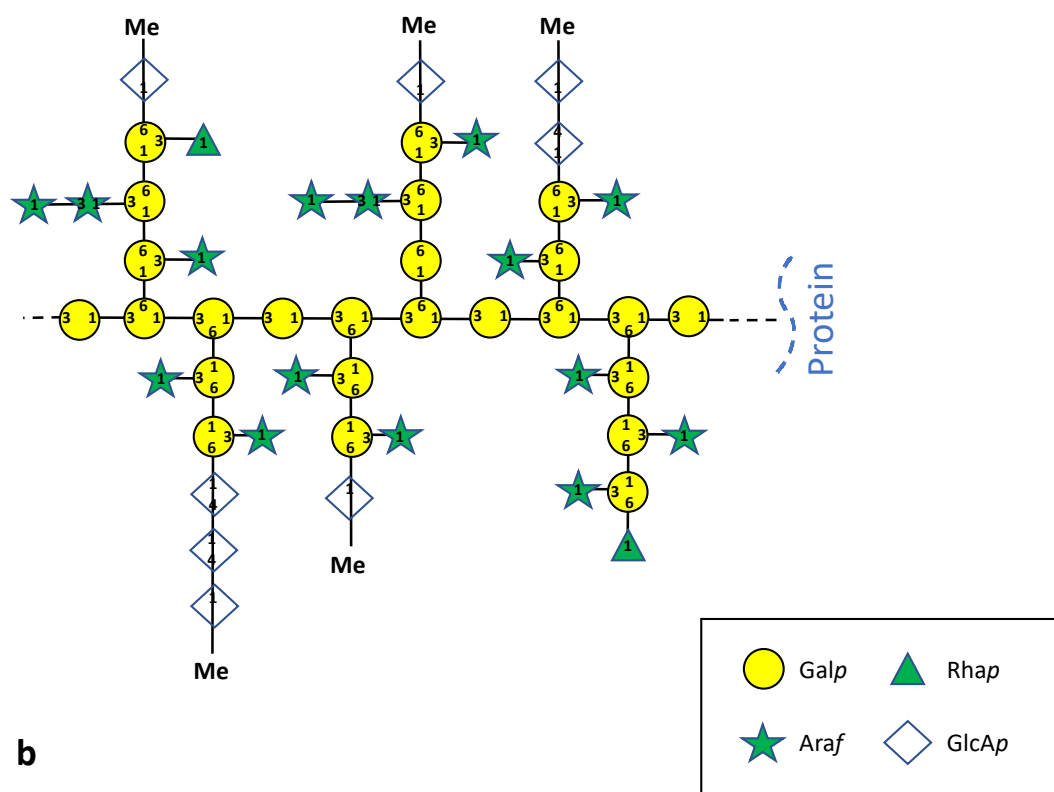
Mass spectrum and fragmentation pattern of C6-di-deuterated 4-OMe Glc in the uronic acid reduced sample of *Z. marina* whole plant AGP after alditol acetate analysis. In green the origins of the five most intense fragments, aside from the acetyl-group ($m/z = 43.0$), are indicated.



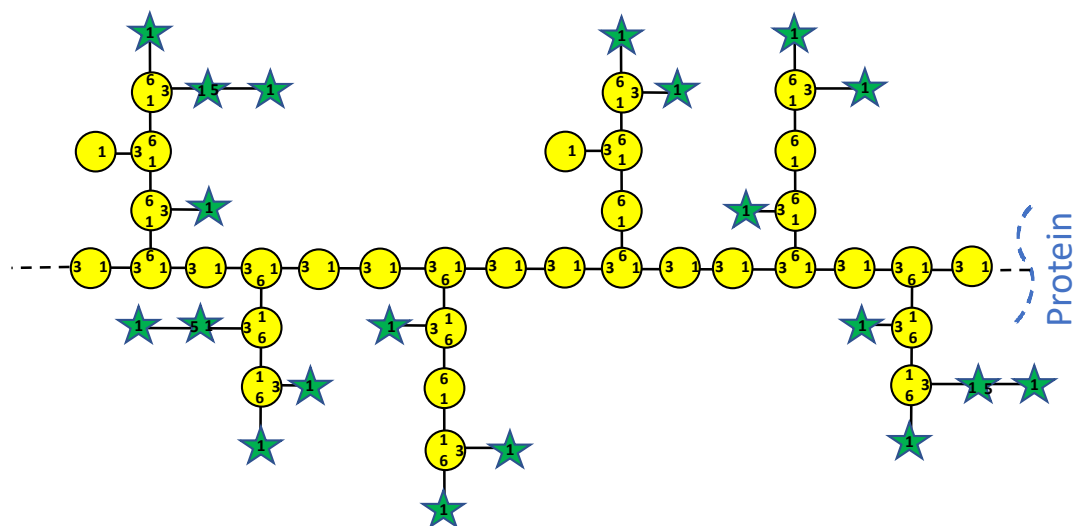
Supplementary Figure S2. Comparison of inferred *Z. marina* and *A. sativa* glycans.

Comparison of **a** *Z. marina* AGP (as presented in the main text) and **b** *A. sativa* AGP based on linkage-type analysis results from ²⁵.

a



b



Supplementary Figure S3. Sequences with diagnostic motifs highlighted.

Sequences identified by MAAB in *Z. marina* with diagnostic motifs highlighted and shown by MAAB class³⁰. Signal peptide in green, GPI anchor in purple and motifs used for classification of sequences are AGP motifs ([ASVTG]P, [ASVTG]PP, [AVTG]PPP) in cyan, Extensin motifs (SP3, SP4, SP5) in red and cross-linking motifs ([FY]XY, KHY, VY[HKDE], VxY, YY) in dark blue and PRP motifs (PPV[QK], PPVx[KT] and KKPCPP) highlighted brown. The total number of sequences identified in each class is shown in brackets. * indicates putative alternative start site.

Class 1: GPI-AGPs (5)

>KMZ66553.1 hypothetical protein ZOSMA_294G00050

MAIIPRYLQHLAVFFVFISVLFATTLAQALTDQSPLASTQPPSTSQLSPTFSA[SPTPTP]TS[SP]SATTFASLPPTP
SPSAAS[TPTI][SPTPSPSP]SAVSTSTIA[PTL]PLAASPLSD[GP]SGLS[SPAE][SPRS][SGSSSLVGCTGVGF]AALVLT
L

>KMZ65764.1 hypothetical protein ZOSMA_30G00450

MAITRTAISLFSIFLLASTSIFAQSPAT[PTL]PPPAPVAS[TPPP]ISVS[SPF]SVVAT[PPPV]VSA[APP]MMVST[PP]MSA
PVPMAAEVPSVETPVD[SPVAD]TPDEAMVPEASMPASPVFSSFA[PNIPN]PEG[SAGSTGAKMNVVGLMGLV]GAAAMI
I

>KMZ63928.1 hypothetical protein ZOSMA_38G00360

MTHRSILTUVFLLSIIASLHDAVFTSAAD[SPTT][SPSPA][SPPP]TATAT[PT]TTTTTS[VPPP]TAT[PT]TTNVTAVQP
PTAT[PT]PTTNVTAVQPPTAT[PT]TSSIT[TPAA][SPTTTA]PAT[PT]TTNV[TP]ISAPVS[SPPT]PTPAPTIT[TPAP]SLP
VE[VPAPAP]TKSKKKPSLPL[APSPSP]DSINL[AP]SSE[AP]GSISDDFLAADTA[DAATGGNGVVGFGLLV]VSLVAVV

>KMZ59493.1 hypothetical protein ZOSMA_68G00800

MGIRIQVTLIAFAILFATVIAQQAPASAPTS[TPGSVS][SPPSVS][SPSSLS][SPPSIS][SPPSLSS][APSS][SPPSMAPPS]
DQ[APS][APMT]PEALASSGSPGS[AP]SISQENNAASFAASWIGAAGSVALVMAYAF

>KMZ56883.1 hypothetical protein ZOSMA_8G00260

MAAPTFSFSAMLLHLLFLFLSLSVFSQA[KMA][SP]ISAISA[APAF]LPE[AP]SSLL[SPYPSY][SPTL][SPD]GSMQPEFPT
PRAEGV[AP]TTSIIITSVQ[SPPN]PDTMVPEAGDDDGFFVL[APV]GYSASIAADTS[SASSRVVAIRLIVVFLVLKSSLI]
LLLVSTFSC

Class 4: non-GPI-AGPs (4)

>KMZ70582.1 hypothetical protein ZOSMA_199G00190

MASHNLLWLLIIACFCSSLSTTTTQAQSQSASPVTLVPVAPVIPVS[SPPPE]KTLAS[PT]TTS[SP]LASPVIPVS[SPPPE]
KTLAS[PT]TTS[SP]LASPTLPVAPVLPVS[SPPPE]KILAS[PT]TPPP[SPSP]IIQ[TP][SPPPE]VLSLPPVSLPPALPI[TPAP]
AALAI[TPAV][SPSPAAD][SP]LAET[TPALAPAP]VHPKHKHKHRRWHRRKHRRHKKHHVM[APAPAP]IPP[SPPSPPAPPEF]
DDYTID[GPSPAP]TDLGGT[SIHG]IYKHGRRLSEGLRLTNFILIPFLFLF

>KMZ62892.1 hypothetical protein ZOSMA_43G00740

MGDLRSTTAAVLLILYFNMSAARILEGGGGGGGGGGGGGGGGGGNGG[VGY]GFGRGAG[VGY]GYS[APGAP]TSS[SP]IYS
QNSTTNQSGCS[APNT][TP][SPPPP]KEPC[APNTTV][TPPP]KEPC[TPNTTV][TPPP]KEPC[APNTTV][TPPP]KEPC[APNTTV]
PPPPPP[SPPVI][SPGPPSPSP][VI][SPGPP]PPVI[SPPTFVE][TPPP]PPNSVD[TPPP]PP[TPVE][TPPP]PPNSVD[TPPP]
P[TPVE][TPPP]PPNSVD[TPPP]PPAPVE[TPPP]PPNSVE[TPPP]PPNSVDMPPPPNSVDMPPPPPP[TPVE][TPPP]PPPP[TPV]
ETFP PPPPLNL[TPPP]PSLLDDEPPPLPPLVE[TPPP]PV[TPS][TP]TTTLPLIGVL[SPKEPC][TPNT][PSL][SPPPE]KEPCS
PNT[PSL][SPPPP]PPPEEPC[TPNT][PSL][SPPPE]KEPC[APNT][PSL][SPP]SHTKIPLQFPFK[TPS]FPHNENGKNWFGSG
SGFGHGAFFGGTYGSGGGGGGGGGGGGGGGGGNDASGYGSGYGNMGYGSFSGNDFMNAAGGGGGGGGGGGGGGFA
GMGN[YDYR]QFGGGGGGGGGGGGGGGGGFGQFGGGGGGGGGGGGGGGGGF

>KMZ58528.1 hypothetical protein ZOSMA_76G00830, partial

MKQKWEAWFLVACDVLLFSTLMDFPVCGGSKGKAL[SPAP]KVISSTLTTEPT[VP]SI[SP]LPSLLP[SP]LFST[VP]MI
PNV[SPF]ISRLHMPP[VPI]APTIPNV[SPSS]LPI[SPSL][VP]IVSEIPNV[SPF]SLPQF[SPSL][VP]MVSEIPNVSHSSLLP
[SPPPE]TPITSTISNV[SPSS]LLPP[SP]SLLPMT[PT]ITSNI[SP]LLHPP[SP]LLPIPAKILNS[SP]QLSLF[SPPPE]ASTAST
APNF[SPSS]LLPH[SP]LLPIPVTKPNLSHPPLPLNPPALTTST[TPK]V[SPF]LLPLP[SPF]VTIA[PTV]HNVS[TP]FLPH
NP SHFLKQ[PPVQ]NIS[TP]LLS[SPPPE]VLI[VP]PLHNIS[APLFP][SPF]LAHI[APT]IENIPSL[VPF]SFPLGR[SP]SQQDRS
SS[AP]IDRHPRTLD[SPPAR]PKNSIPTQS[SPSH][SPSLAP]DASTSAGTASHLSHNHH[SPVRGS][VPVS][SPKPPK]SLTY
[SP]LGSHTRFPNRSKLLHSSA[AP]NSHHHQSRNTSSI[SP]SNDK

>KMZ57387.1 hypothetical protein ZOSMA_86G00220

MDHRSVATVGLLMCIVAMSVGVHAMSPASASAPTITITSEISPTSESFPVADTFSSPTVIESPPTPAFVVDAIPDS
PPSMEVNSPPTPAFVVDTIPTVSIPDSPPFMEVNSPPTPPFVVDAIPPTVSIPDSPPFMEINSPTPAFVVVTS
SPDSPPFVENSPPTPAFVVNAAPPMSSLPDSPPFVENSPPTPAFVVDAAPPMSSLPDSPPFVENSPPTPAFIV
APATPPFITSLPDTPPFVVASVPRSGAPVVASAPRSVLVPTSAPLSPTPAASPTTILSPVPAPVEVTFDGSF
FVDAPAPSSSTSVAPSLSEAFPPGPSPDPAIFADDTARGVKENPMIAGAVITMLMLMSSFAVVFF

Class 2: CL-EXT (1)

>KMZ72364.1 hypothetical protein ZOSMA_166G00640

MTEGRDGPERRRRKFLPSI*MVAAFAVVLVVTNMGVVGSDPYIYS SPPPPYNYE SPPPPYKYE SPPPPYK
YESPPPPYKYE SPPPPYKYK SPPPPYKYE SPPPPYKYE SPPPPYKYE SPPPPYKYE SPPPPYKYE SPP
PPYKYE SPPPPYKYESLPPPYKYE SPPPPYKYE SPPPPYKYE SPPPPYKYE SPPPPYKYK SPPPPYK
YESPPPPYKYE SPPPPYKYK SPPPPYKYE SPPPPYKYE SPPPPYKYE SPPPPYKYK SPPPEVYKYE SPP
PPYKYK SPPPPAYKYE SPPPPYKHE SPPSYHK SPPPKPYPPHHNRIFKVVGVEVCYGCYDSEHSFKSHHKK
KLEGAVVKVTCMKGSKDVIAYGKTMNGMYEITVEDYDVKYGHENCKAMLHAPPKGSSCNMATDIHNGNTGAKL
HIKSRNHVEV VYKSKKFAYAPKTPYKDCYEKRHHHHHHHHHSPVPYCKPPPTYYYK SPPPTPTYYYK SPPPA
PTYHYKPPSPKYYYK SPPPPSPKYHYK SPPPPSPKYHYK SPPPPSPKYHYK SPPPPSPKYHYK SPPPPAP
TYYYKPPSPKYYYK SPP
PPSPKYYYK SPPPPVH SPPPPYYYK SPPPPVH SPPPPYYYK SPPPPVH SPPPPYYYK SPPPEH SPPPPYHYK S
PPEK SPPPPYHYE SPPPEKSEPHFYIYS SPPPEIYK

Class 20: Shared Bias, high EXT (SPn & Y) (1)

>KMZ74158.1 hypothetical protein ZOSMA_133G00070

MRKITASFFLILTALASPGDSADSAKLIGVAECADCGNNAFGSFKGINVAVVCNSEINLVDFKEVAVGEFAGD
GKLSLQLPTTIVDKKCFAHVRSLSKTNPCPTFQNLDNFILSLSSDDQSVYVFGNSDGKVSFSRAACAQKTFWKYE
YFKCPNHPWFYKLYPCNPPPSNPPP VYKNPPP VYKNPPP VYKNPPP VYKNPPP VYKNPPPCNPPPYKKP VYKSP
PVHK SPPPVHKNPPPCNPPPYKKPVDKNP SPYKPPVHNTPPMHMKPPSPKYNPPVHKPAPCNTPIHKPPMFKLPP
IYKPPVYVPKHPKTTTSN

Class 21 Shared Bias, high SPn (1)

>KMZ66204.1 hypothetical protein ZOSMA_2G02140

KPLFVLVLILLFAIVSTVKADERRRCVRGCHRLHRCRNTTYSHGHRHCLRKYRRCIIICNNKLPPPPPPPPSP
PPFSLPSCAHCVRGCYQRLHRCRNTTYHHGHRHCLRRYRRCIIICNNKFPPPPF SPPPPSPPPPFPPPP SPPPF
SPPPPSPPPPSPPPPFPP SPPPPSPPPPFPP SPPPPSPPPPFPP SPPPPSPPPPFPSLPPS SPPPPFPPPLC
FTTLKSCVNGCGHNDPCVLKCSNDYKVCRSNVP IIDHTK

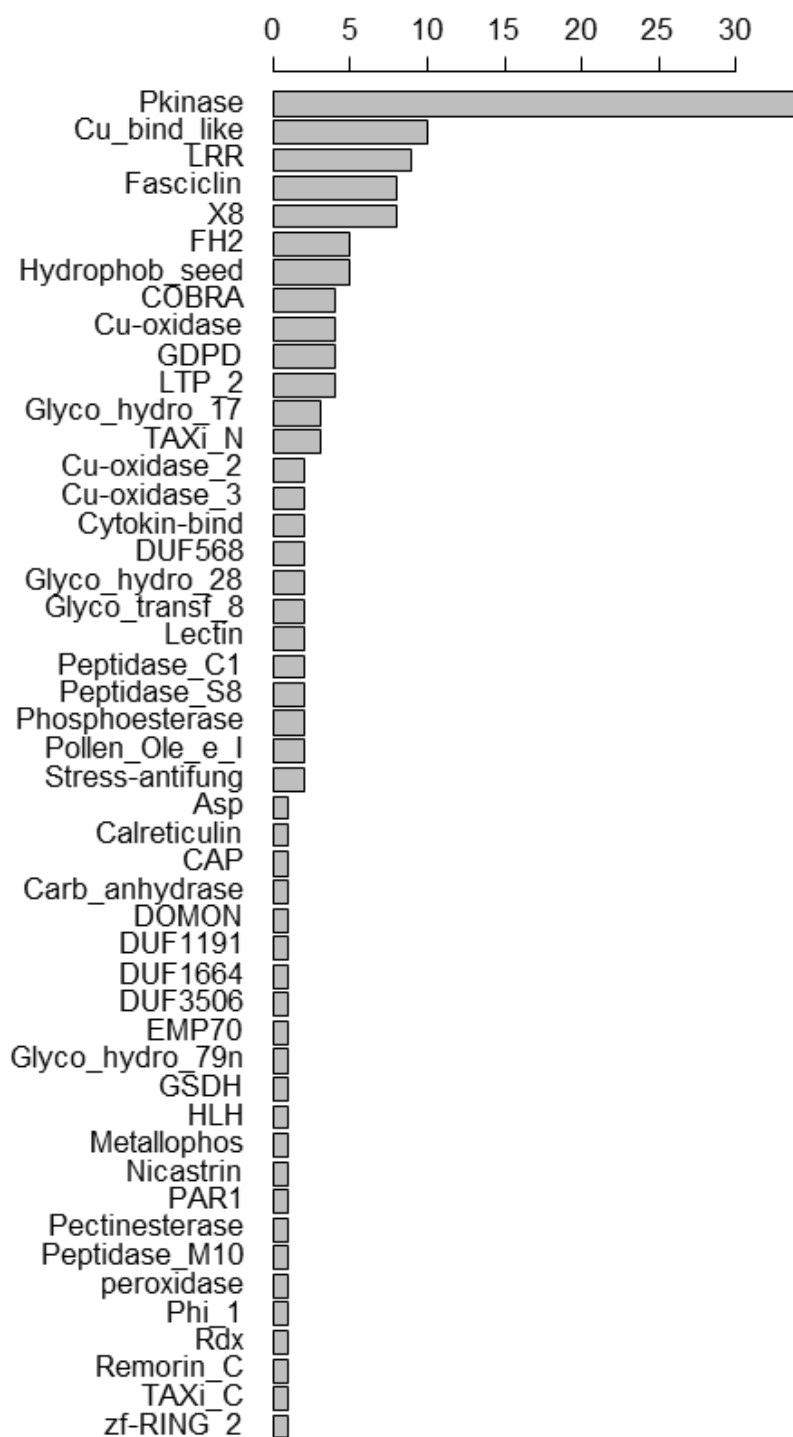
Class 24: <15% motif HRGP (3)

>KMZ58455.1 hypothetical protein ZOSMA_76G00110

MKPCVVSFLVIAAISLQIILVAPLVCGRTLQDFDDQKAYYHNSPPKTSHSGSHSSSHSGSSSFPSPFHQGGCAKTP
SHSSSSSTPKPRDGSYGTPPTPSHGSRSATTPATTPPTSSHATTPPTPSHSTTPSTPSHTTTPATPSIPGFPSIT
ATCDFWRTHPSMIFGILGQWSNIGNLFGFPATSI FGRNPSVPQALGNARNDDGYGALFREGTASLLNSMANPSFPL
TTQVVRDRFNQALSSEKTASAEAQRFRLANEGA

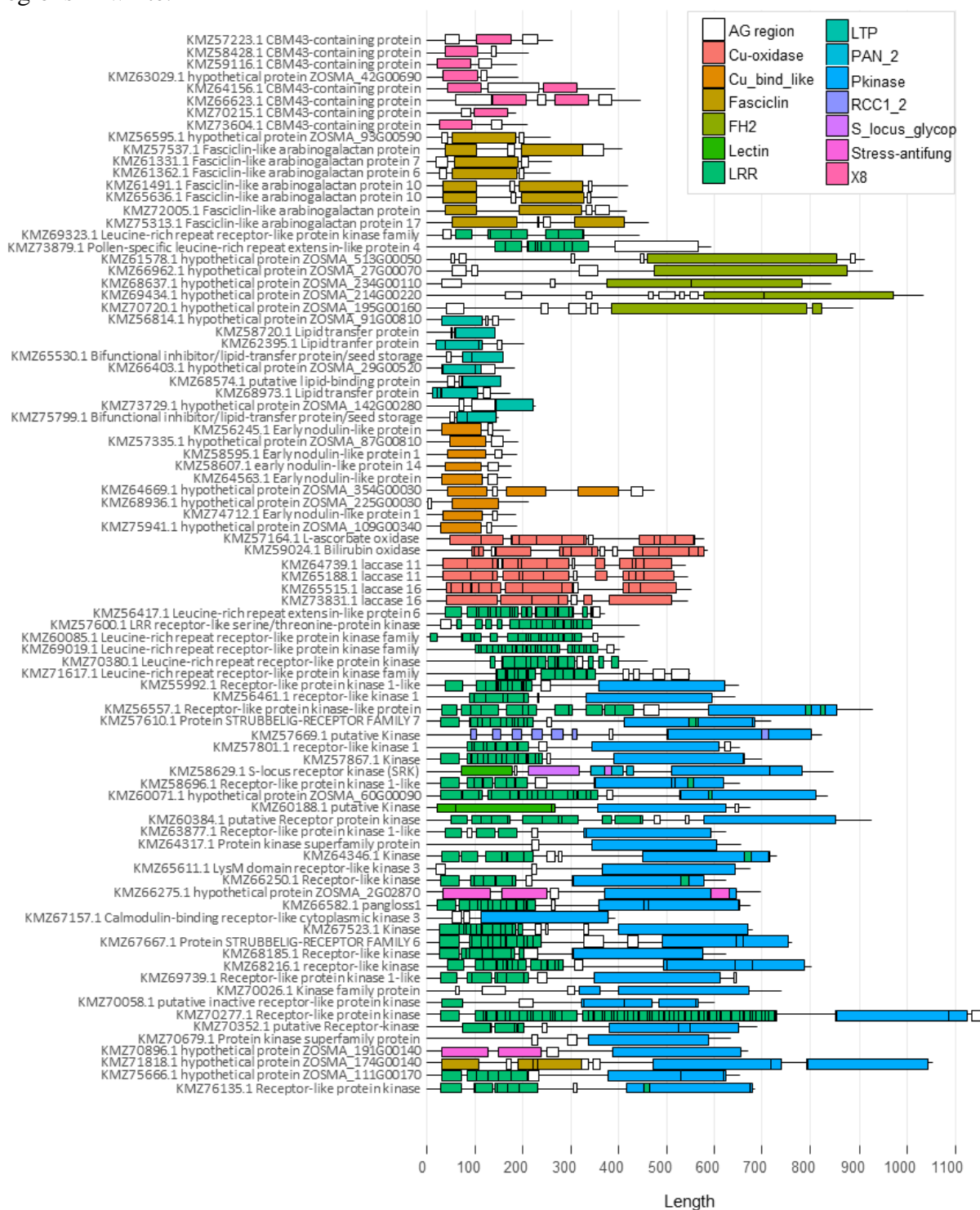
Supplementary Figure S4. Domain families of all chimeric AGPs.

Pfam domains found in sequences containing at least 1 AG region and a predicted signal peptide. For those sequences with multiple Pfam domains, only the largest domain is counted.



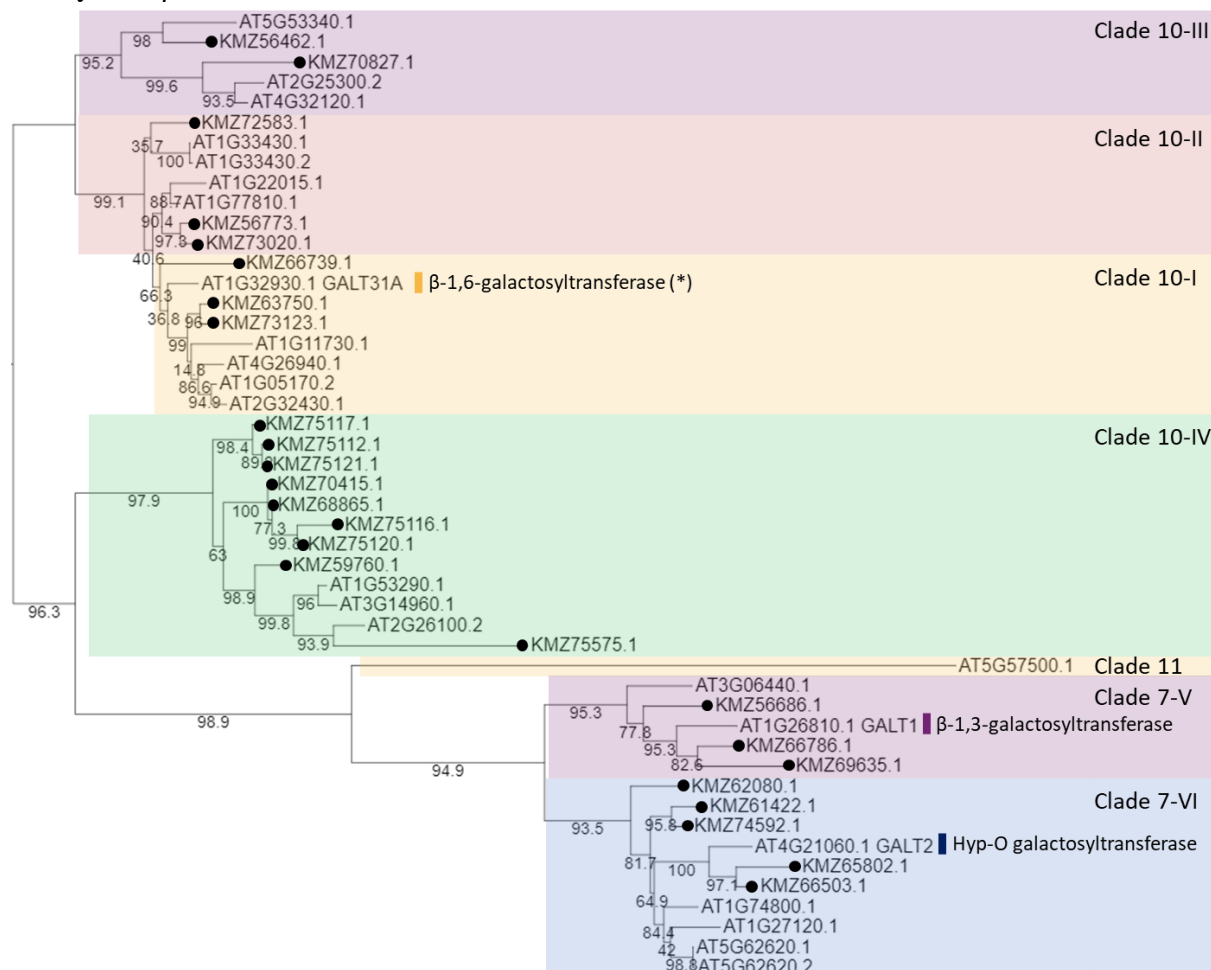
Supplementary Figure S5. Most common chimeric AGP domain architectures.

Architectures in *Z. marina* with domains coloured by their Pfam family (see Fig. 5b). AG regions in white.



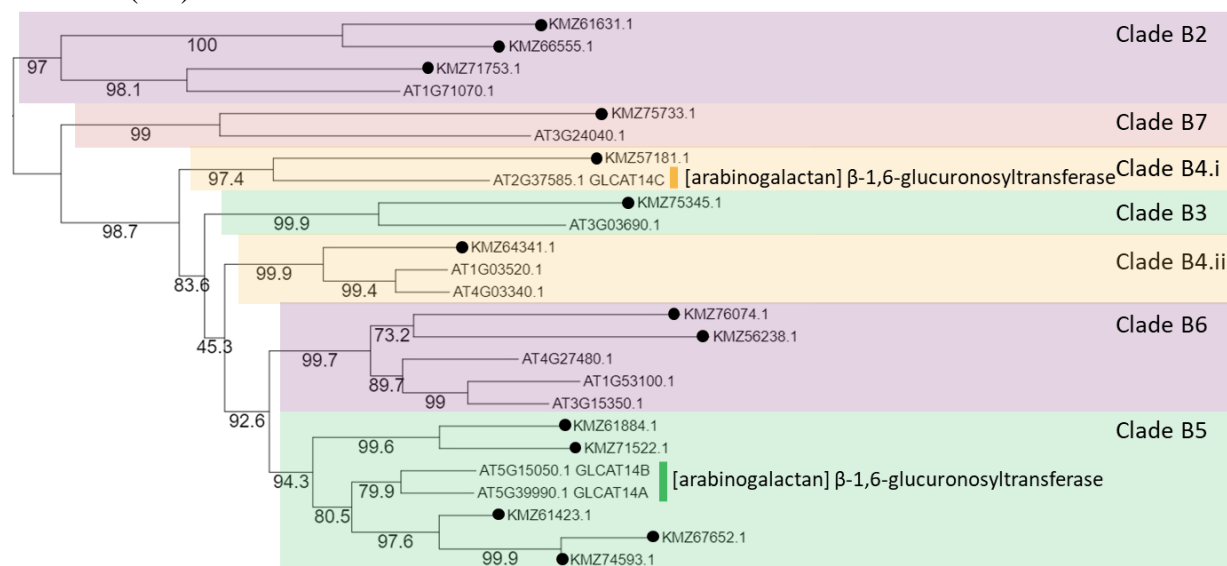
Supplementary Figure S6. Phylogeny of GT31 family members.

Members from *Z. marina* (ZM accession numbers, highlighted with black circles), *Arabidopsis thaliana* (AT). Clades labelled as in ³¹. Genes with known function indicated. * Note: although QMAP8 is described as a β -1,6 GalT in the published literature, unpublished data suggests that it may be a β -1,3 GalT.



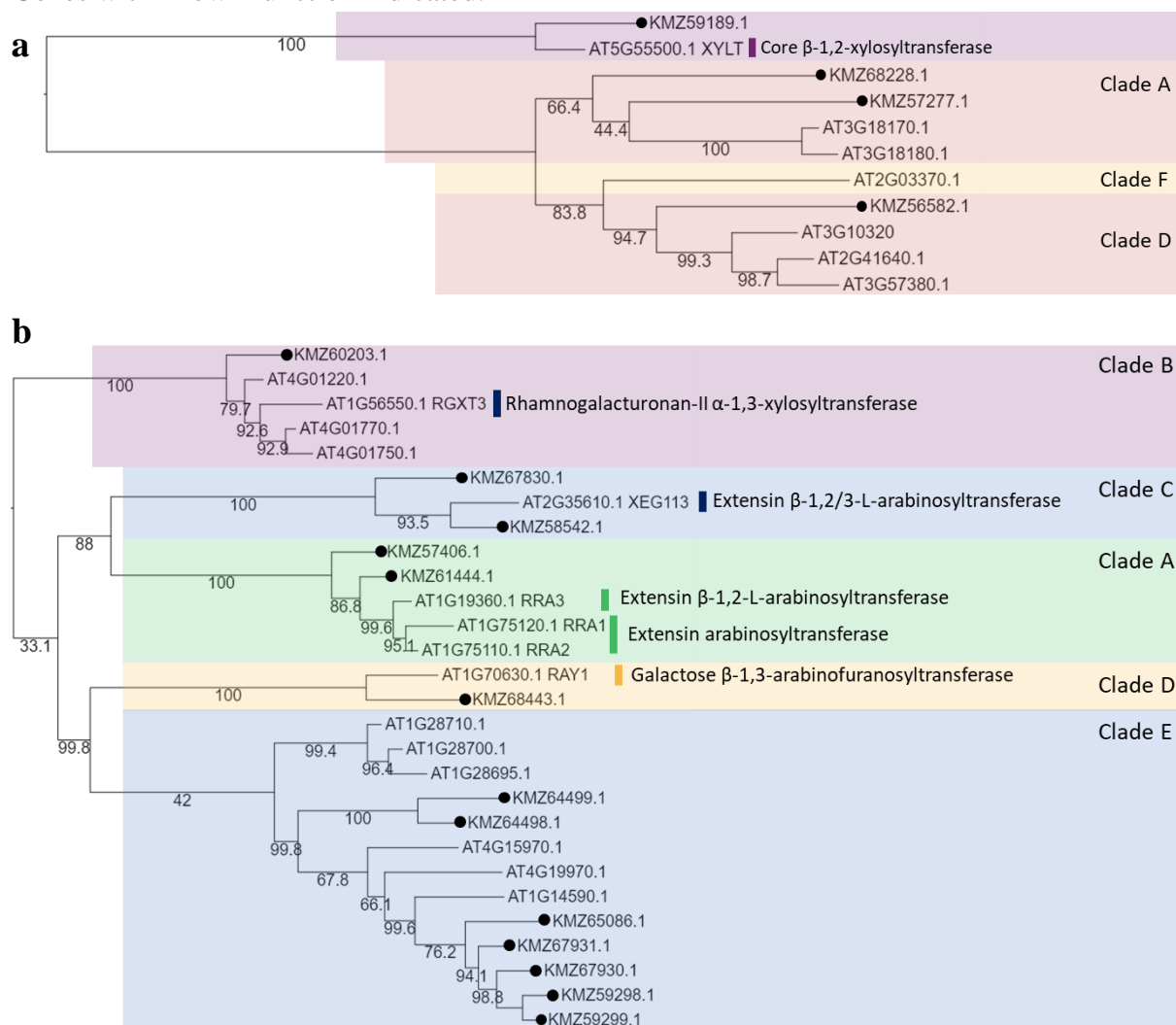
Supplementary Figure S7. Phylogeny of GT14 family members.

Members from *Z. marina* (ZM accession numbers, highlighted with black circles), *Arabidopsis thaliana* (AT). Clades labelled as in ³². Genes with known function indicated.



Supplementary Figure S8. Phylogeny of arabinosyltransferases families.

a Phylogeny of GT61 family members from *Z. marina* (ZM accession numbers, highlighted with black circles), *Arabidopsis thaliana* (AT). Clades labelled as in ³³. Genes with known function indicated. **b** Phylogeny of GT77 family members from *Z. marina* (ZM accession numbers, highlighted with black circles), *Arabidopsis thaliana* (AT). Clades labelled as in ³⁶. Genes with known function indicated.



Supplementary Data Files

DOI: [10.26181/5e8c491575d77](https://doi.org/10.26181/5e8c491575d77)

Supplementary Data S1. Classical HRGPs from MAAB (signalP filtered)

Unaligned sequences of HRGPs as fasta file

Supplementary Data S2. Chimeras by clan (signalP filtered)

Unaligned sequences of chimeric AGPs as fasta files, organised by clan of the globular domain

Supplementary Data S3. Enzyme trimmed alignments and trees

Multiple sequence alignments of enzymes as fasta files and phylogenetic trees as newick files