

Article

The In Silico Characterization of Monocotyledonous α -L-Arabinofuranosidases on the Example of Maize

Alsu Nazipova ^{1,*}, Olga Makshakova ²  and Liudmila Kozlova ¹ 

¹ The Laboratory of Plant Cell Growth Mechanisms, Kazan Institute of Biochemistry and Biophysics FRC Kazan Scientific Center of RAS, Lobachevsky Str. 2/31, 420111 Kazan, Russia

² The Laboratory of Biophysical Chemistry of Nanosystems, Kazan Institute of Biochemistry and Biophysics FRC Kazan Scientific Center of RAS, Lobachevsky Str. 2/31, 420111 Kazan, Russia

* Correspondence: alsu.nazipova@kibb.knc.ru

Abstract: Plant α -L-arabinofuranosidases remove terminal arabinose from arabinose-containing substrates such as plant cell wall polysaccharides, including arabinoxylans, arabinogalactans, and arabinans. In plants, de-arabinosylation of cell wall polysaccharides accompanies different physiological processes such as fruit ripening and elongation growth. In this report, we address the diversity of plant α -L-arabinofuranosidases of the glycoside hydrolase (GH) family 51 through their phylogenetic analysis as well as their structural features. The CBM4-like domain at N-terminus was found to exist only in GH51 family proteins and was detected in almost 90% of plant sequences. This domain is similar to bacterial CBM4, but due to substitutions of key amino acid residues, it does not appear to be able to bind carbohydrates. Despite isoenzymes of GH51 being abundant, in particular in cereals, almost half of the GH51 proteins in *Poales* have a mutation of the acid/base residue in the catalytic site, making them potentially inactive. Open-source data on the transcription and translation of GH51 isoforms in maize were analyzed to discuss possible functions of individual isoenzymes. The results of homology modeling and molecular docking showed that the substrate binding site can accurately accommodate terminal arabinofuranose and that arabinoxylan is a more favorable ligand for all maize GH51 enzymes than arabinan.

Keywords: α -L-arabinofuranosidase; arabinoxylan; maize; homology modeling; molecular docking



Citation: Nazipova, A.; Makshakova, O.; Kozlova, L. The In Silico

Characterization of Monocotyledonous

α -L-Arabinofuranosidases on the Example of Maize. *Life* **2023**, *13*, 266. <https://doi.org/10.3390/life13020266>

Academic Editors: Kousuke Hanada and Yoshiteru Noutoshi

Received: 13 October 2022

Revised: 26 December 2022

Accepted: 14 January 2023

Published: 18 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

α -L-Arabinofuranosidases (EC 3.2.1.55) are exo-active glycoside hydrolases (GHs) that remove α -L-arabinofuranose from the non-reducing end of arabinose-containing substrates. The α -L-arabinofuranosidases have been classified into eight GH families: 2, 3, 10, 43, 51, 54, 62, and 159 [1]. Bacteria and fungi use α -L-arabinofuranosidases to degrade plant cell wall polysaccharides such as arabinoxylan, arabinan, arabinogalactan, and their oligosaccharides [2,3]. These enzymes are also used in cocktails to convert plant biomass for different industrial purposes [4], making their investigation more important from a practical point of view.

In plants, α -L-arabinofuranosidases are known to be involved in modification/degradation of cell wall polysaccharides during various physiological processes such as seed storage mobilization [5], fruit ripening [6,7] and fruit softening [8,9], turnover of arabinogalactan proteins during seed maturation [10], and in the proper release of seed mucilage from secretory cells [11]. In monocotyledons (*Liliopsida*), α -L-arabinofuranosidase activity is increased during cessation of elongation growth and presumably contributes to the arabinoxylan debranching [12–14]. The latter is important for maintaining the correct architecture of plant cell walls, especially primary cell walls of type II, where arabinoxylan serves as the main hemicellulose [15].

In higher plants, enzymes with α -L-arabinofuranosidase activity belong to GH families 3 and 51 [16]. However, members of GH3 demonstrate β -D-glucosidase (EC 3.2.1.21),

β -D-xylosidase (EC 3.2.1.37) or bifunctional β -D-xylosidase/ α -L-arabinofuranosidase activities in dicotyledons [11,17,18] and monocotyledons [19], whereas GH51 enzymes exhibit mostly an α -L-arabinofuranosidase activity with a single example of a bifunctional α -L-arabinofuranosidase/ β -D-xylosidase enzyme [17,20,21]. Some bacterial GH51 enzymes possess endoglucanase (EC 3.2.1.4), xylanase (EC 3.2.1.8), and cellobiohydrolase (EC 3.2.1.91) activities [1,22–24], but no such reports have yet been made for eukaryotes. For simplicity, we will refer to the plant sequences belonging to GH51 as α -L-arabinofuranosidases, as this is most likely, but it should be considered that this may not be true for all the sequences.

The GH51 enzymes are retaining glycosidases that preserve the anomeric configuration of the substrate after hydrolysis. Like most retaining enzymes, they use two glutamic acids as an acid/base and a nucleophile [25]. GH51 enzymes belong to clan A of GHs, so their active site is located at the N-terminal (β/α)₈-barrel. Typical enzymes also possess a C-terminal β -sandwich domain that stabilizes the overall structure [25–30]. However, a recent study of the fungal enzyme MgGH51 showed the presence of an N-terminal domain similar to the carbohydrate-binding module (CBM) of family 4 with unknown function in addition to the aforementioned domains [31].

Plant members of the GH51 family do not have resolved crystal structures, but some have been studied biochemically. The barley enzyme HvAXAH1 removes arabinose from mono- and di-substituted arabinoxylans at high and low rates, respectively, utilizing sugar beet arabinan and larch wood arabinogalactan, but the latter at a low catalytic rate [21]. Rice enzymes OsARAF1 and OsARAF3 are involved in the degradation of primary cell wall arabinoxylans, which was revealed in mutant plants overexpressing the corresponding genes [32]. In the dicotyledonous plant *Arabidopsis thaliana*, the α -L-arabinofuranosidase AtARAF1 is involved in the modification of pectic arabinans as shown in overexpressing and knock-out mutant plants [33,34].

We previously identified eight genes in the maize genome as encoding GH51 α -L-arabinofuranosidases (*ZmaABF1-8*) and showed that the expression of most of them increases with the development of primary maize root cells, as well as the total α -L-arabinofuranosidase activity [14]. However, the *ZmaABF2* gene showed the opposite pattern of gene expression. It was up-regulated in the root cap and meristem [14]. A recent transcriptomic study confirmed this pattern and additionally pointed to co-expression of *ZmaABF1* with genes of secondary cell wall cellulose synthases, indicating the possible involvement of this enzyme in secondary cell wall formation [35]. Cereals have a notably evolved GH51 family [34], but neither the specific targets for these enzymes nor their physiological role are fully understood.

In this study, we re-predicted maize genes encoding putative α -L-arabinofuranosidases of the GH51 family in the most recent to date version of the maize genome. Transcription and translation data for these genes in different organs and tissues of maize were analyzed. Homology modeling using the resolved structures of α -L-arabinofuranosidases as templates and molecular docking experiments were performed to establish substrate preferences for maize GH51 enzymes. Phylogenetic analysis established the evolution of α -L-arabinofuranosidases in plants, revealing a widespread potentially deactivating mutation in the enzymes of some orders of monocotyledonous plants. A domain resembling the carbohydrate binding module that is specific exclusively to GH51 was revealed and its ability to interact with carbohydrates was discussed.

2. Materials and Methods

2.1. Identification of GH51 Proteins in Different Plant Species

The sequences of plant α -L-arabinofuranosidases were recognized in the genomes and transcriptomes of various plants in three ways: (1) searching for PF06964 (Alpha-L-AF_C) Pfam domain name in the Phytozome database v13 (<https://phytozome-next.jgi.doe.gov/>; [36]) (accessed on 1 April 2022); (2) Basic Local Alignment Search Tool (BLAST) search in the One Thousand Plant Transcriptome Project database (1KP, <https://db.cngb.org/onekp/species/>; [37]) (accessed on 20 April 2022) using AtARAF1 of *A. thaliana* as a

query; (3) local searching for the PF06964 domain in preliminarily downloaded genomes using the *hmmsearch* function in the HMMer v3.3 [38]. The domain organization of the sequences found was verified using the InterProScan (<https://www.ebi.ac.uk/interpro/search/sequence/>; [39]) (accessed on 20 April 2022) and/or the web-based tool Batch-CD-search (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>; [40]) (accessed on 20 April 2022).

The PF06964 domain profile was downloaded from the Pfam 32.0 database (<http://pfam.xfam.org/>; [41]) (accessed on 19 April 2022). Plant genomes were downloaded from the EnsemblPlants (<http://plants.ensembl.org/index.html>; [42]) and GenBank (<https://www.ncbi.nlm.nih.gov/protein/>; [43]) databases (accessed on 16 April 2022); *Klebsormidium nitens* alga genome was downloaded from the *K. nitens* NIES-2285 genome project (http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/index.html; [44]) (accessed on 16 April 2022), and plant transcriptomes were downloaded from the 1KP database. The resulting plant GH51 protein sequences were aligned using the Muscle algorithm (<https://www.ebi.ac.uk/Tools/msa/muscle/>; [45]) (accessed on 21 April 2022). Sequences less than 350 residues and/or not containing two conserved G-N-E/Q and S-E-Y-x-V motifs simultaneously were excluded from further analysis. A total of 139 out of 156 sequences remained after this filtering.

Three protein sequences (M8CET9, M8CFW8 of *Aegilops tauschii*, and 56386 of *Coccomyxa subellipsoidea* (listed in Table S1)) were re-predicted in the web tools Augustus (<https://bioinf.uni-greifswald.de/augustus/>; [46]) and FGENSESH (<http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>; [47]) (accessed on 21 April 2022).

The maize genes encoding GH51 proteins were identified as described above in the maize B73 genome downloaded from the GenBank (Zm-B73-REFERENCE-NAM-5.0 assembly). The genes encoding GH51 α -L-arabinofuranosidase in the RefGen B73 AGPv3 5b+ genome assembly (downloaded from the Maize Genome database (MaizeGDB, <https://www.maizegdb.org/>; [48]) (accessed on 2 March 2022) were aligned to the Zm-B73-REFERENCE-NAM-5.0 genome assembly using a BLAST search of translated nucleotide databases using a protein query (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>; [49]) (accessed on 3 March 2022).

The presence of signal peptide in the maize sequences was predicted using the web tool SignalP 6.0 (<https://services.healthtech.dtu.dk/service.php?SignalP>; [50]) (accessed on 5 March 2022).

2.2. GH51 Gene and Protein Expression in Maize

Data on GH51 gene expression in different maize organs were obtained by Walley et al. [51] and Hoopes et al. [52] and downloaded from the MaizeGDB database. The proteome data were retrieved from the study by Walley et al. [51].

2.3. CBM4-like Domain Analysis

To create a CBM4-like domain profile, 11 bacterial, 16 fungal, and 111 plant GH51 sequences (listed in Table S1) were used. The random full-length GH51 sequences of bacteria and fungi were obtained by protein BLAST search using the fragment (1–238 residues from the N-terminal end) of AtARAF1 as a query. Several sequences listed in Lagaert et al.'s (2014) review [2] and the sequence of MgGH51 were also included. The plant sequences were obtained as described above. The domain organization of all sequences was verified using InterProScan. The CBM4-like domain was established between two conserved L-x-V-D and S-x-x-P motifs (Figure S1) in the multiple sequence alignment (Muscle algorithm) of all obtained sequences. The parts of the alignment following the S-x-x-P motif were removed. The truncated alignment was converted into a Pfam domain profile using the *hmmbuild* function in HMMer-3.3.

To estimate the proportion of bacterial, fungal, insect, and plant GH51 proteins which contain a CBM4-like domain, the search of the generated domain profile using the *hmm-*

scan function (HMMer-3.3.) was performed. The protein sequences contained either the PF06964 domain or both the PF02018 and PF06964 domains, which were downloaded from the Uniprot (<https://www.uniprot.org/>; [53]) and Phytozome (<https://phytozome-next.jgi.doe.gov/>; [36]) databases (accessed on 17 May 2022). The sequences longer than 200 residues were used for search. The GH51 protein sequences of insects were found by the BLASTP search of the fragment (1–238 residues from N-terminal end) of AtARAF1 as a query in eukaryotes excluding fungi and plants.

The multiple sequence alignment was followed by a structural superposition determined by super method in the PyMOL Molecular Graphics System v2.2.0 [54]. All spatial structures were visualized in PyMOL.

The multiple sequence alignment of bacterial CBMs of family 4 and CBM4-like domains of MgGH51 and some plant GH51 proteins was performed using the Guidance 2 server with default parameters (<http://guidance.tau.ac.il/>; [55]) (accessed on 17 May 2022).

2.4. Phylogenetic Analysis

Sequences were aligned using the Muscle algorithm and the resulting alignments were subjected to phylogenetic analysis in IQ-TREE v2.1.3 [56]. The best-fit model of sequence evolution was computed using Model Finder (IQ-TREE) [57] and selected using a Bayesian Information Criterion. To construct phylogenetic dendrograms, an ultra-fast bootstrap branch support based on 10,000 replicates [58] was used. Trees were visualized in the iTOL v6.5.3 web service (<https://itol.embl.de/>; [59]) (accessed on 26 April and 17 May 2022). Abbreviated names of genes and proteins were used; full names are listed in Tables S1 and S2. Phylogenetic analysis of plant sequences of monocotyledons was preceded by the removal of ambiguous regions of multiple alignments using Gblocks v0.91b [60] with the following parameters: the maximum number of contiguous unconserved positions was 15; the minimum block length was 2; the allowed gap position was set as half (only positions where 50% or more of the sequences have a gap are treated as a gap position [60]).

2.5. Homology Models of GH51 Proteins

The crystal structure of MgGH51 from *Meripilus giganteus* co-crystallized with arabinose (PDB ID: 6zpy) was used as a template for homology modeling of plant proteins in MODELLER v. 9.15 [61]. The arabinose ligand was copied to the resultant models from the template to shape the highly conservative binding site of the model as in X-ray structure. The protein sequences of GH51 proteins from maize and AtARAF1 from *A. thaliana* were obtained as described above. The sequence and structure of MgGH51 and Tx-Abf from *Thermobacillus xylanilyticus* (PDB ID: 2vrq) were obtained from the Protein Data Bank (PDB, <http://www.rcsb.org/>; [62]) (accessed on 2 April 2022). Protein sequence pairwise alignment was performed in the ClustalW service (<https://www.ebi.ac.uk/Tools/msa/clustalo/>; [63]) (accessed on 2 April 2022). The 36, 74, 36, and 42 amino acid residues from the N-terminal end of ZmaABF1, ZmaABF2, ZmaABF5, and AtARAF1, respectively, were left out during homology modeling due to the presence of signal peptides (ZmaABF1, ZmaABF5, and AtARAF1) and the absence of a homologous fragment in the MgGH51 sequence (ZmaABF2). The resulting least standard deviation models were examined using PROCHECK [64] and Verfiy3D [65] services.

2.6. Docking of GH51 Proteins with Different Oligosaccharides

Structures of the xylan oligosaccharide (β (1,4)-D-xylopyranose- $[\alpha$ (1,3)-L-arabinofuranose]-di- β (1,4)-D-xylopyranoside (XAXX)) from the Tx-Abf-XAXX complex crystal structure (PDB ID: 2vrq) and free α -L-arabinofuranose co-crystallized with the MgGH51 structure were used for docking. The structures of the pectin arabinan fragment (tetra- α (1,5)-L-arabinofuranoside) and arabinogalactan II fragment (β (1,6)-D-galactopyranose- $[\alpha$ (1,3)-L-arabinofuranose]-di- β (1,6)-D-galactopyranoside) were constructed in the Glycam web tool (<https://glycam.org/cb/>) (accessed on 5 April 2022), then the puckering and the primary hydroxyl position in the terminal arabinofuranose was adjusted to that in the crystal

structures. Molecular docking was performed in Autodock v4.2 software [66]. During docking, protein molecules remained rigid, while the linkages in the ligand were allowed to rotate sequentially. At the first step, the glycosidic linkages were kept flexible so that the terminal arabinosyl residue could adopt a position in the binding site and the xylosyl, galactosyl, or arabinosyl residues of the main chain could interact with the protein surface in close proximity. The best poses were used as initial geometries at the second step, where the hydroxyl groups of sugars were flexible to saturate hydrogen bonds with the protein. The docking poses with the most negative binding energy were selected for subsequent comparative analysis. Binding energies were calculated using Autodock as follows:

$$\Delta G_{\text{binding}} = \Delta E_{\text{vdW}} + \Delta E_{\text{Hbond}} + \Delta E_{\text{elec}} + \Delta E_{\text{desolv}} + \Delta E_{\text{tors}} \quad (1)$$

where ΔE_{vdW} is the contribution from van der Waals contacts, ΔE_{elec} —from electrostatic interactions, ΔE_{Hbond} —hydrogen bonds, ΔE_{desolv} —desolvation, ΔE_{tors} —entropic contribution.

The resulting plant structures in complexes with different ligands were analyzed and visualized in PyMOL software.

3. Results

3.1. The Identification of GH51 Genes in the Maize Genome

Eight maize genes (*ZmaABF1-8*) putatively encoding GH51 family enzymes were previously identified in the RefGenB73 AGPv3 genome assembly [14]. In the Zm-B73-REFERENCE-NAM-5.0 assembly [67] deposited in the GenBank, we identified five such genes and one pseudogene (*LOC109939272*) that matched *GRMZM2G393002* (*ZmaABF3*) (Table 1). The protein sequences encoded by *GRMZM2G120132* (*ZmaABF6*), *GRMZM2G452353* (*ZmaABF7*), and *GRMZM2G131440* (*ZmaABF8*) genes of the older genome version all displayed high similarity with that of *LOC109944448* of the newer assembly. However, the search of the *GRMZM2G120132* protein in the GenBank revealed no information about the encoding gene, whereas both *GRMZM2G452353* and *GRMZM2G131440* corresponded to the *LOC109944448* gene. So, probably *GRMZM2G120132* (*ZmaABF6*) does not really exist, and *GRMZM2G452353* (*ZmaABF7*) and *GRMZM2G131440* (*ZmaABF8*) are the same gene designated as *LOC109944448* in the latest genome assembly. We kept the names given in the first study [14] to corresponding genes from the newer genome assembly (Table 1). The *LOC109944448* was named as *ZmaABF7* because the *GRMZM2G452353* displayed a higher identity percent (99.4%) to translated *LOC109944448* compared to *GRMZM2G131440* (87.6%). However, *LOC109944448* encodes a short protein of 272 amino acids, and attempts to re-predict the gene were unsuccessful.

Table 1. The maize genes encoding proteins of GH51 family. aa, the number of amino acids of corresponding proteins; nd, no data.

RefGenB73 AGPv3		Zm-B73-REFERENCE-NAM-5.0			
Gene Locus IDs	Gene Name Used in the Study by Kozlova et al. (2015) [14]	Gene Locus/ Protein IDs	Gene Names Used in Present Study	Chromosome Number	The Protein Length, aa
<i>GRMZM2G006130</i>	<i>ZmaABF1</i>	<i>LOC100193890</i> / <i>NP_001337023</i>	<i>ZmaABF1</i>	3	656
<i>GRMZM2G077299</i>	<i>ZmaABF2</i>	<i>LOC100282597</i> / <i>XP_035822820</i>	<i>ZmaABF2</i>	4	675
<i>GRMZM2G393002</i>	<i>ZmaABF3</i>	<i>LOC109939272</i> (pseudogene)	nd	3	nd
<i>GRMZM2G097277</i>	<i>ZmaABF4</i>	<i>LOC100193089</i> / <i>NP_001345560</i>	<i>ZmaABF4</i>	2	663
<i>GRMZM2G180889</i>	<i>ZmaABF5</i>	<i>LOC103634588</i> / <i>XP_035819499</i>	<i>ZmaABF5</i>	1	670
<i>GRMZM2G120132</i>	<i>ZmaABF6</i>	nd	nd	nd	nd
<i>GRMZM2G452353</i>	<i>ZmaABF7</i>	<i>LOC109944448</i> / <i>XP_020404813</i>	<i>ZmaABF7</i>	2	272
<i>GRMZM2G131440</i>	<i>ZmaABF8</i>				

Four (*ZmaABF1*, *ZmaABF2*, *ZmaABF4*, and *ZmaABF5*) of five identified maize genes encoded the full-length proteins (Table 1). All of them had an Alpha-L-AF_C domain (IPR010720/PF06964) at the C-terminus and a glycoside hydrolase superfamily domain (Glycoside_hydrolase_SF, IPR017853) in the middle part of the sequence (Figure 1). Amino acid sequences of *ZmaABF1*, *ZmaABF2*, and *ZmaABF4* proteins also contained a galactose-binding-like superfamily domain (Galactose-bd-like_sf, IPR008979) in the N-terminal region (Figure 1). Unlike the other sequences, *ZmaABF2* did not have a signal peptide (Figure 1).

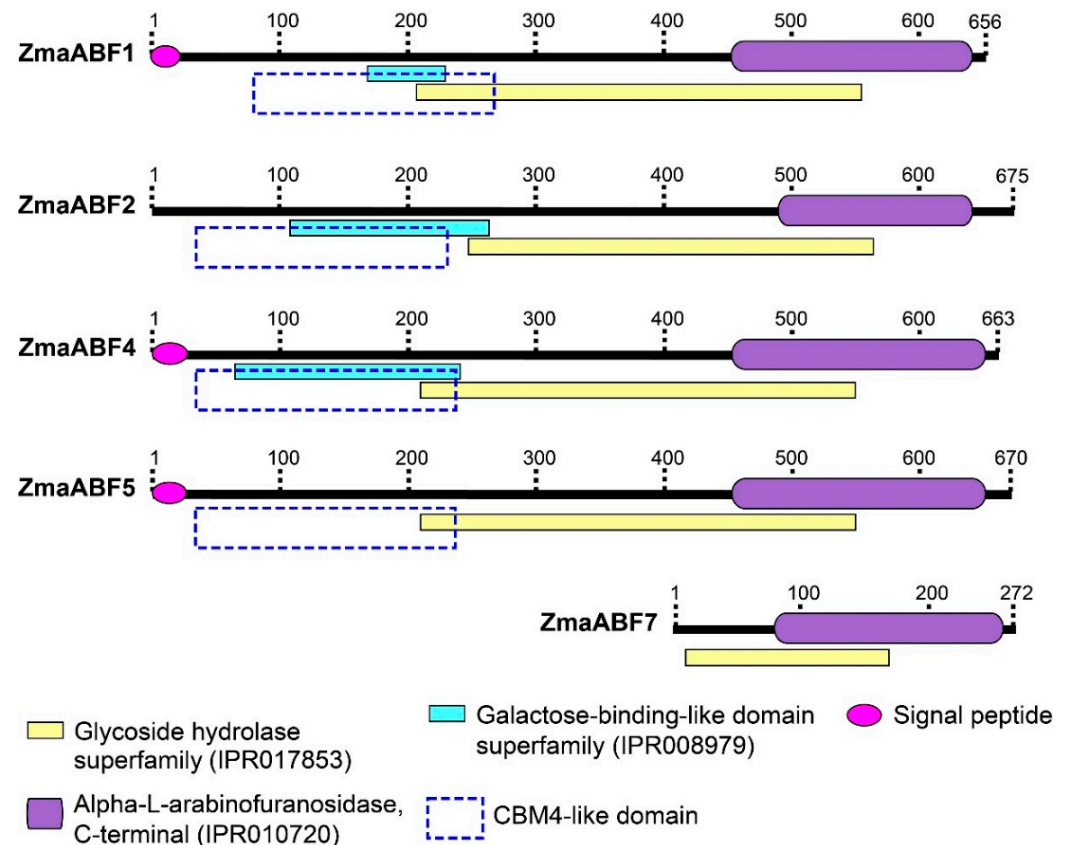


Figure 1. Domain organization of GH51 proteins encoded in Zm-B73-REFERENCE-NAM-5.0 maize genome assembly. The numbers above the domain schemes indicate the number of amino acid residues.

3.2. Abundance of GH51 Transcripts and Proteins in Maize Plants

In a multi-omics study by Walley and colleagues [51], transcripts of the *ZmaABF1* gene were found in almost all 23 maize tissues analyzed. They were particularly abundant in silks, seed embryos, and mature leaves (Figure 2).

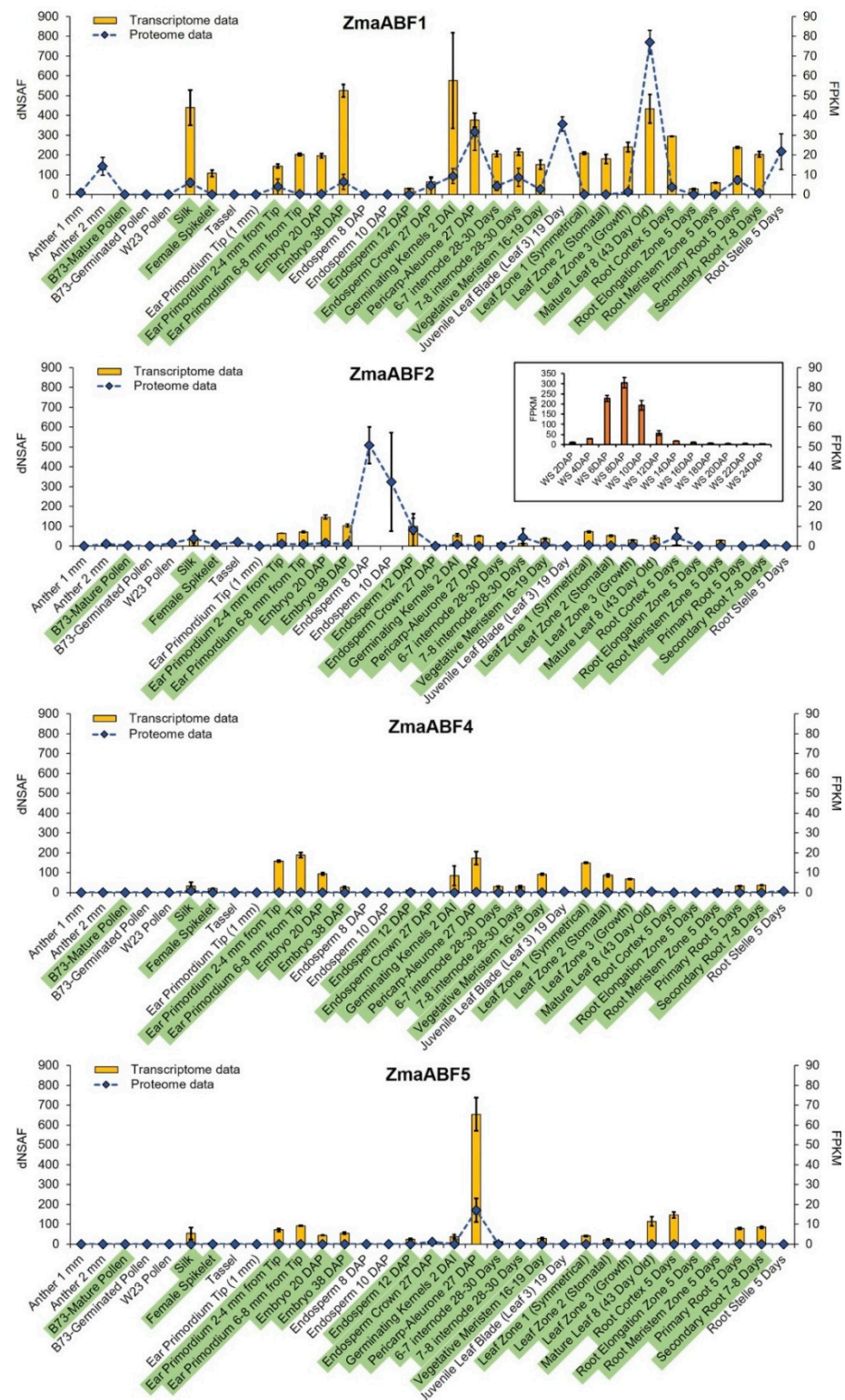


Figure 2. Expression levels of α -L-arabinofuranosidase genes of the GH51 family in various maize tissues according to Walley et al. [51] and Hoopes et al. [52]. Transcriptomic data are presented as FPKM (fragments per kilobase of transcript per million fragments mapped) values, proteomic data are presented as dNSAF (distributed normalized spectral abundance factor) values, both with standard deviations. The green background for some sample names marks those for which Walley and colleagues [51] obtained both transcriptomic and proteomic data, whereas for the remaining samples only proteomic data were obtained. WS, whole seed; DAP, days after pollination; DAI, days after imbibition.

The amount of the corresponding protein was high in anthers, seed pericarp and aleurone, in the juvenile and mature leaves, and in the root stele (Figure 2). The expression of the *ZmaABF2* gene was significantly lower in maize tissues compared with the *ZmaABF1* gene (Figure 2). However, transcript and protein levels were very high in the seed endosperm on days 8–10 after pollination (Figure 2, inset). The transcript level of the *ZmaABF4* gene was comparable to that of the *ZmaABF2* gene, but the corresponding protein product was detected only in silk and in the 5-day-old root stele at a very low level (Figure 2). The *ZmaABF5* gene also had low levels of mRNA in maize tissues, except for the seed pericarp and aleurone layer, where the corresponding protein was also detected at a significant level (Figure 2).

According to the expression data, *ZmaABF1* protein serves as the main isoform of α -L-arabinofuranosidases employed by maize in different tissues and at different developmental stages. *ZmaABF2* is required during the short period of endosperm development. *ZmaABF4* and *ZmaABF5* have moderate levels of transcripts across different tissues, but *ZmaABF5* protein has been detected only in the aleurone and pericarp. The fine regulation of the set of expressed α -L-arabinofuranosidases in some tissues or during different development processes indicates possible differences in their functions. Such functions must be structurally and evolutionarily determined.

3.3. Phylogeny of GH51 Proteins of Various Plant Species

We searched for sequences encoding GH51 proteins in genomes and transcriptomes of different plants to assess their variability during evolution. All sequences found were checked on the proper length and presence of necessary domains and motifs (Table S1). Such sequences are absent in the transcriptomes of red (*Rhodophyta*) and brown (*Phaeophyceae*) algae, and algae of *Chromista* and *Glaucophyta* phylum, deposited in the 1KP database. Thus, only the protein sequences of green plants were included in the resulting phylogenetic tree (Figure 3).

Sequences of GH51 are divided into three major clades: one of them includes *Chlorophyta* algae and the other two closely related clades include *Charophyta* algae and *Embryophyta* (land plants) representatives (Figure 3). The *Embryophyta* clade, in turn, is subdivided into several subclades, distinguishing hornworts, liverworts, mosses, ferns, lycophytes, gymnosperms, dicots, and monocotyledons (Figure 3). Most of the tree branches within the subclades are not very long, suggesting that the GH51 proteins of plants are highly conserved within the family.

The number of genes encoding GH51 proteins varies in different plant genomes. The genomes and transcriptomes of the *Chlorophyta* algae contain one or two genes encoding full-length GH51 proteins, while the *Charophyta* algae contain only one protein (Figure 3; Table S1). Land plants possess one or two sequences, except for the transcriptome of *Thuja plicata* (Figure 3; Table S1) and cereals, which have many isoforms. Basal monocotyledons (members of *Acorales*, *Alismatales*, and *Liliales* orders) possess only one or two genes encoding GH51 family members, whereas in other orders the number of isoforms has increased during evolution. The *Eleusine coracana* and *Oryza sativa* contain the greatest number of GH51 genes, nine and seven, respectively (Figure 4, Table S2).

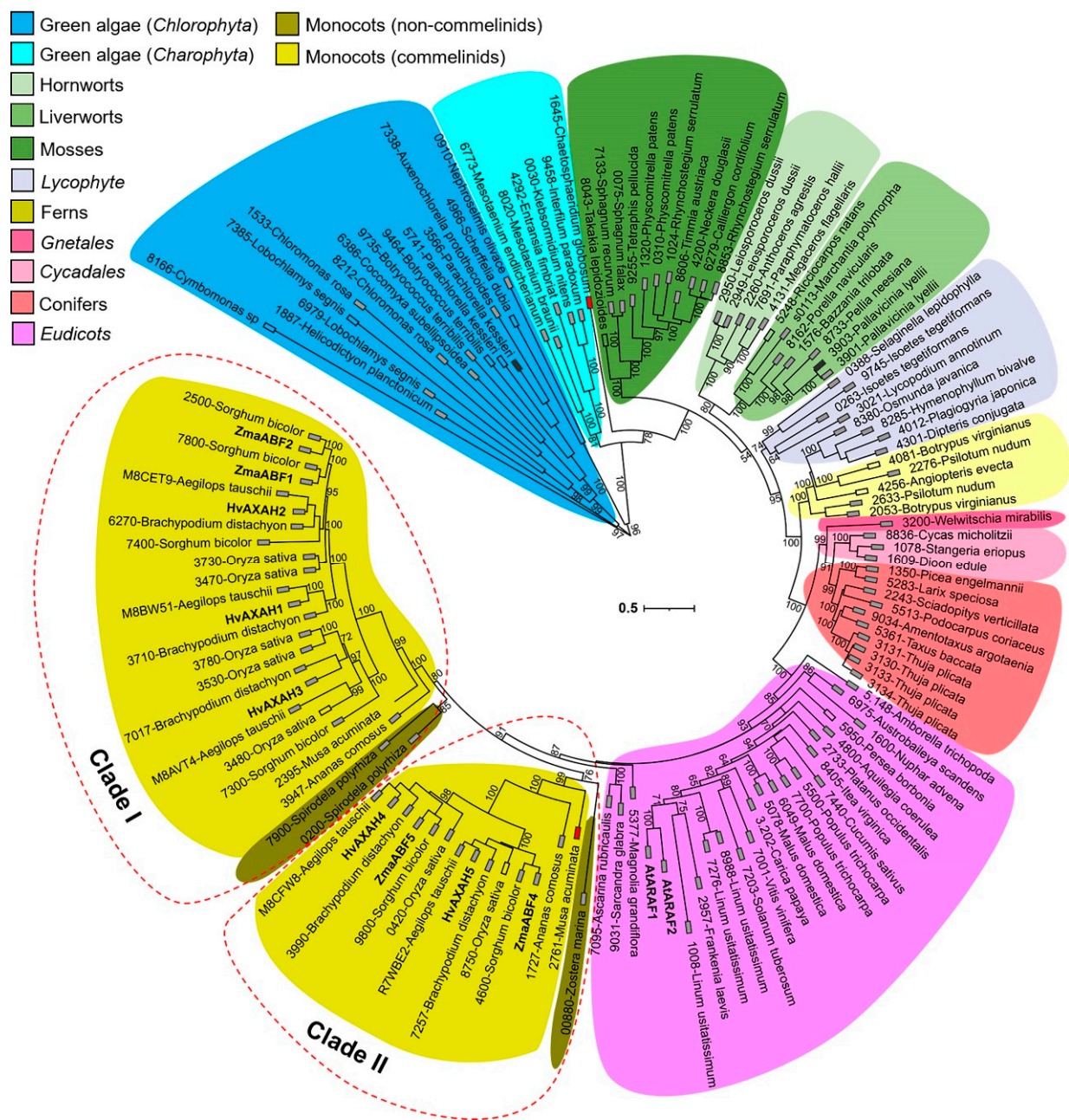
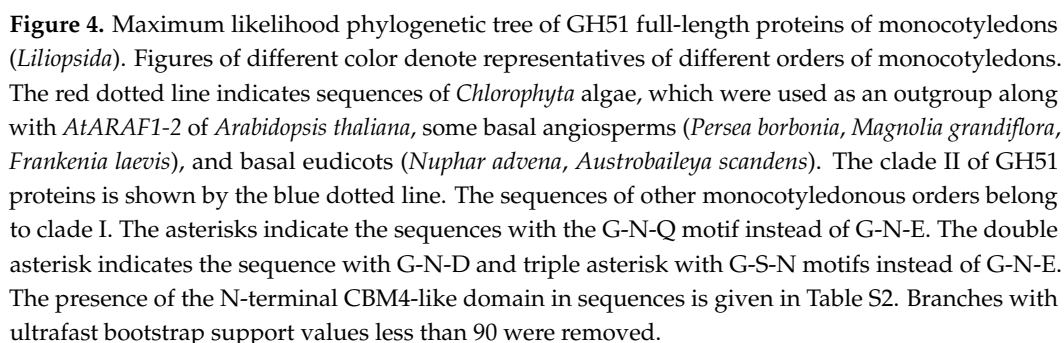


Figure 3. Maximum likelihood phylogenetic tree of plant full-length protein sequences of the GH51 family in circular representation. Characterized plant enzymes are highlighted in bold [14,34,68]. Red dotted lines surround two clades of monocotyledons. Colored rectangles next to the protein names show the presence of a third N-terminal CBM4-like domain in addition to α -L-arabinofuranosidase (Alpha-L-AF_C (PF06964)) and Glycoside_hydrolase_SF (IPR017853) domains. Among them, the gray rectangles show proteins with a CBM4-like domain. Black and red rectangles show sequences with CBM4-like domain also predicted in the InterProScan as galactose-bd-like_sf or CBM_4_9 (PF02018) domains, respectively. Blank rectangles indicate the absence of N-terminal CBM4-like domain. The architecture of the domains of each sequence as well as the full names of the proteins are given in Table S1. Numbers indicate the ultrafast bootstrap support values.



During evolution, the GH51 proteins of monocotyledons were divided into two large clades (Figure 3), both of which include sequences of both commelinids and non-commelinids. Commelinids include *Arecales*, *Zingiberales*, *Commelinales*, and *Poales* orders and are distinguished by the special composition of the primary cell wall and the presence of ferulic acids [69,70].

The clade II includes sequences with a substitution of acid/base glutamic acid (E) (underlined) for glutamine (Q) in the conserved G-N-E motif [71], while in clade I this substitution is absent (Figure 4). However, the *Zostera marina* (eelgrass) enzyme joined clade II despite the absence of acid/base substitution, and one enzyme of *Spirodela polyrhiza* (duckweed) with a substitution joined clade I (Figure 4). The phylogenetic analysis of a broader list of monocotyledons confirmed the division of *Liliopsida* GH51 proteins into clades I and II (Figure 4). Clade II includes enzymes of the *Poales*, *Asparagales*, and *Zingiberales* orders, and in almost all of them the catalytic acid/base E is replaced by Q.

3.4. N-Terminal Domain of GH51 α -L-arabinofuranosidases

Three out of four full-length maize GH51 sequences contained a domain at the N-terminus that was annotated as galactose-binding-like superfamily domain (Figure 1). Many other plant sequences used for phylogenetic analysis were also predicted to contain this domain or a carbohydrate binding module of the family 4 (CBM4) domain (PF02018) (Figure 3; Table S1). A homologous N-terminal domain was found in the MgGH51 enzyme of the fungus *M. giganteus*. It forms a ten-strand β -sandwich structure similar to the bacterial CBM4 domains [31] (Figure 5A,B, highlighted in purple). However, unlike the bacterial CBM4s that bind cellulose, laminarin, and xylan oligosaccharides [72–75], the N-terminal domain of MgGH51 was unable to bind saccharides and was named a CBM4-like domain [31]. We used this name in our study.

The N-terminal sequences of ZmaABF1 and AtARAF1, which are homologous to the CBM4-like domain of MgGH51, were used for BLASTP search in the GenBank database, excluding *Viridiplantae* (green plants). The search resulted in bacterial, fungal, and six insect (*Hexapoda*) sequences all belonging to the GH51 family indicating the specificity of the CBM4-like domain to that glycoside hydrolase family. The insect sequences have 30% identity with plant, 35–40% identity with fungal, and 41–48% identity with bacterial enzymes, so horizontal gene transfer is most likely.

We created the CBM4-like domain profile in HMMer-3.3 using multiple alignment of 16 fungal, 11 bacterial, and 111 plant sequences possessing the N-terminal domain sequence. The beginning and the end of the CBM4-like domain were established by L-x-V-D and S-x-x-P conserved motifs, respectively (Figure S1). The S-x-x-P motif is located on the loop separating the CBM4-like domain from the others in the MgGH51 structure [31]. The generated CBM4-like domain profile was represented in 119 of 133 (89.5%) GH51 plant sequences (Table S1) that were used for phylogenetic analysis earlier (Figure 3).

To estimate the proportion of GH51 sequences that contain a CBM4-like domain in bacteria, fungi, and insects, we downloaded all available protein sequences of these organisms that were longer than 200 residues and contained either the PF06964 domain or both the PF02018 and PF06964 domains from the Uniprot database. Among the 13584 bacterial and 2189 fungal GH51 protein sequences, 3362 (24.7%) and 1354 (61.8%), respectively, possessed a generated CBM4-like domain profile. All insect sequences (6 out of 6, 100%) contained a CBM4-like domain. A total of 866 of 1027 (84.3%) and 402 of 437 (91.9%) plant sequences with the PF06964 domain deposited in the Uniprot and Phytozome databases, respectively, possessed the CBM4-like domain. The superposition of the MgGH51 arabinofuranosidase carrying a CBM4-like domain and bacterial GH51 arabinofuranosidase (Tx-Abf) lacking this domain is shown in Figure 5A.

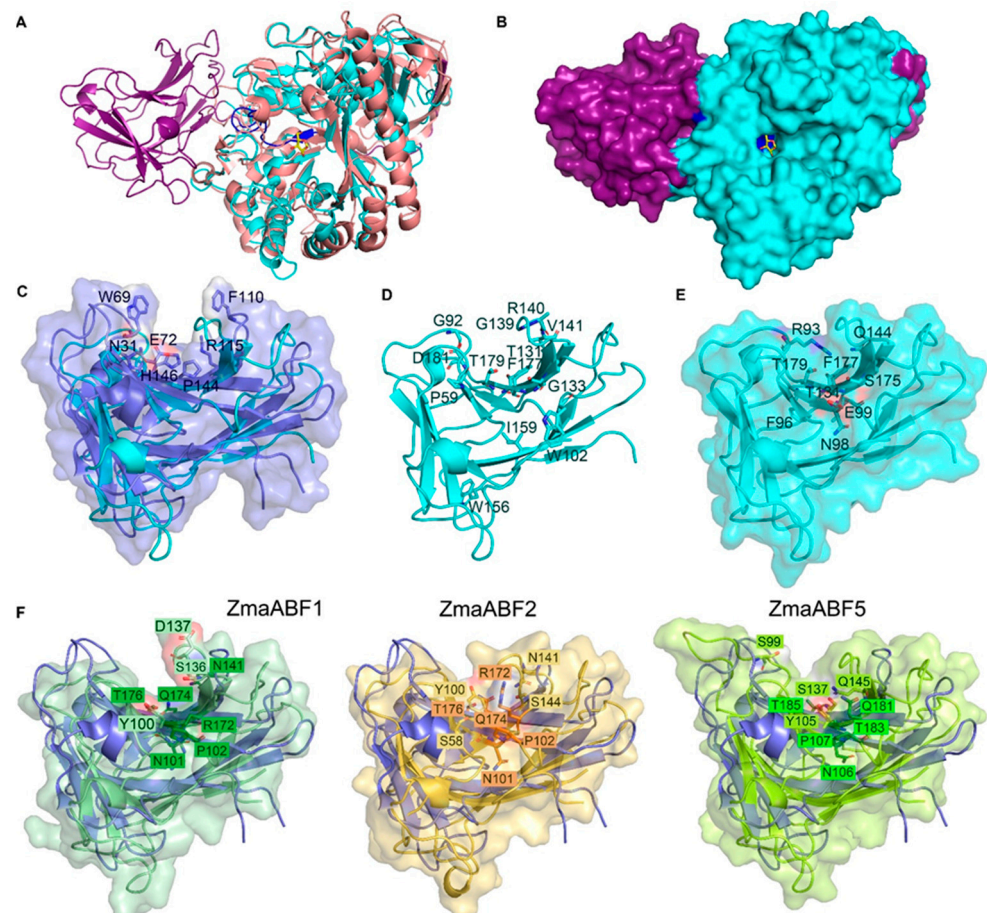


Figure 5. Spatial structures of bacterial and fungal arabinofuranosidases of GH51 family (A,B); bacterial CBM4 and fungal and plant CBM4-like domains (C–F). (A) The cartoon representation of superimposition of bacterial GH51 Tx-Abf (PDB ID: 2vrq, salmon) from *Thermobacillus xylanilyticus* and fungal MgGH51 (PDB ID: 6pzy, cyan) from *Meripillus giganteus*. The CBM4-like domain of MgGH51 (MgCBM4-like) is highlighted in purple. The conserved G-x-x-F-E-x-I-x-x-G-x-G-G motif is highlighted in blue. Arabinose co-crystallized with the MgGH51 structure is represented as sticks with yellow C atoms and red O atoms. (B) MgGH51 in the surface representation and arabinose in sticks representation colored in the same way as in (A). (C) Superimposed CBM4-2 from *Rhodotermus marinus* xylanase (RmCBM4-2; PDB ID: 1k42; dark purple) and CBM4-like domain from MgGH51 (blue). CBM4-2 from *R. marinus* is shown as a cartoon and surface. RmCBM4-2 amino acids involved in xylooligosaccharide recognition are shown as purple sticks with red O atoms, blue N atoms, and purple C atoms. (D) MgCBM4-like domain in cartoon representation. Amino acid residues of the MgCBM4-like domain homologous to the substrate-binding residues of bacterial CBMs of family 4 are shown as sticks with red O atoms, blue N atoms, and cyan C atoms. (E) MgCBM4-like domain in cartoon and surface representations. Amino acids with functional groups in close proximity to the putative substrate-binding cleft are shown as sticks with red O atoms, blue N atoms, and cyan C atoms. (F) Superimposed structures of RmCBM4-2 (cartoon representation; dark purple) and the CBM4-like domains of maize GH51 proteins (cartoon and surface representations of dark green, yellow-orange and lime colors for ZmaABF1, ZmaABF2, and ZmaABF5, respectively). Amino acid residues of maize CBM4-like domains homologous to (lighter labels) and those occupying the same spatial positions (darker labels) as the substrate-binding residues of bacterial CBMs of family 4 are shown as sticks.

Phylogenetic analysis of the 138 CBM4-like domain sequences used to create the domain profile showed that the sequences of fungi and land plants are the most distant from each other (Figure S2). The bacterial N-terminal domains alternate with those of the algae *Chloro-*

phyta and, probably, *Chlorophyta* acquired it during gene horizontal transfer. The N-terminal domain of the algae *Charophyta* occupies the same clade as *Lycophyte* (Figure S2).

To establish the differences between bacterial CBM4s and CBM4-like domains, we aligned their protein sequences. The sequences of bacterial CBMs of family 4 with resolved crystal structures [72–75] were compared with those of CBM4-like domains of some bacteria, plants, and fungi (Figure S3). None of the amino acid residues which were reported as crucial for recognition and accommodation of carbohydrate substrates in bacterial CBM4s ([72–75]; Figure S3, green background) were conserved in CBM4-like sequences tested (Figure S3). In contrast, the G-x-x-F-E-x-I-x-x-G-x-G-G motif which is conserved in CBM4-like domains (Figure S1, black frame) is absent or significantly altered in bacterial CBM4s (Figure S3, black frame). This motif is located in the β -strand that permeates the entire spatial structure of the MgGH51 enzyme and forms the bottom of its catalytic pocket (Figure 5A,B, highlighted in blue) [31].

The polar amino acid residues of bacterial CBMs of family 4 interact with the OH groups of sugars, and the aromatic radicals of tryptophan (W), tyrosine (Y), and phenylalanine (F) residues with the sugar rings through a stacking-like interaction [76]. These amino acids form a cleft-like structure in which oligosaccharides are placed, such as in *Rhodotermus marinus* xylanase (RmCBM4-2) [73] (Figure 5C). In the MgGH51 and ZmaABFs, most of the residues homologous to these substrate-binding amino acids of CBM4-like domain are aliphatic (Figure S3, Figure 5D). Moreover, the arginine (R93) of the MgCBM4-like surface is located on a loop next to the putative cleft-like pocket and crosses it, making the MgCBM4-like surface rather flat (Figure 5E) in contrast to resolved bacterial CBM4s (Figure 5C).

The CBM4-like domain of ZmaABF1 compared to the fungal CBM4-like demonstrates a cleft-like surface similar to that of RmCBM4-2 (Figure 5F); however, it is narrower than that in RmCBM4-2 and, probably, cannot accommodate the sugar. In the CBM4-like domains of ZmaABF2 and ZmaABF5, the cleft-like surface is almost absent as in the CBM4-like of MgGH51. The putative cleft is populated by residues with different functional groups (Figure 5F). Thus, according to alignments and modeling, neither the CBM4-like domain of MgGH51 nor the CBM4-like domains of maize GH51 proteins are capable of binding oligosaccharides in a manner similar to that displayed by bacterial CBM4s.

3.5. Three-Dimensional Models of GH51 Proteins and Molecular Docking

Three-dimensional models of ZmaABF1, ZmaABF2, and ZmaABF5, and additionally AtARAF1, were constructed using the crystal structure of the fungus MgGH51 as a template. ZmaABF4 was discarded because of the lack of significant accumulation of transcripts and corresponding proteins in maize tissues (Figure 2; [51]). The sequence identity between maize proteins, AtARAF1, and MgGH51 ranged from 32.5 (ZmaABF2) to 34.3% (ZmaABF5), which is sufficient for homologous modeling [77]. All models of maize proteins, AtARAF1 of *A. thaliana*, structures of fungal MgGH51, and bacterial Tx-Abf with replacement of catalytic acid/base E with Q and wild type enzyme were subjected to molecular docking with a number of substrates (Table 2). In silico prediction of the specificity of glycoside hydrolases to a particular substrate is effective [78], as has already been demonstrated, including for the α -L-arabinofuranosidases of GH51 [79]. However, the results of such modeling should be treated with caution, since they are derived from predicted protein structures. During homology modeling, the orientation of the arabinosyl residue in the active site was preserved as in the X-ray structure of the MgGH51-arabinose complex. Thus, a single permissible binding mode of the terminal arabinose was observed in all proteins; therefore, the difference in binding energies is mainly due to the interaction of the protein with the oligosaccharide backbone.

Table 2. The molecular docking results for plant, bacterial, and fungal GH51 enzymes. The wild type of Tx-Abf is marked by asterisk. Xylp, xylopyranose; Araf, arabinofuranose; Galp, galactopyranose; ND, not determined.

	The Docking Score ΔG (kcal mol ^{−1})		
	$\beta(1,4)$ -D-Xylp-[$\alpha(1,3)$ -L-Araf]- $\beta(1,4)$ -di-D-Xylp	tetra- $\alpha(1,5)$ -L-Araf	$\beta(1,6)$ -D-Galp-[$\alpha(1,3)$ -L-Araf]- $\beta(1,6)$ -D-di-Galp
Tx-Abf (2vrq)	−8.46	−9.23	−6.25
Tx-Abf * (2vrk)	−8.11	ND	ND
MgGH51 (6zpy)	−7.18	−4.96	ND
ZmaABF1	−6.57	−4.52	ND
ZmaABF2	−5.52	−1.83	−3.50
ZmaABF5	−6.21	−3.82	ND
AtARAF1	−7.00	−2.69	ND

Among the studied proteins, bacterial enzyme Tx-Abf displayed the highest affinity to the fragment of arabinan (Table 2). Both Tx-Abf (with E/Q mutation and wild type enzyme) and fungal MgGH51 also exhibited high affinity to the monosubstituted fragment of xylan; however, judging from slightly lower ΔG absolute value, arabinan was a less favorable ligand for MgGH51 than for Tx-Abf (Table 2). The arabinogalactan II fragment was the least favorable substrate for Tx-Abf among those tested (Table 2). The Tx-Abf is the xylan-utilizing enzyme as it was also shown biochemically [70].

Maize proteins had lower binding energies when interacting with monosubstituted xylan when compared with bacterial or fungal enzymes. Their interaction with arabinan was even less favorable (Table 2). Surprisingly, the model of AtARAF1 also displayed higher affinity to monosubstituted xylan than that to its known substrate [34], arabinan (Table 2). All attempts to simulate the interaction of plant GH51 models with disubstituted xylan fragments were unsuccessful.

In maize models, the conservative catalytic E342 and E419 formed the hydrogen bonds to the arabinose moiety of XAXX (Q349 and E426 in ZmaABF5) (Figure 6A,B). The E23, N341 (N348 in ZmaABF5), and Y392 (Y399 in ZmaABF5) also formed hydrogen bonds to arabinose and xylose (Figure 6B). The same amino acids were involved in the binding of arabinan fragments in maize models; however, in ZmaABF1, this substrate was additionally bound by H288 and E226 residues (Figure 6C) and in ZmaABF5 by N230 (data not shown). Probably, these additional hydrogen bonds from H288 and E226 provide a higher affinity of ZmaABF1 to arabinan compared to other maize GH51 models (Table 2).

The structure of the catalytic (β/α)₈-barrel domain in maize models is almost the same as that of MgGH51, except for ZmaABF2. The ZmaABF2 protein sequence lacks a fragment of 19 residues which forms a β -strand in the catalytic (β/α)₈-barrel domain of ZmaABF1, ZmaABF5, and MgGH51 (Figure 6D). In MgGH51, this β -strand is involved in the formation of a flat surface of a xylan-binding catalytic groove with a deep pocket for arabinose (Figure 5B). This surface is organized similarly in ZmaABF1 and ZmaABF5 (Figure 6A). The absence of the 19-residue fragment makes that surface of ZmaABF2 shallow and more heterogeneous than that in MgGH51 (Figure 6A). The putative arabinose-binding pocket in ZmaABF2 is looser and not as deep as in ZmaABF1 and ZmaABF5 (Figure 6A). Nevertheless, according to molecular docking results, both the narrow catalytic grooves of ZmaABF1 and ZmaABF5 and the wide groove of ZmaABF2 are unable to accept the xylooligosaccharide di-substituted with arabinose (data not shown).

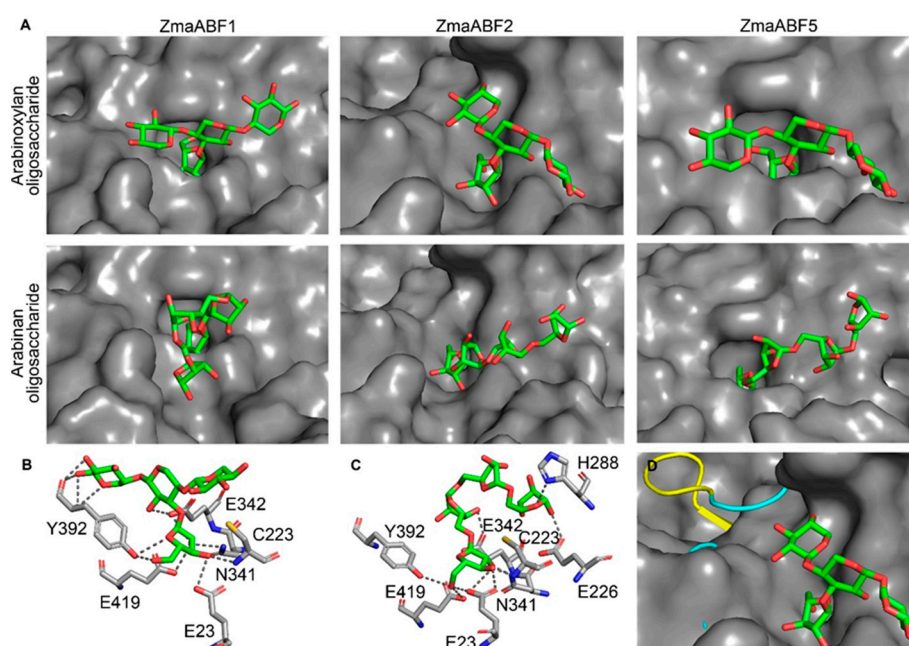


Figure 6. Three-dimensional models of ZmaABF1, ZmaABF2, and ZmaABF5 of the GH51 family and molecular docking results. **(A)** The spatial models of ZmaABF1, ZmaABF2, and ZmaABF5 in surface representation with the docked $\beta(1,4)$ -D-xylopyranose- $[\alpha(1,3)$ -L-arabinofuranose]-di- $\beta(1,4)$ -D-xylopyranoside or tetra- $\alpha(1,3)$ -L-arabinofuranoside, shown as sticks, where C atoms are green and O atoms are red. **(B)** The amino acid residues of ZmaABF1 that form the hydrogen bonds with arabinoxylan and arabinan oligosaccharides are shown as sticks with red O atoms, blue N atoms, yellow S atoms. C atoms are gray in amino acids and green in sugars. The bonds are shown by gray dotted lines. **(C,D)** The superimposition of ZmaABF2 (in surface mode) on the ZmaABF1 (in cartoon mode), where 19-residue insertion is highlighted in yellow. The docked arabinoxylan oligosaccharide is shown as sticks, where C atoms are green and O atoms are red.

4. Discussion

The five maize genes encoding GH51 proteins were identified in the latest version of the maize genome (Table 1) compared to the eight identified previously [14]. The gene identification results depend on the genome assembly quality. The identification and analysis of maize GH51 genes are useful for further studies of these plant enzymes, which may include the construction of recombinant proteins and other techniques of molecular biology.

During early plant evolution, α -L-arabinofuranosidases of the GH51 family probably appeared in the genomes of the green algae *Chlorophyta*, since we did not find genes for these proteins in red, brown algae, and other basal eukaryotes in used databases. However, the GH51 proteins are also presented in fungi and some insects. Despite the absence of GH51 proteins, arabinose is present in the cell walls of brown algae, for example, as part of AGPs arabinogalactans [80], however, arabinose is a minor monosaccharide in the cell walls of red algae compared to brown and green algae [81]. In contrast, in *Chlorophyta* algae, arabinose is the major terminal-linked monosaccharide of carbohydrates [82]. Polymers that form the shell structure of the green alga *Botryococcus* have arabinose and galactose in significant quantities [83]. Arabinose is a part of arabinomannan [84], a unique type of arabinogalactan [85], and glucuronan with arabinose-containing side chains [86] and N-glycosylation chains in *Chlorella* [87]. The increase in the amount of arabinose in various polymers in *Chlorophyta* could form the demand for new α -L-arabinofuranosidases and result in the emergence of the GH51 family. However, the apparent absence of corresponding genes in other algae and basal eukaryotes may be the result of scarce genomic data for them compared to *Viridiplantae*.

Since their appearance in *Chlorophyta*, GH51 α -L-arabinofuranosidases have moderately changed during plant evolution (Figure 3). A few whole-genome duplication events that occurred in monocotyledonous plants of the order *Poales* [88] probably lead to an increase in the number of GH51 genes in cereals (Figures 3 and 4; Tables S1 and S2). For example, barley has five genes encoding GH51 proteins, sorghum has six, and rice has seven (Figures 3 and 4; Tables S1 and S2). The increased number of putative α -L-arabinofuranosidases in cereals, along with a large number of α -L-arabinosyl- and β -D-xylosyltransferases [89,90], is probably necessary for the synthesis and modification of arabinoxylans with a complex side chain pattern [91]. The idea of the special development of GH51 in species possessing complex xylans in cell walls is supported by the fact that the number of genes encoding GH51 is also increased in the majority of analyzed representatives of the order *Asparagales* (Figure 4). The *Asparagales* are non-commelinid, i.e., their primary cell walls are of the so-called type I and are abundant in pectins and xyloglucans, but not in arabinoxylans and mixed-linkage glucans characteristic of type II cell walls common to commelinids [69]. Despite this, *Asparagales* have cell wall xylans containing a terminal 1,2-linked arabinopyranoside as a side substituent in addition to the typical glucuronic acid [92]. Thus, the increased amount of α -L-arabinofuranosidases in *Asparagales* may also be directed toward cell wall xylan modification.

A progressive step in GH51 protein evolution seems to have been the division of monocot proteins into two large clades I and II (Figures 3 and 4). Clade II includes sequences of only three orders (*Asparagales*, *Zingiberales*, and *Poales*) of monocotyledons (Figure 4). The absence of clade II sequences in genomes of other monocot orders indicates that this clade probably could have arisen during the whole-genome duplication of the *Asparagales*, *Zingiberales*, and *Poales* which took place approx. 66 million years ago [88]. These orders separated from each other about 120 million years ago [93]. This means that their molecular evolution had long been independent by the time of the genome duplication 66 million years ago. In this context, it is particularly interesting that most of the representatives of clade II carry the catalytic acid substitution (Figure 4). A similar replacement of acid/base glutamic acid with glutamine in Tx-Abf resulted in a more than 100-fold decrease in specific activity compared to native enzyme in vitro [71]. According to our docking study, this decrease in activity should not be due to a decrease in affinity to the substrate, since the indicated replacement had almost no effect on it (Table 2). Some enzymes, such as the GH1 myrosinases or GH92 α -mannosidases, can function efficiently in the absence of an acid/base catalyst [94,95], probably by using the substrate's own acid catalyst [96]. The possibility of using a similar mechanism by GH51 enzymes from clade II monocotyledons should be studied separately. Another possibility is that the catalytic site of clade II GH51 is indeed disrupted, which has already been reported for some glycoside hydrolases [97]. The significance of acid/base substitution in plant GH51 proteins requires further studies.

The enzymes of clade I seem to be more important and more active compared to clade II members. Barley HvAXAH1-2 enzymes belonging to clade I (Figure 3) showed high catalytic rates in vitro, and the corresponding genes were expressed at high levels in many tissues [68], as were maize GH51 genes (*ZmaABF1*; *ZmaABF2*) from clade I (Figures 2 and 3). The *ZmaABF1* protein and corresponding transcripts were found in almost all tissues analyzed by Walley and colleagues [51]. Protein accumulation was high in mature leaves, seed pericarp, and the root stele, but not in the meristem and elongation growth zones of root (Figure 2) [51]. This is consistent with the recently described co-expression of *ZmaABF1* with the genes of secondary cell wall cellulose synthases in maize root [35], since mature leaves, seed pericarp, and root stele also develop secondary cell walls. The modeled maize *ZmaABF1* displayed the highest affinity to both arabinoxylan and arabinan fragments among other maize enzymes studied, comparable to estimates for MgGH51 (Table 2), suggesting that *ZmaABF1* is efficiently adapted to handle both these substrates. At the same time, the catalytic site of *ZmaABF1* does not seem to be able to accommodate the arabinose of disubstitution of the arabinoxylan backbone. Thus, *ZmaABF1* appears to serve

as the main isoform of GH51 α -L-arabinofuranosidases employed by maize in different tissues, especially those developing secondary cell walls. Monosubstituted xylan moieties probably serve the main target for this enzyme; however, its affinity to α -L(1,5)-arabinan is also significant.

ZmaABF2 appears to be a seed-specific enzyme. The *ZmaABF2* gene is expressed at low levels in most maize tissues [51,52], with the exception of seeds at stage 6–10 DAP (Figure 2) [52], where a peak of expression was found in the endosperm at 6 DAP [98]. ZmaABF2 protein accumulation also peaked only in the endosperm at 8–10 DAP (Figure 2) [51]. During 3–12 DAP, the maize endosperm is cellularized, suggesting intense cell division and formation of the aleurone layer (6–10 DAP) [99]. Cellularization is accompanied by the deposition of cell wall components as observed in barley seeds at the 6 DAP stage [100]. The modeled ZmaABF2 exhibited higher affinity for fragments of arabinoxylan and arabinogalactan than that for arabinan (Table 2), indicating that the latter may be the least favorable substrate. The absence of a predicted signal peptide in ZmaABF2 (Figure 1) may indicate that this enzyme accumulates first within some seed cells separately from its primary substrates and then acts on them only at the stage of seed germination. ZmaABF2 differs from other maize GH51 proteins by the absence of a 19-amino acid fragment in the protein sequence (Figure S4), which led to the formation of an additional hole near the catalytic pocket; however, it did not affect the catalytic pocket itself (Figure 6D). Nevertheless, ZmaABF2 with changed topology of surface is unable to bind the di-substituted arabinose fragment of xylan in silico (data not shown). This deletion of 19 residues is not present in ZmaABF2 protein sequences in earlier assemblies of the maize genome, so it is difficult to characterize this deletion as unique to ZmaABF2 among other maize enzymes. In general, among GH51 α -L-arabinofuranosidases, only a few, mostly fungal, can detach arabinose from di-substituted xylans [101]. Plant enzymes, if they are capable of mediating this reaction, catalyze it at a low rate [21] and, apparently, these enzymes are absent among maize GH51 proteins.

The barley clade II members, HvAXAH3-5 (Figure 3), showed negligible in vitro catalytic rates against different arabinose-containing polysaccharides and low expression levels of the corresponding genes [68]. Transcription of maize clade II members (*ZmaABF4* and *ZmaABF5* genes) was also low, and accumulation of the corresponding protein was negligible, except for significant accumulation of the ZmaABF5 protein in the seed pericarp (Figure 2; [51]). The acid/base mutation did not occur long ago by evolutionary standards, but it has been preserved in the genomes of various monocotyledons for millions of years. The GH51 proteins of clade II may have minor functions in cereals or require specific conditions for catalysis, or they may play a role as reserve enzymes in cereals. Another possibility is that, having lost their hydrolytic function, these proteins have acquired some other function [97,102]. Anyway, the physiological significance of the mutated GH51 proteins requires additional studies.

The third N-terminal CBM4-like domain was found only in sequences of plant, fungal, and bacterial GH51 proteins and it is probably unique to the enzymes of this family. This domain is present in 84–92% (depending on the search database) of analyzed plant GH51 sequences (Table S1), but only in 62% and 25% of fungal and bacterial sequences, respectively. The CBM4-like domain of plant GH51 family proteins appeared in green (*Chlorophyta*) ancient algae during evolution and passed to *Charophyta* algae and land plants (Figure 3; Table S1). The sequences of CBM4-like domains of plants and fungi are the most remote from each other according to phylogenetic analysis results (Figure S2). The bacterial CBM4-like domains occupy an intermediate position between fungi and plants; however, the domain of analyzed *Chlorophyta* algae is closely related to some bacteria (Figure S2).

The evolution of CBM4-like domains and whole GH51 sequences in land plants is generally similar. CBM4-like domains of angiosperms are distant from gymnosperms, which, in turn, are distant from ferns and parts of *Lycophyte* with significant values of ultrafast bootstrap support (Figure S2), as in the case of whole sequences (Figure 3). CBM4-like domains of liverworts and hornworts are more basal (Figure S2), although in the

case of whole GH51 sequences, mosses share this position with them (Figure 3). CBM4-like domains of mosses, parts of *Lycophyte*, and the alga *Charophyta* are interspersed with each other with low branch support values, indicating minor differences in the CBM4-like domain sequences in these groups (Figure S2). CBM4-like domains of *Charophyta* algae have evolved considerably compared to *Chlorophyta*. This coincides with the theory suggesting the *Charophyta* are the closest ancestors of land plants [44,103]. The topologies of phylogenetic trees constructed separately for CBM4-like domains (Figure S2) and for full-length GH51 proteins of monocotyledons (Figure 3) are similar (data not shown). However, the clade II is more basal compared to clade I (data not shown) in tree for CBM4-like domains (Figure S2) in contrast to whole-sequence phylogenetic analysis (Figure 4).

The various CBMs in the glycoside hydrolases facilitate substrate recognition and increase affinity, especially for insoluble surfaces such as cellulose, in microbial enzymes. The replacement of native CBMs to another one is able to change the substrate specificity of glycoside hydrolase [104]. The CBMs of family 4 belong to the type B of CBMs, which recognize the glycan chains up to six sugar moieties. The binding cleft of CBMs of this type can be either deep or shallow [76]. The aromatic and charged amino acid residues are responsible for ligand binding and orientation [76]. The surface of the CBM4-like domain in the α -L-arabinofuranosidase of MgGH51 and maize GH51 proteins is shallow and devoid of necessary residues (Figure 5C–F). The CBM4-like domain in these proteins is probably incapable of binding saccharides. However, this domain was retained in the GH51 sequences of plants and evolved with them. Thus, the presence of the CBM4-like domain is probably important for plant GH51 proteins, but the understanding of its function requires additional biochemical studies.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/life13020266/s1>, Figure S1: Multiple sequence alignment of CBM4-like domains of GH51 proteins of different origin.; Figure S2: Maximum likelihood phylogenetic tree of the N-terminal CBM4-like domain of fungal, bacterial and plant origin; Figure S3: Multiple sequence alignment of CBM4-like domains of bacterial, fungal and plant origin with those bacterial CBM4 domain sequences which crystal structures were resolved; Figure S4: Multiple sequence alignment of the protein sequences of maize GH51 proteins and AtARAF1 of *Arabidopsis thaliana* with the sequences of bacterial Tx-Abf (PDB ID: 2vrq) from *Thermobacillus xylanilyticus* and fungal MgGH51 (6zpy) from *Meripillus giganteus*. Table S1: Protein sequences used for phylogenetic analysis of the GH51 proteins, creation of the CBM4-like domain profile, and phylogenetic analysis of the CBM4-like domain. Domain organization of plant GH51 proteins. Table S2: Protein sequences used for phylogenetic analysis of the GH51 proteins of monocotyledons.

Author Contributions: Conceptualization, A.N. and L.K.; writing—original draft preparation, A.N. and L.K.; writing—review and editing, A.N., O.M. and L.K.; formal analysis, A.N. and O.M.; visualization, A.N. and O.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the RUSSIAN SCIENCE FOUNDATION, project number 18-14-00168 (L.K., A.N.), and by financial support from the government assignment for the FRC Kazan Scientific Center of RAS (O.M.).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Drula, E.; Garron, M.-L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The Carbohydrate-Active Enzyme Database: Functions and Literature. *Nucleic Acids Res.* **2022**, *50*, D571–D577. [CrossRef]
2. Lagaert, S.; Pollet, A.; Courtin, C.M.; Volckaert, G. β -Xylosidases and α -L-Arabinofuranosidases: Accessory Enzymes for Arabinoxylan Degradation. *Biotechnol. Adv.* **2014**, *32*, 316–332. [CrossRef]

3. Mechelke, M.; Koeck, D.E.; Broecker, J.; Roessler, B.; Krabichler, F.; Schwarz, W.H.; Zverlov, V.V.; Liebl, W. Characterization of the Arabinoxylan-Degrading Machinery of the Thermophilic Bacterium *Herbinix hemicellulosilytica*—Six New Xylanases, Three Arabinofuranosidases and One Xylosidase. *J. Biotechnol.* **2017**, *257*, 122–130. [\[CrossRef\]](#)
4. Poria, V.; Saini, J.K.; Singh, S.; Nain, L.; Kuhad, R.C. Arabinofuranosidases: Characteristics, Microbial Production, and Potential in Waste Valorization and Industrial Applications. *Bioresour. Technol.* **2020**, *304*, 123019. [\[CrossRef\]](#)
5. Minic, Z.; Do, C.T.; Rihoey, C.; Morin, H.; Lerouge, P.; Jouanin, L. Purification, Functional Characterization, Cloning, and Identification of Mutants of a Seed-Specific Arabinan Hydrolase in *Arabidopsis*. *J. Exp. Bot.* **2006**, *57*, 2339–2351. [\[CrossRef\]](#)
6. Tateishi, A.; Mori, H.; Watari, J.; Nagashima, K.; Yamaki, S.; Inoue, H. Isolation, Characterization, and Cloning of α -L-Arabinofuranosidase Expressed during Fruit Ripening of Japanese Pear. *Plant Physiol.* **2005**, *138*, 1653–1664. [\[CrossRef\]](#)
7. Brummell, D.A. Cell Wall Disassembly in Ripening Fruit. *Funct. Plant Biol.* **2006**, *33*, 103–119. [\[CrossRef\]](#)
8. Brummell, D.A.; Dal Cin, V.; Crisosto, C.H.; Labavitch, J.M. Cell Wall Metabolism during Maturation, Ripening and Senescence of Peach Fruit. *J. Exp. Bot.* **2004**, *55*, 2029–2039. [\[CrossRef\]](#)
9. Nobile, P.M.; Wattedled, F.; Quecini, V.; Girardi, C.L.; Lormeau, M.; Laurens, F. Identification of a Novel α -L-Arabinofuranosidase Gene Associated with Mealiness in Apple. *J. Exp. Bot.* **2011**, *62*, 4309–4321. [\[CrossRef\]](#)
10. Kotake, T.; Tsuchiya, K.; Aohara, T.; Konishi, T.; Kaneko, S.; Igarashi, K.; Samejima, M.; Tsumuraya, Y. An α -L-Arabinofuranosidase/ β -D-Xylosidase from Immature Seeds of Radish (*Raphanus sativus* L.). *J. Exp. Bot.* **2006**, *57*, 2353–2362. [\[CrossRef\]](#)
11. Arsovski, A.A.; Popma, T.M.; Haughn, G.W.; Carpita, N.C.; McCann, M.C.; Western, T.L. AtBXL1 Encodes a Bifunctional Beta-D-Xylosidase/ α -L-Arabinofuranosidase Required for Pectic Arabinan Modification in *Arabidopsis* Mucilage Secretory Cells. *Plant Physiol.* **2009**, *150*, 1219–1234. [\[CrossRef\]](#)
12. Obel, N.; Porchia, A.C.; Scheller, H.V. Dynamic Changes in Cell Wall Polysaccharides during Wheat Seedling Development. *Phytochemistry* **2002**, *60*, 603–610. [\[CrossRef\]](#)
13. Suzuki, M.; Kato, A.; Nagata, N.; Komeda, Y. A Xylanase, AtXyn1, Is Predominantly Expressed in Vascular Bundles, and Four Putative Xylanase Genes Were Identified in the *Arabidopsis thaliana* Genome. *Plant Cell Physiol.* **2002**, *43*, 759–767. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Kozlova, L.V.; Gorshkov, O.V.; Mokshina, N.E.; Gorshkova, T.A. Differential Expression of α -L-Arabinofuranosidases during Maize (*Zea mays* L.) Root Elongation. *Planta* **2015**, *241*, 1159–1172. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Carpita, N.C. Structure and Biogenesis of the Cell Walls of Grasses. *Annu. Rev. Plant Biol.* **1996**, *47*, 445–476. [\[CrossRef\]](#)
16. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes Database (CAZy): An Expert Resource for Glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238. [\[CrossRef\]](#)
17. Minic, Z.; Rihoey, C.; Do, C.T.; Lerouge, P.; Jouanin, L. Purification and Characterization of Enzymes Exhibiting Beta-D-Xylosidase Activities in Stem Tissues of *Arabidopsis*. *Plant Physiol.* **2004**, *135*, 867–878. [\[CrossRef\]](#)
18. Tateishi, A.; Kamiyoshihara, Y.; Matsuno, J.; Miyohashi, F.; Shiba, H.; Kanayama, Y.; Watanabe, K.; Nomura, K.; Inoue, H. Heterologous Expression of Tomato Glycoside Hydrolase Family 3 α -L-arabinofuranosidase/ β -xylosidases in Tobacco Suspension Cultured Cells and Synergic Action of a Family 51 Isozyme under Antisense Suppression of the Enzyme. *Physiol. Plant.* **2014**, *150*, 238–251. [\[CrossRef\]](#)
19. Lee, R.C.; Hrmova, M.; Burton, R.A.; Lahnstein, J.; Fincher, G.B. Bifunctional Family 3 Glycoside Hydrolases from Barley with α -L-Arabinofuranosidase and Beta-D-Xylosidase Activity—Characterization, Primary Structures, and COOH-Terminal Processing. *J. Biol. Chem.* **2003**, *278*, 5377–5387. [\[CrossRef\]](#)
20. Ferré, H.; Broberg, A.; Duus Jens Ø and Thomsen, K.K.; Duus, J.Ø.; Thomsen, K.K. A Novel Type of Arabinoxylan Arabinofuranohydrolase Isolated from Germinated Barley: Analysis of Substrate Preference and Specificity by Nano-probe NMR. *Eur. J. Biochem.* **2000**, *267*, 6633–6641. [\[CrossRef\]](#)
21. Lee, R.C.; Burton, R.A.; Hrmova, M.; Fincher, G.B. Barley Arabinoxylan Arabinofuranohydrolases: Purification, Characterization and Determination of Primary Structures from cDNA Clones. *Biochem. J.* **2001**, *356*, 181–189. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Eckert, K.; Schneider, E. A Thermoacidophilic Endoglucanase (CelB) from *Alicyclobacillus acidocaldarius* Displays High Sequence Similarity to Arabinofuranosidases Belonging to Family 51 of Glycoside Hydrolases. *Eur. J. Biochem.* **2003**, *270*, 3593–3602. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Grant, S.; Sorokin, D.Y.; Grant, W.D.; Jones, B.E.; Heaphy, S. A Phylogenetic Analysis of Wadi El Natrun Soda Lake Cellulase Enrichment Cultures and Identification of Cellulase Genes from These Cultures. *Extremophiles* **2004**, *8*, 421–429. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Boyce, A.; Walsh, G. Characterisation of a Novel Thermostable Endoglucanase from *Alicyclobacillus vulcanalis* of Potential Application in Bioethanol Production. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 7515–7525. [\[CrossRef\]](#)
25. Hövel, K.; Shallom, D.; Niefind, K.; Belakhov, V.; Shoham, G.; Baasov, T.; Shoham, Y.; Schomburg, D. Crystal Structure and Snapshots along the Reaction Pathway of a Family 51 α -L-Arabinofuranosidase. *EMBO J.* **2003**, *22*, 4922–4932. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Taylor, E.J.; Smith, N.L.; Turkenburg, J.P.; D'souza, S.; Gilbert, H.J.; Davies, G.J. Structural Insight into the Ligand Specificity of a Thermostable Family 51 Arabinofuranosidase, Ara f 51, from *Clostridium thermocellum*. *Biochem. J.* **2006**, *395*, 31–37. [\[CrossRef\]](#) [\[PubMed\]](#)

27. Paës, G.; Skov, L.K.; O'Donohue, M.J.; Rémond, C.; Kastrup, J.S.; Gajhede, M.; Mirza, O. The Structure of the Complex between a Branched Pentasaccharide and *Thermobacillus xylanilyticus* GH-51 Arabinofuranosidase Reveals Xylan-Binding Determinants and Induced Fit. *Biochemistry* **2008**, *47*, 7441–7451. [\[CrossRef\]](#)
28. Souza, T.A.C.B.; Santos, C.R.; Souza, A.R.; Oldiges, D.P.; Ruller, R.; Prade, R.A.; Squina, F.M.; Murakami, M.T. Structure of a Novel Thermostable GH51 α -L-arabinofuranosidase from *Thermotoga petrophila* RKU-1. *Protein Sci.* **2011**, *20*, 1632–1637. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Im, D.-H.; Kimura, K.; Hayasaka, F.; Tanaka, T.; Noguchi, M.; Kobayashi, A.; Shoda, S.; Miyazaki, K.; Wakagi, T.; Fushinobu, S. Crystal Structures of Glycoside Hydrolase Family 51 α -L-Arabinofuranosidase from *Thermotoga maritima*. *Biosci. Biotechnol. Biochem.* **2012**, *76*, 423–428. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Dos Santos, C.R.; de Giuseppe, P.O.; de Souza, F.H.M.; Zanphorlin, L.M.; Domingues, M.N.; Pirolla, R.A.S.; Honorato, R.V.; Tonoli, C.C.C.; de Moraes, M.A.B.; de Matos Martins, V.P. The Mechanism by Which a Distinguishing Arabinofuranosidase Can Cope with Internal Di-Substitutions in Arabinoxylans. *Biotechnol. Biofuels* **2018**, *11*, 223. [\[CrossRef\]](#)
31. McGregor, N.G.S.; Turkenburg, J.P.; Mørkeberg Krogh, K.B.R.; Nielsen, J.E.; Artola, M.; Stubbs, K.A.; Overkleeft, H.S.; Davies, G.J. Structure of a GH51 α -L-Arabinofuranosidase from *Meripilus giganteus*: Conserved Substrate Recognition from Bacteria to Fungi. *Acta Crystallogr. Sect. D Struct. Biol.* **2020**, *76*, 1124–1133. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Sumiyoshi, M.; Nakamura, A.; Nakamura, H.; Hakata, M.; Ichikawa, H.; Hirochika, H.; Ishii, T.; Satoh, S.; Iwai, H. Increase in Cellulose Accumulation and Improvement of Saccharification by Overexpression of Arabinofuranosidase in Rice. *PLoS ONE* **2013**, *8*, e78269. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Fulton, L.M.; Cobbett, C.S. Two Alpha-L-Arabinofuranosidase Genes in *Arabidopsis thaliana* Are Differentially Expressed during Vegetative Growth and Flower Development. *J. Exp. Bot.* **2003**, *54*, 2467–2477. [\[CrossRef\]](#)
34. Chávez Montes, R.A.; Ranocha, P.; Martinez, Y.; Minic, Z.; Jouanin, L.; Marquis, M.; Saulnier, L.; Fulton, L.M.; Cobbett, C.S.; Bitton, F.; et al. Cell Wall Modifications in *Arabidopsis* Plants with Altered α -L-Arabinofuranosidase Activity. *Plant Physiol.* **2008**, *147*, 63–77. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Nazipova, A.; Gorshkov, O.; Eneyskaya, E.; Petrova, N.; Kulminskaya, A.; Gorshkova, T.; Kozlova, L. Forgotten Actors: Glycoside Hydrolases during Elongation Growth of Maize Primary Root. *Front. Plant Sci.* **2021**, *12*, 802424. [\[CrossRef\]](#)
36. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N. Phytozome: A Comparative Platform for Green Plant Genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [\[CrossRef\]](#)
37. Leebens-Mack, J.H.; Barker, M.S.; Carpenter, E.J.; Deyholos, M.K.; Gitzendanner, M.A.; Graham, S.W.; Grosse, I.; Li, Z.; Melkonian, M.; Mirarab, S.; et al. One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants. *Nature* **2019**, *574*, 679–685. [\[CrossRef\]](#)
38. Eddy, S.R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7*, e1002195. [\[CrossRef\]](#)
39. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240. [\[CrossRef\]](#)
40. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R. CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [\[CrossRef\]](#)
41. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J. Pfam: The Protein Families Database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Bolser, D.; Staines, D.M.; Pritchard, E.; Kersey, P. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. In *Plant Bioinforma. Methods and Protocols*, 2nd ed.; Humana Press: New York, NY, USA, 2016; Volume 1374, pp. 115–140. [\[CrossRef\]](#)
43. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W.; Pruitt, K.D.; Sayers, E.W.; Lipman, D.J.; et al. GenBank. *Nucleic Acids Res.* **2012**, *41*, D37. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Hori, K.; Maruyama, F.; Fujisawa, T.; Togashi, T.; Yamamoto, N.; Seo, M.; Sato, S.; Yamada, T.; Mori, H.; Tajima, N. *Klebsormidium flaccidum* Genome Reveals Primary Factors for Plant Terrestrial Adaptation. *Nat. Commun.* **2014**, *5*, 3978. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Madeira, F.; Pearce, M.; Tivey, A.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and Sequence Analysis Tools Services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50*, W276–W279. [\[CrossRef\]](#)
46. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: *Ab Initio* Prediction of Alternative Transcripts. *Nucleic Acids Res.* **2006**, *34*, W435–W439. [\[CrossRef\]](#)
47. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D. Automatic Annotation of Eukaryotic Genes, Pseudogenes and Promoters. *Genome Biol.* **2006**, *7*, S10. [\[CrossRef\]](#)
48. Woodhouse, M.R.; Cannon, E.K.; Portwood, J.L.; Harper, L.C.; Gardiner, J.M.; Schaeffer, M.L.; Andorf, C.M. A Pan-Genomic Approach to Genome Databases Using Maize as a Model System. *BMC Plant Biol.* **2021**, *21*, 385. [\[CrossRef\]](#)
49. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and Applications. *BMC Bioinformatics* **2009**, *10*, 421. [\[CrossRef\]](#)
50. Teufel, F.; Almagro Armenteros, J.J.; Johansen, A.R.; Gíslason, M.H.; Pihl, S.I.; Tsirigos, K.D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 Predicts All Five Types of Signal Peptides Using Protein Language Models. *Nat. Biotechnol.* **2022**, *40*, 1023–1025. [\[CrossRef\]](#)

51. Walley, J.W.; Sartor, R.C.; Shen, Z.; Schmitz, R.J.; Wu, K.J.; Urich, M.A.; Nery, J.R.; Smith, L.G.; Schnable, J.C.; Ecker, J.R.; et al. Integration of Omic Networks in a Developmental Atlas of Maize. *Science* **2016**, *353*, 814–818. [CrossRef]
52. Hoopes, G.M.; Hamilton, J.P.; Wood, J.C.; Esteban, E.; Pasha, A.; Vaillancourt, B.; Provart, N.J.; Buell, C.R. An Updated Gene Atlas for Maize Reveals Organ-specific and Stress-induced Genes. *Plant J.* **2019**, *97*, 1154–1167. [CrossRef] [PubMed]
53. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef]
54. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015. Available online: <https://pymol.org/2/> (accessed on 13 January 2023).
55. Sela, I.; Ashkenazy, H.; Katoh, K.; Pupko, T. GUIDANCE2: Accurate Detection of Unreliable Alignment Regions Accounting for the Uncertainty of Multiple Parameters. *Nucleic Acids Res.* **2015**, *43*, W7–W14. [CrossRef] [PubMed]
56. Nguyen, L.-T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
57. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
58. Minh, B.Q.; Nguyen, M.A.T.; von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **2013**, *30*, 1188–1195. [CrossRef]
59. Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef]
60. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [CrossRef]
61. Šali, A.; Blundell, T.L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [CrossRef]
62. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]
63. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J. Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]
64. Laskowski, R.A.; Rullmann, J.A.C.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M. AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR* **1996**, *8*, 477–486. [CrossRef] [PubMed]
65. Lüthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of Protein Models with Three-Dimensional Profiles. *Nature* **1992**, *356*, 83–85. [CrossRef] [PubMed]
66. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.;Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [CrossRef]
67. Hufford, M.B.; Seetharam, A.S.; Woodhouse, M.R.; Chougule, K.M.; Ou, S.; Liu, J.; Ricci, W.A.; Guo, T.; Olson, A.; Qiu, Y. De Novo Assembly, Annotation, and Comparative Analysis of 26 Diverse Maize Genomes. *Science* **2021**, *373*, 655–662. [CrossRef]
68. Laidlaw, H.K.; Lahnstein, J.; Burton, R.A.; Fincher, G.B.; Jobling, S.A. Analysis of the Arabinoxylan Arabinofuranohydrolase Gene Family in Barley Does Not Support Their Involvement in the Remodelling of Endosperm Cell Walls during Development. *J. Exp. Bot.* **2012**, *63*, 3031–3045. [CrossRef]
69. Carpita, N.C.; Gibeaut, D.M. Structural Models of Primary-Cell Walls in Flowering Plants—Consistency of Molecular-Structure with the Physical-Properties of the Walls during Growth. *Plant J.* **1993**, *3*, 1–30. [CrossRef]
70. Nyffeler, R.; Eggli, U. Introduction to the Classification of Monocotyledons BT—Monocotyledons. In *Monocotyledons. Illustrated Handbook of Succulent Plants*; Eggli, U., Nyffeler, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–6.
71. Debeche, T.; Bliard, C.; Debeire, P.; O'Donohue, M.J. Probing the Catalytically Essential Residues of the α -L-Arabinofuranosidase from *Thermobacillus xylanilyticus*. *Protein Eng.* **2002**, *15*, 21–28. [CrossRef]
72. Boraston, A.B.; Nurizzo, D.; Notenboom, V.; Ducros, V.; Rose, D.R.; Kilburn, D.G.; Davies, G.J. Differential Oligosaccharide Recognition by Evolutionarily-Related β -1, 4 and β -1, 3 Glucan-Binding Modules. *J. Mol. Biol.* **2002**, *319*, 1143–1156. [CrossRef]
73. Simpson, P.J.; Jamieson, S.J.; Abou-Hachem, M.; Karlsson, E.N.; Gilbert, H.J.; Holst, O.; Williamson, M.P. The Solution Structure of the CBM4-2 Carbohydrate Binding Module from a Thermostable *Rhodothermus marinus* Xylanase. *Biochemistry* **2002**, *41*, 5712–5719. [CrossRef]
74. Alahuhta, M.; Xu, Q.; Bomble, Y.J.; Brunecky, R.; Adney, W.S.; Ding, S.-Y.; Himmel, M.E.; Lunin, V.V. The Unique Binding Mode of Cellulosomal CBM4 from *Clostridium thermocellum* Cellobiohydrolase A. *J. Mol. Biol.* **2010**, *402*, 374–387. [CrossRef] [PubMed]
75. Alahuhta, M.; Luo, Y.; Ding, S.-Y.; Himmel, M.E.; Lunin, V.V. Structure of CBM4 from *Clostridium Thermocellum* Cellulase K. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2011**, *67*, 527–530. [CrossRef] [PubMed]
76. Boraston, A.B.; Bolam, D.N.; Gilbert, H.J.; Davies, G.J. Carbohydrate-Binding Modules: Fine-Tuning Polysaccharide Recognition. *Biochem. J.* **2004**, *382*, 769–781. [CrossRef]
77. Baker, D.; Sali, A. Protein Structure Prediction and Structural Genomics. *Science* **2001**, *294*, 93–96. [CrossRef]
78. Perez, S.; Makshakova, O. Multifaceted Computational Modeling in Glycoscience. *Chem. Rev.* **2022**, *122*, 15914–15970. [CrossRef] [PubMed]
79. Yang, Y.; Zhang, L.; Guo, M.; Sun, J.; Matsukawa, S.; Xie, J.; Wei, D. Novel α -L-Arabinofuranosidase from *Cellulomonas fimi* ATCC 484 and Its Substrate-Specificity Analysis with the Aid of Computer. *J. Agric. Food Chem.* **2015**, *63*, 3725–3733. [CrossRef]

80. Hervé, C.; Siméon, A.; Jam, M.; Cassin, A.; Johnson, K.L.; Salmeán, A.A.; Willats, W.G.T.; Doblin, M.S.; Bacic, A.; Kloareg, B. Arabinogalactan Proteins Have Deep Roots in Eukaryotes: Identification of Genes and Epitopes in Brown Algae and Their Role in *Fucus serratus* Embryo Development. *New Phytol.* **2016**, *209*, 1428–1441. [\[CrossRef\]](#)
81. Arad, S.; Levy-Ontman, O. Red Microalgal Cell-Wall Polysaccharides: Biotechnological Aspects. *Curr. Opin. Biotechnol.* **2010**, *21*, 358–364. [\[CrossRef\]](#)
82. Baudalet, P.-H.; Ricochon, G.; Linder, M.; Muniglia, L. A New Insight into Cell Walls of *Chlorophyta*. *Algal Res.* **2017**, *25*, 333–371. [\[CrossRef\]](#)
83. Weiss, T.L.; Roth, R.; Goodson, C.; Vitha, S.; Black, I.; Azadi, P.; Rusch, J.; Holzenburg, A.; Devarenne, T.P.; Goodenough, U. Colony Organization in the Green Alga *Botryococcus braunii* (Race B) Is Specified by a Complex Extracellular Matrix. *Eukaryot. Cell* **2012**, *11*, 1424–1440. [\[CrossRef\]](#)
84. Pieper, S.; Unterrieser, I.; Mann, F.; Mischnick, P. A New Arabinomannan from the Cell Wall of the Chlorococcal Algae *Chlorella Vulgaris*. *Carbohydr. Res.* **2012**, *352*, 166–176. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Suárez, E.R.; Kralovec, J.A.; Noseda, M.D.; Ewart, H.S.; Barrow, C.J.; Lumsden, M.D.; Grindley, T.B. Isolation, Characterization and Structural Determination of a Unique Type of Arabinogalactan from an Immunostimulatory Extract of *Chlorella pyrenoidosa*. *Carbohydr. Res.* **2005**, *340*, 1489–1498. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Sugimoto, I.; Onimatsu, H.; Fujie, M.; Usami, S.; Yamada, T. VAL-1, a Novel Polysaccharide Lyase Encoded by Chlorovirus CVK2. *FEBS Lett.* **2004**, *559*, 51–56. [\[CrossRef\]](#)
87. Mócsai, R.; Blaukopf, M.; Svehla, E.; Kosma, P.; Altmann, F. The N-Glycans of *Chlorella Sorokiniana* and a Related Strain Contain Arabinose but Have Strikingly Different Structures. *Glycobiology* **2020**, *30*, 663–676. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Zhang, L.; Wu, S.; Chang, X.; Wang, X.; Zhao, Y.; Xia, Y.; Trigiano, R.N.; Jiao, Y.; Chen, F. The Ancient Wave of Polyploidization Events in Flowering Plants and Their Facilitated Adaptation to Environmental Stress. *Plant. Cell Environ.* **2020**, *43*, 2847–2856. [\[CrossRef\]](#)
89. Anders, N.; Wilkinson, M.D.; Lovegrove, A.; Freeman, J.; Tryfona, T.; Pellny, T.K.; Weimar, T.; Mortimer, J.C.; Stott, K.; Baker, J.M. Glycosyl Transferases in Family 61 Mediate Arabinofuranosyl Transfer onto Xylan in Grasses. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 989–993. [\[CrossRef\]](#)
90. Kozlova, L.V.; Nazipova, A.R.; Gorshkov, O.V.; Petrova, A.A.; Gorshkova, T.A. Elongating Maize Root: Zone-Specific Combinations of Polysaccharides from Type I and Type II Primary Cell Walls. *Sci. Rep.* **2020**, *10*, 10956. [\[CrossRef\]](#)
91. Rumpagaporn, P.; Reuhs, B.L.; Kaur, A.; Patterson, J.A.; Keshavarzian, A.; Hamaker, B.R. Structural Features of Soluble Cereal Arabinoxylan Fibers Associated with a Slow Rate of in vitro Fermentation by Human Fecal Microbiota. *Carbohydr. Polym.* **2015**, *130*, 191–197. [\[CrossRef\]](#)
92. Pena, M.J.; Kulkarni, A.R.; Backe, J.; Boyd, M.; O'Neill, M.A.; York, W.S.; O'Neill, M.A.; York, W.S. Structural Diversity of Xylans in the Cell Walls of Monocots. *Planta* **2016**, *244*, 589–606. [\[CrossRef\]](#)
93. Jiao, Y.; Li, J.; Tang, H.; Paterson, A.H. Integrated Syntenic and Phylogenomic Analyses Reveal an Ancient Genome Duplication in Monocots. *Plant Cell* **2014**, *26*, 2792–2802. [\[CrossRef\]](#)
94. Burmeister, W.P.; Cottaz, S.; Driguez, H.; Iori, R.; Palmieri, S.; Henrissat, B. The Crystal Structures of Sinapis Alba Myrosinase and a Covalent Glycosyl–Enzyme Intermediate Provide Insights into the Substrate Recognition and Active-Site Machinery of an S-Glycosidase. *Structure* **1997**, *5*, 663–676. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Tiels, P.; Baranova, E.; Piens, K.; De Visscher, C.; Pynaert, G.; Nerinckx, W.; Stout, J.; Fudalej, F.; Hulpiau, P.; Tännler, S. A Bacterial Glycosidase Enables Mannose-6-Phosphate Modification and Improved Cellular Uptake of Yeast-Produced Recombinant Human Lysosomal Enzymes. *Nat. Biotechnol.* **2012**, *30*, 1225–1231. [\[CrossRef\]](#)
96. Zechel, D.L.; Withers, S.G. Glycosidase Mechanisms: Anatomy of a Finely Tuned Catalyst. *Acc. Chem. Res.* **2000**, *33*, 11–18. [\[PubMed\]](#)
97. Naumoff, D.G. GH10 Family of Glycoside Hydrolases: Structure and Evolutionary Connections. *Mol. Biol.* **2016**, *50*, 132–140. [\[CrossRef\]](#)
98. Li, G.; Wang, D.; Yang, R.; Logan, K.; Chen, H.; Zhang, S.; Skaggs, M.I.; Lloyd, A.; Burnett, W.J.; Laurie, J.D. Temporal Patterns of Gene Expression in Developing Maize Endosperm Identified through Transcriptome Sequencing. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7582–7587. [\[CrossRef\]](#)
99. Sabelli, P.A.; Larkins, B.A. The Development of Endosperm in Grasses. *Plant Physiol.* **2009**, *149*, 14–26. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Wilson, S.M.; Burton, R.A.; Doblin, M.S.; Stone, B.A.; Newbigin, E.J.; Fincher, G.B.; Bacic, A. Temporal and Spatial Appearance of Wall Polysaccharides during Cellularization of Barley (*Hordeum vulgare*) Endosperm. *Planta* **2006**, *224*, 655. [\[CrossRef\]](#)
101. Koutaniemi, S.; Tenkanen, M. Action of Three GH51 and One GH54 α -Arabinofuranosidases on Internally and Terminally Located Arabinofuranosyl Branches. *J. Biotechnol.* **2016**, *229*, 22–30. [\[CrossRef\]](#)
102. Kozlova, L.V.; Mokshina, N.E.; Nazipova, A.R.; Gorshkova, T.A. Systemic Use of “Limping” Enzymes in Plant Cell Walls. *Russ. J. Plant Physiol.* **2017**, *64*, 808–821. [\[CrossRef\]](#)

103. Jensen, J.K.; Busse-Wicher, M.; Poulsen, C.P.; Fangel, J.U.; Smith, P.J.; Yang, J.; Peña, M.; Dinesen, M.H.; Martens, H.J.; Melkonian, M. Identification of an Algal Xylan Synthase Indicates That There Is Functional Orthology between Algal and Plant Cell Wall Biosynthesis. *New Phytol.* **2018**, *218*, 1049–1060. [[CrossRef](#)]
104. Araki, R.; Ali, M.K.; Sakka, M.; Kimura, T.; Sakka, K.; Ohmiya, K. Essential Role of the Family-22 Carbohydrate-binding Modules for β -1, 3-1, 4-glucanase Activity of *Clostridium stercorarium* Xyn10B. *FEBS Lett.* **2004**, *561*, 155–158. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.