Mycologia Iranica 2026, Vol. 13 (No. 1): 25–38 JIrMS

Data Analysis

Bioinformatic analysis of domain and motif architecture of cellulase enzymes in five CAZy families of *Ascomycota*

Abbas Rahimi Galugahi[1], Mohammad Ali Tajick Ghanbary[1]✉, Valiollah Babaeizad[1], Mohammad Ali Barimani Varandi[1], Samira Shahbazi[2]

[1]*Department of Plant Protection. Faculty of Crop Sciences, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran*
[2]*Nuclear Agriculture Research School, Nuclear Science and Technology Research Institute (NSTRI), Karaj, Iran*

10.22092/mi.2026.371792.1338

ABSTRACT

Cellulase enzymes degrade cellulose, the most abundant structural polysaccharide in plant biomass, thereby playing a central role in fungal ecology and biotechnology. Here, we conducted a bioinformatic analysis of cellulase enzymes across the phylum *Ascomycota* to examine their diversity, domain architecture, conserved motifs, and evolutionary relationships. We retrieved cellulase protein sequences from 69 representative Ascomycota species in the UniProtKB database, analyzed them using InterProScan, COBALT, and the MEME Suite, and conducted phylogenetic reconstruction in MEGA11. Our domain analysis identified five major CAZy families: glycoside hydrolase families GH5, GH7, GH10, GH11, and auxiliary activity family AA9, demonstrating the structural diversity found in ascomycete cellulases. Motif analysis revealed multiple conserved and family-specific motifs, indicating strong conservation of catalytic cores and variability in peripheral regions. Multiple sequence alignment and phylogenetic analyses showed that cellulase proteins clustered clearly within CAZy families, with high bootstrap support, suggesting that domain and motif architectures provide robust phylogenetic signals. Together, these results clarify the structural and evolutionary diversity of ascomycete cellulases and establish a foundation for future functional, ecological, and biotechnological studies of fungal cellulases.

KEYWORDS
COBALT, InterProScan, MEME Suite, NCBI, UniProtKB.

INTRODUCTION

Cellulose is the most abundant organic polymer in plant biomass and plays a central role in global carbon cycling. Microbial activity largely drives its degradation in both natural and agricultural ecosystems (Lynd et al. 2002). However, the efficient enzymatic breakdown of cellulose is limited by the complex chemical composition and highly ordered structure of plant cell walls. This resistance is known as biomass recalcitrance (Himmel et al. 2007). Structurally, cellulose microfibrils act as the main load-bearing parts of plant cell walls. They provide mechanical stability and support plant growth by counteracting internal turgor pressure (Somerville 2006).

Filamentous fungi within the phylum *Ascomycota* are key contributors to the degradation of plant biomass, as they secrete sophisticated and efficient enzymatic systems capable of decomposing cellulose and other structural polysaccharides of the plant cell wall (de Vries and Visser 2001). More broadly, microbial communities harbor a remarkable diversity of carbohydrate-active enzymes (CAZymes), underscoring their fundamental role in the decomposition of complex plant cell wall polymers, including cellulose (El Kaoutari et al. 2013). The enzymatic hydrolysis of cellulose requires the coordinated action of multiple enzyme classes, including endoglucanases, cellobiohydrolases, and β-glucosidases, whose synergistic interactions enable the effective degradation of this highly recalcitrant substrate (Horn et al. 2012). In fungi, cellulases act on insoluble cellulose via distinct catalytic mechanisms,

Received: 17 Dec. 2025 Revised: 25 June 2026 Accepted: 27 June 2026 Published online: 30 June 2026
✉ Corresponding Author: Mohammad Ali Tajick Ghanbary; Email: m.tajick@sanru.ac.ir
Mycologia Iranica is licensed under a "Creative Commons Attribution 4.0 International (CC-BY 4.0)"
Published by Iranian Mycological Society (IrMS)—https://mij.areeo.ac.ir

including processive and non-processive modes, enabling the concerted degradation of both crystalline and amorphous regions of cellulose (Payne et al. 2015).

Glycoside hydrolases are classified into families based on amino acid sequence similarity, a framework that reflects shared structural and mechanistic features despite differences in substrate specificity (Henrissat and Davies 1997). The CAZy database provides a curated, sequence-based classification system that is widely regarded as the reference resource for CAZyme annotation, integrating functional, structural, and mechanistic data for carbohydrate-active enzymes (Lombard et al. 2014). Although enzymes within a glycoside hydrolase family may differ in substrate preference, modular organization, and auxiliary domains, they usually retain conserved catalytic cores, indicating a common evolutionary origin (Naumoff 2011).

At the genomic level, gene duplication, diversification, and loss generate functional diversity and drive the expansion of enzymatic repertoires (Hage and Rosso 2021). In fungi, comparative genomic analyses show that variation in cellulase complements reflects lineage-specific expansions and contractions of CAZyme families, which relate to adaptation to distinct ecological niches and nutritional strategies (Eastwood et al. 2011). Although much research describes the biochemical properties, functional roles, and industrial relevance of fungal cellulases (Lynd et al. 2002, Payne et al. 2015), most studies focus on individual enzyme families, select model organisms, or genome-wide CAZyme inventories, rarely providing detailed protein-level comparisons across multiple cellulase families (Eastwood et al. 2011, Lombard et al. 2014).

In particular, integrated analyses that simultaneously examine conserved domain architecture, motif organization, and sequence-based relationships across multiple cellulase families within *Ascomycota* remain limited. Although large-scale reviews and comparative genomic studies have substantially advanced our understanding of cellulase diversity and classification (Henrissat and Davies 1997, Lombard et al. 2014), they often lack systematic evaluation of conserved motifs and domain organization at the level of individual cellulase proteins across different CAZy families. Therefore, the present study aims to provide an integrated bioinformatic analysis of conserved domains, motif architecture, and evolutionary relationships of cellulase enzymes across 69 representative ascomycete species belonging to five major CAZy families (GH5, GH7, GH10, GH11, and AA9). By combining domain prediction, motif discovery, and phylogenetic reconstruction, this work establishes a structured framework for comparative protein-level characterization of fungal cellulases and contributes to a deeper understanding of their structural diversity and evolutionary organization within *Ascomycota*.

MATERIALS AND METHODS

Data collection and sequence retrieval

Cellulase protein sequences were retrieved from the reviewed (Swiss-Prot) section of UniProtKB for representative ascomycete species selected to encompass the major taxonomic groups within the phylum *Ascomycota* (Table 1).

Sequence retrieval used UniProtKB advanced query filters, beginning with a taxonomic restriction to phylum *Ascomycota*, and reviewed annotation status and functional annotation for cellulase activity. These filters ensured relevance and annotation quality for the target group. Subsequently, selection relied on enzyme names, EC numbers, and CAZy family assignments to target cellulase candidates. Retrieved entries were then manually curated to exclude non-cellulases, ambiguous or incomplete annotations, partial sequences, and low-quality entries, thereby improving dataset quality. Sequence selection emphasized completeness, conserved catalytic domains, and classification within cellulase-related CAZy families (GH5, GH7, GH10, GH11, AA9), thereby allowing the inclusion of proteins with varying lengths and domain architectures. When multiple isoforms were present for a species, one representative was chosen based on annotation quality and conservation of key catalytic domains. Finally, all selected sequences were downloaded in FASTA format for downstream analyses (Coudert et al. 2023).

Identification of conserved domains

Conserved functional domains in cellulase protein sequences were identified with InterProScan version 5. InterProScan integrates databases such as Pfam, SMART, and PROSITE, enabling comprehensive detection of domains and signatures linked to cellulase activity. Domain annotations were used to assign sequences to CAZy families and describe family-specific architectures.

Multiple sequence alignment was performed using COBALT (NCBI), which was used to visualize and compare conserved regions previously annotated by InterProScan. Rather than predicting domains independently, COBALT evaluated alignment consistency and the conservation of these annotated catalytic domains (Papadopoulos and Agarwala 2007, Jones et al. 2014).

Motif discovery analysis

Conserved sequence motifs were identified using the MEME Suite version 5.5.1. Motif discovery was performed using the MEME algorithm under the zero-or-one-occurrence-per-sequence (ZOOPS) model. The minimum and maximum motif widths were set to 6 and 50 amino acids, respectively, and the maximum number of motifs allowed in each search was limited to 20. Motif significance was evaluated using MEME's internal statistical framework, and only statistically supported motifs were retained for downstream analyses. Although up to 20 motifs were permitted during the motif

discovery process, the number, length, and positional distribution of significant motifs varied among GH5, GH7, GH10, GH11, and AA9 cellulase proteins, with each CAZy family exhibiting a characteristic motif profile (Bailey et al. 2015).

Table 1. List of ascomycete species used in this study.

Organism	Length (aa)	Order	UniProt accession	Lifestyle
Arthroderma benhamiae	413	*Onygenales*	D4AJL7	Pathogenic
Aspergillus aculeatus	540	*Eurotiales*	O59843	Saprotrophic
Aspergillus awamori	211	*Eurotiales*	P55328	Saprotrophic
Aspergillus clavatus	353	*Eurotiales*	A1C4H2	Opportunistic pathogen
Aspergillus flavus	367	*Eurotiales*	B8MXJ7	Pathogenic
Aspergillus fumigatus	250	*Eurotiales*	Q4WP32	Opportunistic pathogen
Aspergillus kawachii	332	*Eurotiales*	Q96WQ8	Saprotrophic
Aspergillus niger	412	*Eurotiales*	A2R5N0	Saprotrophic
Aspergillus oryzae	250	*Eurotiales*	Q2UGM5	Saprotrophic
Aspergillus tamarii	371	*Eurotiales*	A0A5N6UV50	Saprotrophic
Aspergillus terreus	360	*Eurotiales*	Q0CEU4	Opportunistic pathogen
Aspergillus tubingensis	211	*Eurotiales*	P55331	Saprotrophic
Aureobasidium pullulans	361	*Dothideales*	Q2PGV8	Saprotrophic
Blumeria graminis	426	*Erysiphales*	Q96V64	Pathogenic
Botryotinia fuckeliana	227	*Helotiales*	B3VSG7	Pathogenic
Byssochlamys spectabilis	198	*Eurotiales*	P81536	Saprotrophic
Candida albicans	438	*Saccharomycetales*	P29717	Pathogenic
Candida oleophila	425	*Serinales*	Q8NKF9	Yeast-Saprotroph
Claviceps purpurea	216	*Hypocreales*	O74716	Pathogenic
Cochliobolus carbonum	456	*Pleosporales*	Q00328	Pathogenic
Collariella virescens	274	*Sordariales*	A0A223GEC9	Saprotrophic
Cryphonectria parasitica	452	*Diaporthales*	Q00548	Pathogenic
Emericella nidulans	357	*Eurotiales*	Q5BCX8	Saprotrophic
Fusarium oxysporum	429	*Hypocreales*	P46237	Pathogenic
Fusarium oxysporum f. sp. *lycopersici*	327	*Hypocreales*	B3A0S5	Pathogenic
Geotrichum candidum	288	*Dipodascales*	A0A0J9XKT0	Saprotrophic
Gibberella zeae	327	*Hypocreales*	I1S3T9	Pathogenic
Humicola insolens	402	*Sordariales*	P56680	Saprotrophic
Hypocrea jecorina	418	*Hypocreales*	P07982	Saprotrophic
Hypocrea rufa	223	*Hypocreales*	Q9UVF9	Saprotrophic
Kluyveromyces lactis	429	*Saccharomycetales*	Q12628	Yeast-Saprotroph
Lachancea kluyveri	439	*Saccharomycetales*	Q875R9	Yeast-Saprotroph
Malbranchea cinnamomea	334	*Onygenales*	A0A5J6BJT3	Saprotrophic
Neosartorya fischeri	348	*Eurotiales*	A1DBS6	Saprotrophic
Neurospora crassa	390	*Sordariales*	Q7SDR1	Saprotrophic
Penicillium canescens	328	*Eurotiales*	Q5S7A8	Saprotrophic
Penicillium chrysogenum	353	*Eurotiales*	P29417	Saprotrophic
Penicillium citrinum	217	*Eurotiales*	Q2PGY1	Saprotrophic
Penicillium digitatum	216	*Eurotiales*	K9GHZ3	Pathogenic
Penicillium janthinellum	537	*Eurotiales*	Q06886	Saprotrophic
Penicillium oxalicum	310	*Eurotiales*	E7EF85	Saprotrophic
Penicillium parvum	249	*Eurotiales*	A0A6C0M6J9	Saprotrophic
Penicillium simplicissimum	302	*Eurotiales*	P56588	Saprotrophic
Pichia angusta	435	*Pichiales*	Q12626	Yeast-Saprotroph
Podospora anserina	373	*Sordariales*	B2B3C0	Saprotrophic
Pyricularia grisea	331	*Magnaporthales*	Q01176	Pathogenic
Pyricularia oryzae	424	*Magnaporthales*	G4NK46	Pathogenic
Robillarda sp.	385	*Amphisphaeriales*	P23044	Saprotrophic
Saccharomyces cerevisiae	448	*Saccharomycetales*	P23776	Yeast-Saprotroph
Saccharomyces uvarum	448	*Saccharomycetales*	Q876J3	Yeast-Saprotroph
Schizosaccharomyces pombe	407	*Schizosaccharomycetales*	Q9URU6	Yeast-saprotroph
Schwanniomyces occidentalis	425	*Serinales*	Q12700	Yeast-Saprotroph
Talaromyces funiculosus	282	*Eurotiales*	Q8J0K5	Saprotrophic
Talaromyces pinophilus	310	*Eurotiales*	A0A478ECY3	Saprotrophic
Talaromyces purpureogenus	329	*Eurotiales*	Q9P8J1	Saprotrophic
Talaromyces verruculosus	328	*Eurotiales*	A0A482A9N4	Saprotrophic
Thermoascus aurantiacus	228	*Eurotiales*	G3XAP7	Saprotrophic
Thermomyces lanuginosus	225	*Eurotiales*	O43097	Saprotrophic
Thermothelomyces thermophilus	342	*Sordariales*	G2QAB5	Saprotrophic
Trichoderma harzianum	505	*Hypocreales*	Q9P8P3	Endophytic
Trichoderma koningii	513	*Hypocreales*	P62695	Endophytic
Trichoderma longibrachiatum	463	*Hypocreales*	Q12714	Saprotrophic
Trichoderma parareesei	266	*Hypocreales*	A0A0S0FTT9	Saprotrophic
Verticillium dahliae	223	*Glomerellales*	G2X4G0	Pathogenic
Verticillium longisporum	287	*Glomerellales*	A0A0G4KL96	Pathogenic
Wickerhamomyces anomalus	498	*Phaffomycetales*	O93939	Yeast-Saprotroph
Yarrowia lipolytica	421	*Dipodascales*	Q12725	Yeast-Saprotroph

Multiple Sequence Alignment

Multiple sequence alignment of cellulase amino acid sequences was performed using the ClustalW algorithm. The resulting alignments were manually inspected to assess overall alignment quality and to evaluate conservation patterns, with particular emphasis on catalytic regions and conserved domains identified through InterProScan and motif analyses (Jones et al. 2014, Tamura et al. 2021).

Phylogenetic Analysis

Phylogenetic relationships among cellulase enzymes were inferred using MEGA version 11 software. Neighbor-joining phylogenetic trees were constructed from amino acid alignments, with pairwise evolutionary distances estimated using the Poisson correction model. Gaps and missing data were addressed by pairwise deletion, and branch robustness was assessed using bootstrap analysis with 1000 replicates (Felsenstein 1985).

A cellulase protein sequence from *Phanerochaete chrysosporium* served as an outgroup to root the phylogenetic trees. This species was chosen because it is a phylogenetically distant basidiomycete with homologous cellulase sequences, making it suitable for this analysis (Tamura et al. 2021).

RESULTS

Domain architecture of cellulase enzymes in *Ascomycota*

Analysis of cellulase amino acid sequences using InterProScan revealed multiple conserved catalytic domains from well-characterized CAZy families. Five major domain types were identified in the dataset: glycoside hydrolase family 5 (GH5), glycoside hydrolase family 7 (GH7), glycoside hydrolase family 10 (GH10), glycoside hydrolase family 11 (GH11), and auxiliary activity family 9 (AA9, formerly GH61). GH5 domains appeared in 17 sequences, GH7 in 10, GH10 in 7, GH11 in 13, and AA9 domains in 18 (Table 2). The domain distributions among ascomycete species highlight the diversity of cellulase enzymes within this phylum (Fig. 1). InterProScan annotations were further supported by COBALT alignment analyses, which revealed consistent conservation patterns across all sequences (Fig. 2). These consistent results support the reliability of domain assignments and highlight the strong conservation of catalytic cores across cellulase families. The domain architectures indicate that cellulase families vary not only in catalytic domain type but also in modular organization. GH5, GH7, GH10, and GH11 proteins mainly have a single catalytic domain. In contrast, AA9 proteins often exhibit distinct features associated with oxidative cellulose degradation. The consistent domain boundaries within each family indicate strong structural conservation of catalytic cores, even with substantial taxonomic diversity in *Ascomycota*.

Conserved motif identification and distribution

Motif discovery analysis with the MEME Suite identified multiple conserved motifs in cellulase protein sequences. Each glycoside hydrolase and auxiliary activity family showed a unique motif profile. These profiles reflect conserved functional and structural features. Several motifs were highly conserved within families, while others were partly shared among related families. This pattern suggests evolutionary conservation of catalytic elements (Fig. 3). Although up to 20 motifs were allowed in the MEME search, the number of significant motifs varied across cellulase families. The count ranged from about 6 to 10 motifs per family, showing family-specific patterns and conserved positions within catalytic regions. The number, length, and locations of motifs varied among GH5, GH7, GH10, GH11, and AA9 proteins. These differences contributed to the separation of cellulase families seen in motif analyses (Fig. 3). Such variation is likely linked to differences in substrate specificity and catalytic mechanisms. Motif analysis showed that most conserved motifs localize within catalytic cores. More variable motifs appear in peripheral regions that differ among families. These patterns support a distinction between conserved functional elements key for catalysis and variable regions that may affect substrate interaction, specificity, or structural stability.

Multiple sequence alignment and conservation patterns

Multiple sequence alignment revealed strong conservation within catalytic domains, mainly in active-site and motif-rich regions (Fig. 4). In contrast, terminal and inter-domain linker regions showed much higher sequence variability. This pattern suggests that residues important for enzyme function are under evolutionary constraint. At the same time, peripheral regions remain flexible and may help enable functional diversity.

Phylogenetic analysis of cellulase enzymes

Phylogenetic analysis revealed clear clustering of cellulase enzymes by CAZy family. High bootstrap support separated GH5, GH7, GH10, GH11, and AA9 proteins into well-defined clades (Fig. 5). Inside each clade, cellulases from related Ascomycota species often clustered together. This pattern shows lineage-associated diversification. The results suggest that both domain composition and evolutionary history shape the phylogeny of fungal cellulases.

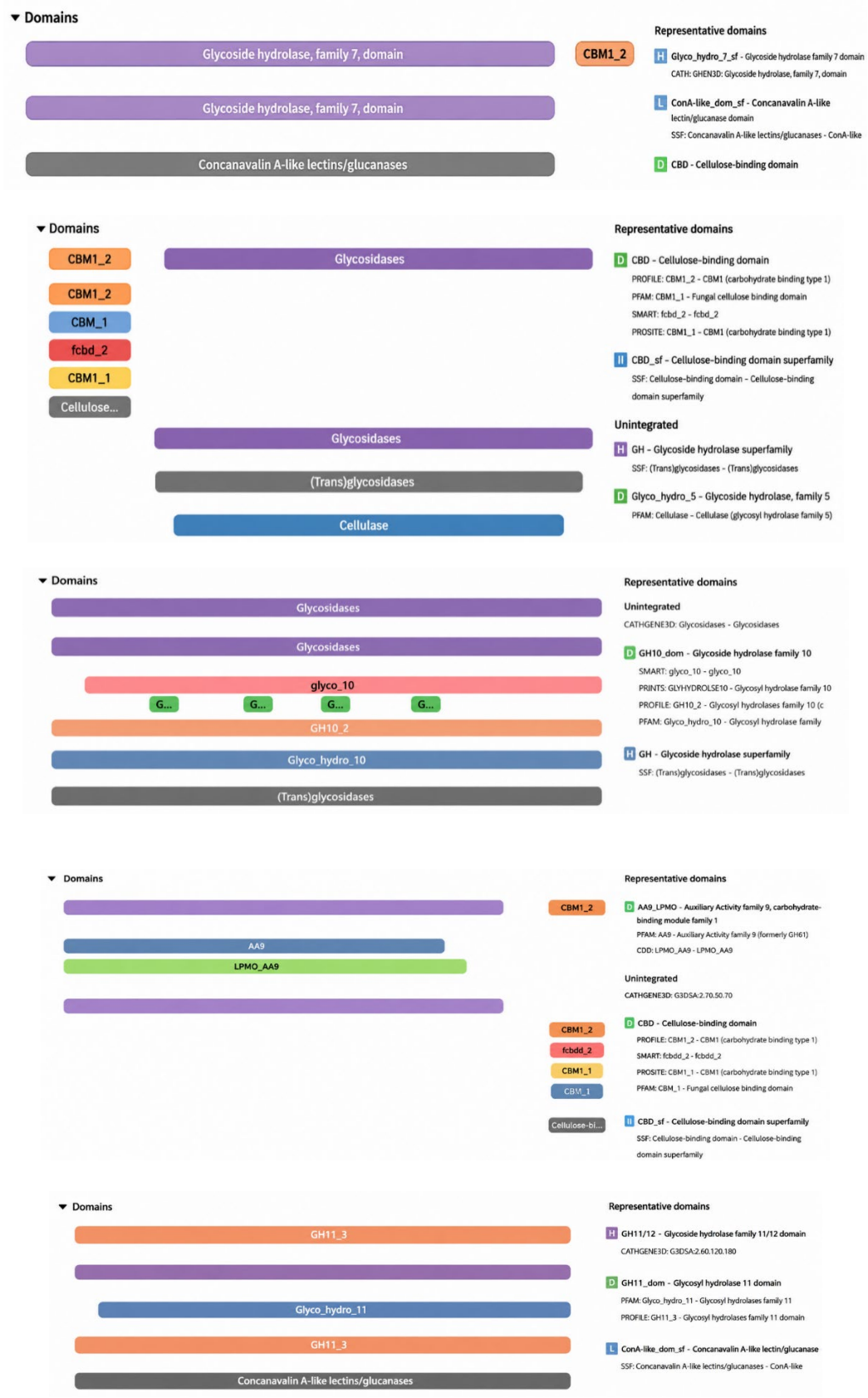

Fig. 1. Domain architectures of cellulase enzymes from representative ascomycetes species identified using InterProScan. Conserved catalytic domains corresponding to CAZy families (GH5, GH7, GH10, GH11, and AA9) illustrate family-specific modular organization, with distinct architectures observed for AA9 proteins.

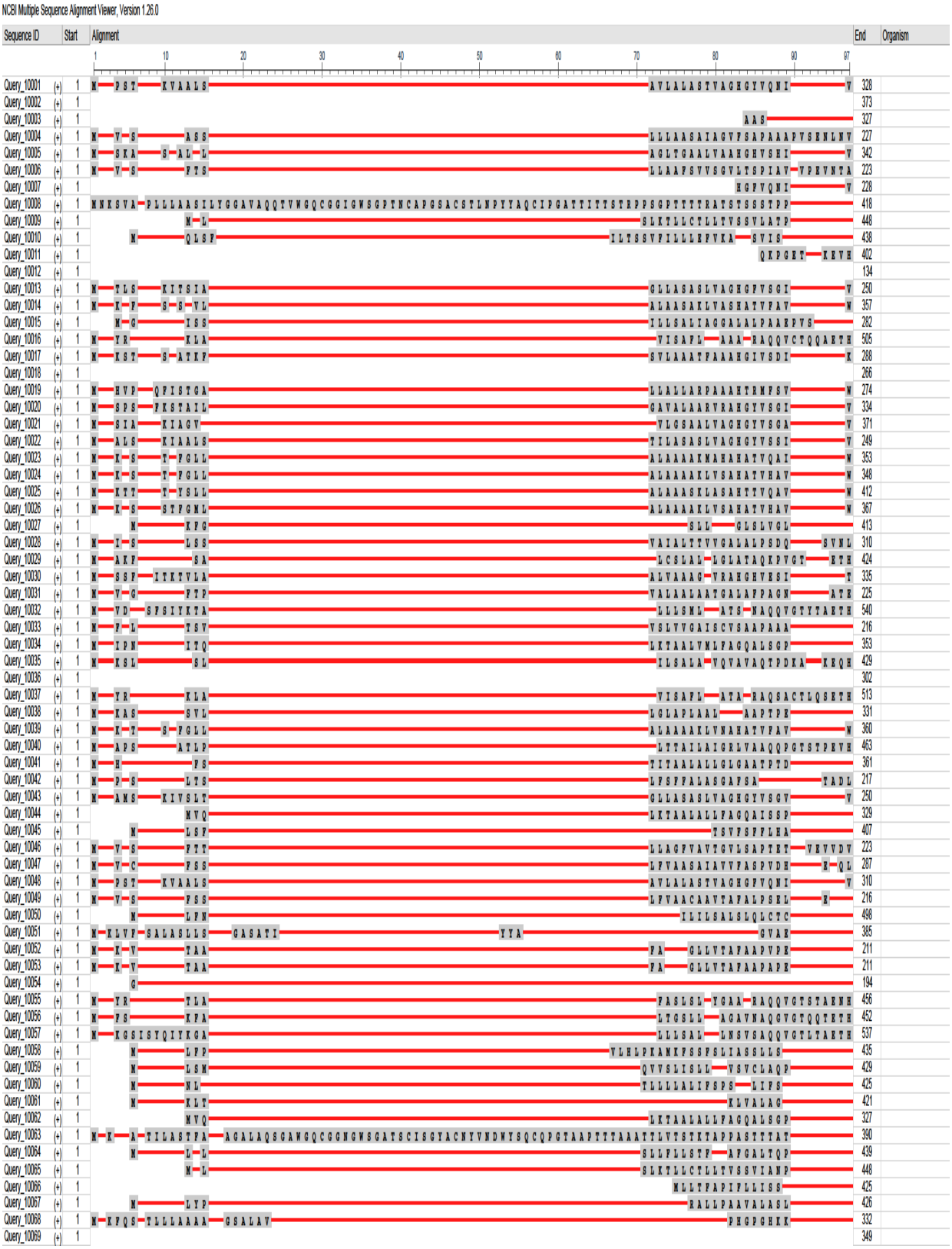

Fig. 2. Conserved regions in cellulase protein sequences visualized by multiple sequence alignment using the COBALT tool. Conserved blocks correspond to InterProScan-annotated catalytic domains and show strong positional conservation across cellulase families.

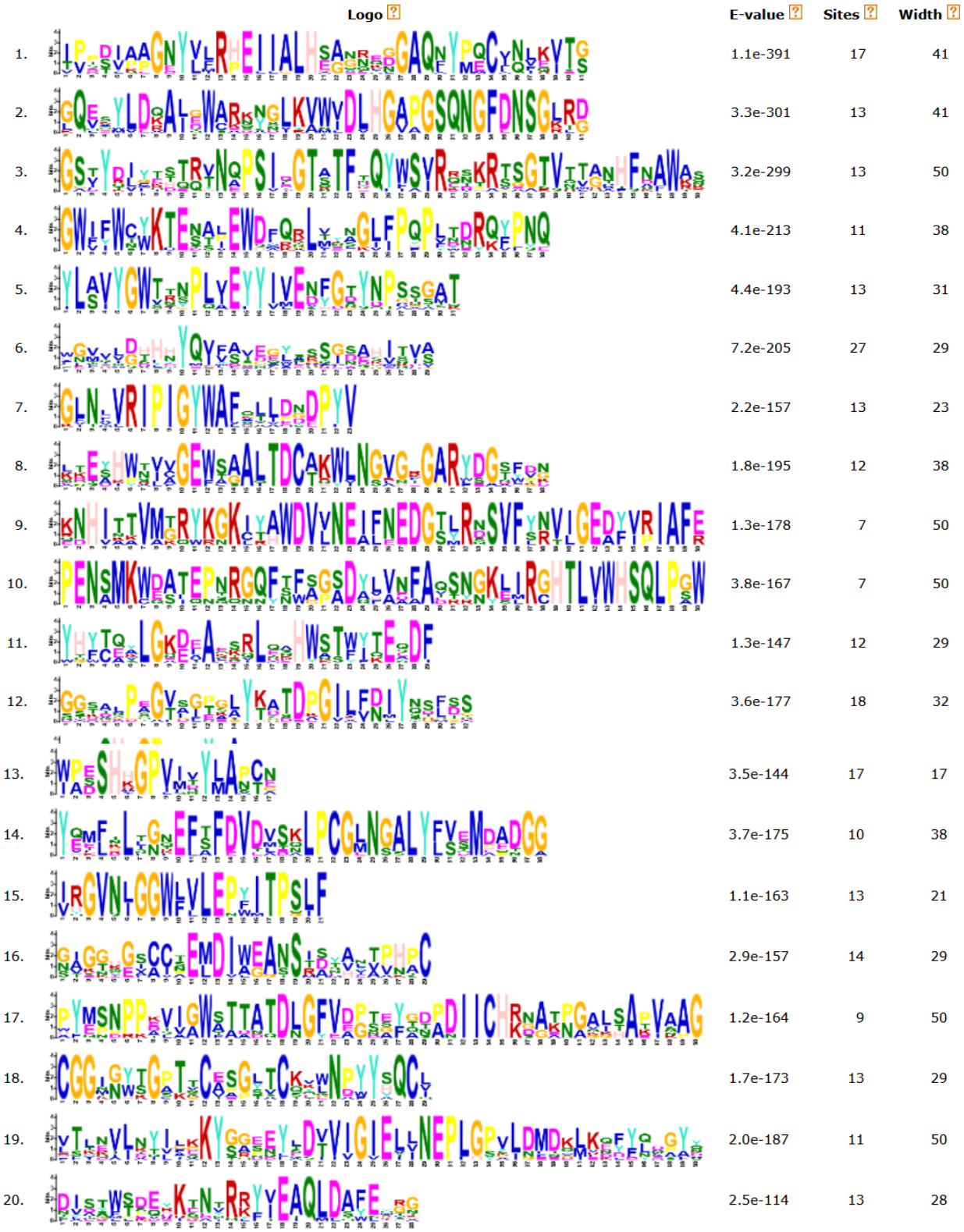

Fig. 3. Sequence logos of conserved motifs identified in cellulase proteins using MEME Suite version 5.5.1. Motif patterns reveal family-specific conservation signatures within GH5, GH7, GH10, GH11, and AA9 cellulases**.**

Fig. 4. Positional distribution of conserved motifs in cellulase protein sequences identified by MEME analysis. Differences in motif arrangement among cellulase families highlight variability outside conserved catalytic regions.

Fig. 4. Continued

Table 2. The list of identified domains using InterProScan.

Domains identified	Number of sequences
Glycoside hydrolase family 5 (GH5) catalytic domain (PF00150)	17
Glycoside hydrolase family 7 (GH7) catalytic domain (PF00840)	10
Glycoside hydrolase family 10 (GH10) domain (PF00331)	7
Glycoside hydrolase family 11 (GH11) domain (PF00457)	13
Auxiliary activity family 9 (AA9, formerly GH61) domain (PF03443)	18

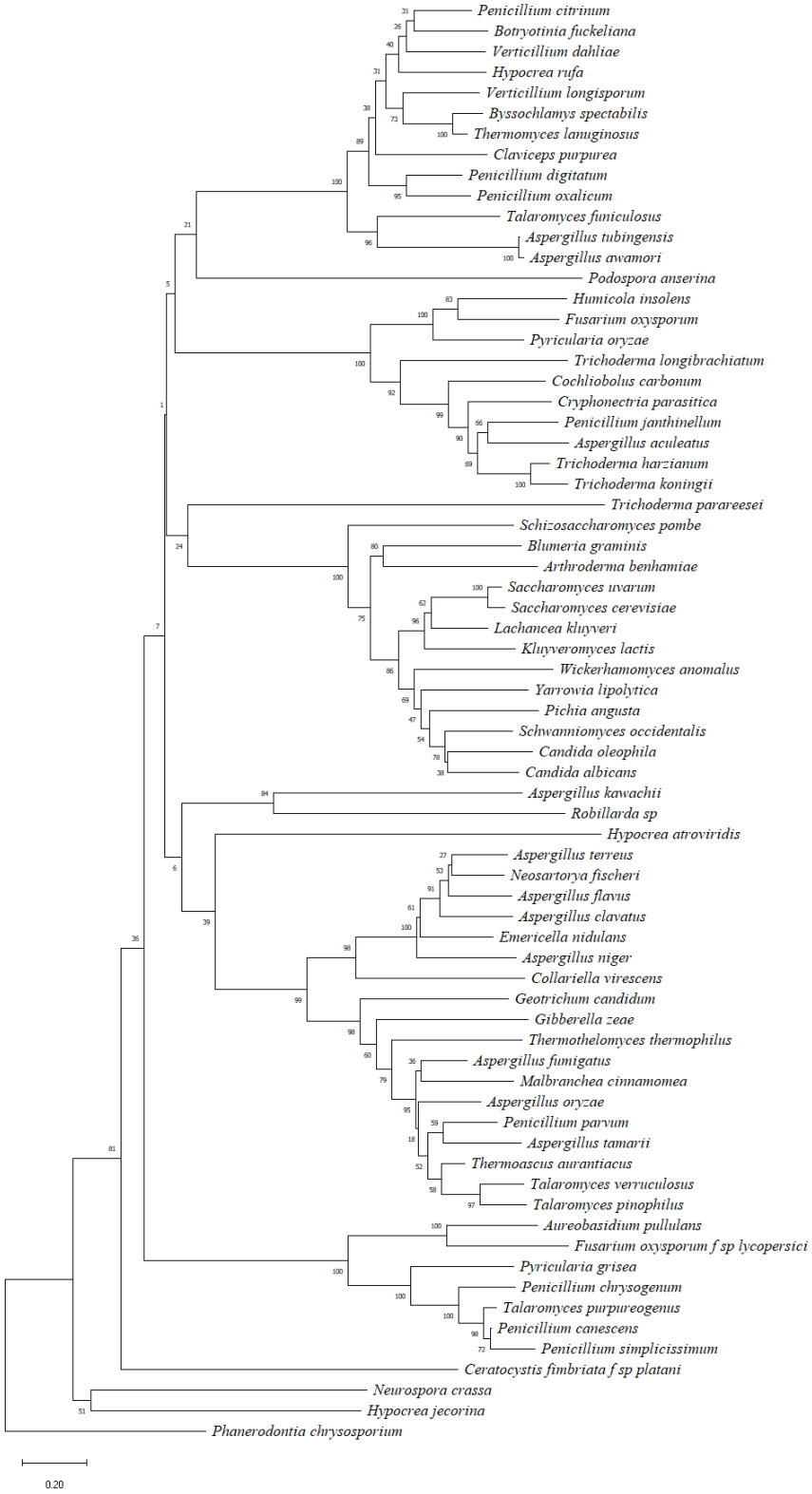

Fig. 5. Phylogenetic relationships of 69 cellulase proteins from Ascomycota inferred using the Neighbor-Joining method based on amino acid alignments. The sequences clustered into distinct, well-supported clades corresponding to CAZy families. Bootstrap values are shown at the nodes**.**

DISCUSSION

Cellulase enzymes constitute a central component of fungal strategies for plant cell wall degradation and carbon acquisition. In the present study, a comprehensive comparative bioinformatic analysis revealed substantial diversity in cellulase enzymes among Ascomycota species, encompassing five major CAZy families (GH5, GH7, GH10, GH11, and AA9). The coexistence of multiple cellulase families within Ascomycota likely reflects long-term evolutionary adaptation to diverse ecological niches and heterogeneous plant-derived substrates, a pattern widely documented in genome- and CAZyme-scale comparative studies of fungi (Hage and Rosso 2021, Resl et al. 2022).

The conserved domain architectures and family-specific motif patterns identified in this study are consistent with established structure–function relationships described for fungal cellulases. GH7 cellulases, which play a central role in the degradation of crystalline cellulose, exhibit highly conserved catalytic motifs associated with their processive mode of action, in agreement with previous biochemical and structural studies. In contrast, AA9 enzymes displayed distinct architectural features characteristic of lytic polysaccharide monooxygenases (LPMOs), reflecting their oxidative mechanism of cellulose cleavage and their dependence on copper-mediated catalysis (Ipsen et al. 2021; Contato and Conte-Junior 2025). The observed variability in motif composition and positional distribution within the GH10 and GH11 families likely reflects differences in substrate specificity and enzymatic versatility, which are known to distinguish endo-acting xylanases and related enzymes within these families.

Across all cellulase families examined, strong conservation of catalytic cores was evident, underscoring the essential functional constraints imposed on residues directly involved in catalysis. Increased sequence variability in terminal regions, linker segments, and peripheral motifs suggests that evolutionary diversification has fine-tuned enzymatic properties, including substrate accessibility, protein interactions, and environmental stability. These patterns have been reported for carbohydrate-active enzymes in diverse fungal lineages, supporting the generality of this evolutionary principle (Hage and Rosso 2021).

Phylogenetic analysis demonstrated robust clustering of cellulase sequences by CAZy family, with high bootstrap support clearly separating the GH and AA families. Within individual family-level clades, enzymes from taxonomically related *Ascomycota* species frequently clustered together, indicating lineage-associated diversification. Such phylogenetic patterns are consistent with gene duplication followed by functional specialization, a dominant evolutionary mechanism shaping fungal CAZyme repertoires in response to ecological pressures and substrate availability (Hage and Rosso 2021; Wang et al. 2022). Although horizontal gene transfer has been implicated in the evolution of certain fungal metabolic pathways, the strong correspondence observed here between phylogenetic relationships and species taxonomy does not support widespread horizontal transfer of cellulase genes among the analyzed Ascomycota species.

The motif distributions identified here support a functional perspective of cellulase evolution. Conserved motifs within catalytic regions likely sustain enzymatic activity, while variable motifs and linker regions may enhance substrate flexibility and enzyme interactions. Integrative motif-based analyses, especially with structural data, effectively elucidate functional divergence among CAZymes and could further refine understanding of cellulase diversity (Ipsen et al. 2021; Resl et al. 2022).

From an applied perspective, the broad diversity of cellulase families and architectures across *Ascomycota* highlights the substantial biotechnological potential of fungal cellulases. Differences in catalytic mechanisms, substrate preferences, and structural organization among GH and AA families offer a valuable resource for industrial enzyme discovery and optimization. Key application areas include biomass conversion, biofuel production, and sustainable agriculture. Future studies integrating structural modeling, gene expression analyses, and experimental validation will be essential to fully explore the functional potential of the cellulase repertoires identified in this study.

In conclusion, this study provides a comprehensive comparative analysis of cellulase enzymes across five major CAZy families (GH5, GH7, GH10, GH11, and AA9) in *Ascomycota*. By integrating domain architecture characterization, conserved motif discovery, and phylogenetic reconstruction, the study reveals both the structural conservation and evolutionary diversification of fungal cellulases. The clear family-specific patterns identified here demonstrate the value of combining multiple bioinformatic approaches for protein-level characterization of CAZymes. Beyond improving our understanding of cellulase evolution, these findings provide a useful framework for future studies aimed at enzyme discovery, functional prediction, and biotechnological exploitation of fungal cellulases in biomass conversion **and** related industrial applications.

ACKNOWLEDGMENTS

The authors sincerely acknowledge Sari Agricultural Sciences and Natural Resources University for its support and for providing a valuable academic environment that facilitated this research. We are grateful to the university for its commitment to advancing scientific knowledge and promoting academic excellence. The authors also appreciate the assistance and encouragement provided throughout the course of this study, which contributed to its successful completion.

AUTHOR CONTRIBUTION

All authors contributed to the conception and design of the study. Abbas Rahimi Galugahi performed the bioinformatic analyses, curated the dataset, interpreted the data, prepared the figures and tables, and drafted the manuscript. Mohammad Ali Tajick Ghanbary contributed to the conceptualization, supervision, validation, and critical revision of the manuscript. Valiollah Babaeizad contributed to the bioinformatics workflow, software implementation, and data analysis. Mohammad Ali Barimani Varandi contributed to the study design, investigation, and visualization. Samira Shahbazi contributed to the methodology, data interpretation, and manuscript revision. All authors read and approved the final manuscript.

DATA AVAILABILITY

The datasets analyzed during the current study were obtained from publicly available databases. The curated datasets and supporting bioinformatic analysis files are available from the corresponding author upon reasonable request.

FUNDING

The author declares that no financial support was received during this research.

DECLARATION

The authors declare no conflicts of interest.

ETHICAL APPROVAL

This study was based exclusively on publicly available data and did not involve human participants, animals, or biological samples.

REFERENCES

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. Nucleic Acids Research. 43(W1): W39-W49. https://doi.org/10.1093/nar/gkv416.

Contato AG, Conte-Junior CA. 2025. Fungal lytic polysaccharide monooxygenases (LPMOs): Functional adaptation and biotechnological perspectives. Eng. 6(8): 177-177. https://doi.org/10.3390/eng6080177.

Coudert E, Gehant S, de Castro E, Pozzato M, Baratin D, Neto T, Sigrist CJA, Redaschi N, Bridge A. 2023. Annotation of biologically relevant ligands in UniProtKB using ChEBI. Bioinformatics. 39: btac803-btac803. https://doi.org/10.1093/bioinformatics/btac803.

de Vries RP, Visser J. 2001. *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. Microbiology and Molecular Biology Reviews. 65(4): 497-522. https://doi.org/10.1128/MMBR.65.4.497-522.2001.

Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Watkinson SC, Hibbett DS. 2011. The plant cell wall–decomposing machinery underlies the functional diversity of forest fungi. Science. 333(6043): 762-765. https://doi.org/10.1126/science.1205411.

El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. 2013. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nature Reviews Microbiology. 11(7): 497-504. https://doi.org/10.1038/nrmicro3050.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39(4): 783-791. https://doi.org/10.2307/2408678.

Hage H, Rosso M-N. 2021. Evolution of fungal carbohydrate-active enzyme portfolios and adaptation to plant cell-wall polymers. Journal of Fungi. 7(3): 185-185. https://doi.org/10.3390/jof7030185.

Henrissat B, Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. Current Opinion in Structural Biology. 7(5): 637-644. https://doi.org/10.1016/S0959-440X(97)80072-3.

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. Science. 315(5813): 804-807. https://doi.org/10.1126/science.1137016.

Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH. 2012. Novel enzymes for the degradation of cellulose. Biotechnology for Biofuels. 5: 45-45. https://doi.org/10.1186/1754-6834-5-45.

Ipsen JØ, Hallas-Møller M, Brander S, Lo Leggio L, Johansen KS. 2021. Lytic polysaccharide monooxygenases and other histidine-brace copper proteins: Structure, oxygen activation and biotechnological applications. Biochemical Society Transactions. 49(1): 531-540. https://doi.org/10.1042/BST20201031.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Hunter S. 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 30(9): 1236-1240. https://doi.org/10.1093/bioinformatics/btu031.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research. 42(D1): D490-D495. https://doi.org/10.1093/nar/gkt1178.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiology and Molecular Biology Reviews. 66(3): 506-577. https://doi.org/10.1128/MMBR.66.3.506-577.2002.

Naumoff DG. 2011. Hierarchical classification of glycoside hydrolases. Biochemistry (Moscow). 76(6): 622-635. https://doi.org/10.1134/S0006297911060022.

Papadopoulos JS, Agarwala R. 2007. COBALT: A constraint-based alignment tool for multiple protein sequences. Bioinformatics. 23(9): 1073-1079. https://doi.org/10.1093/bioinformatics/btm076.

Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Beckham GT. 2015. Fungal cellulases. Chemical Reviews. 115(3): 1308-1448. https://doi.org/10.1021/cr500351c.

Resl P, Bujold AR, Tagirdzhanova G, Meidl P, Freire Rallo S, Kono M, Spribille T. 2022. Large differences in carbohydrate degradation and transport potential among lichen fungal symbionts. Nature Communications. 13: 2634-2634. https://doi.org/10.1038/s41467-022-30218-6.

Somerville C. 2006. Cellulose synthesis in higher plants. Annual Review of Cell and Developmental Biology. 22: 53-78. https://doi.org/10.1146/annurev.cellbio.22.022206.160206.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular evolutionary genetics analysis version 11. Molecular Biology and Evolution. 38(7): 3022-3027. https://doi.org/10.1093/molbev/msab120.

Wang Y, Wu J, Yan J, Guo M, Xu L, Hou L, Zou Q. 2022. Comparative genome analysis of plant ascomycete fungal pathogens with different lifestyles reveals distinctive virulence strategies. BMC Genomics. 23(1): 34-34. https://doi.org/10.1186/s12864-021-08165-1.

قارچ شناسی ایران جلد ۱۳، شماره ۱، ۱۴۰۵

واکاوی داده

بررسی بیوانفورماتیکی معماری دامنه‌ها و موتیف‌های آنزیم‌های سلولاز در پنج خانواده CAZy شاخه آسکومیکوتا

عباس رحیمی گلوگاهی[۱]، محمد علی تاجیک قنبری [۱] ✉ ، ولی اله بابایی زاد[۱]، محمد علی بریمانی ورندی[۱]، سمیرا شهبازی[۲]

[۱] گروه گیاه‌پزشکی، دانشکده علوم زراعی، دانشگاه کشاورزی و منابع طبیعی ساری، ساری، ایران

[۲] پژوهشکده کشاورزی هسته‌ای، پژوهشگاه علوم و فنون هسته‌ای، کرج، ایران

doi 10.22092/mi.2026.371792.1338

چکیده

آنزیم‌های سلولاز نقش اساسی در تجزیه سلولز، به‌عنوان فراوان‌ترین پلی‌ساکارید ساختمانی در زیست‌توده گیاهی، ایفا می‌کنند و از اهمیت بالایی در بوم‌شناسی قارچ‌ها و کاربردهای زیست‌فناورانه برخوردارند. در این پژوهش، یک تحلیل جامع زیست‌اطلاعاتی به منظور بررسی تنوع، معماری دامنه‌ها، موتیف‌های حفاظت‌شده و روابط تکاملی آنزیم‌های سلولاز در شاخه *Ascomycota* انجام شد. توالی‌های اسید آمینه‌ای پروتئین‌های سلولاز متعلق به ۶۹ گونه نماینده از آسکومیست‌ها از پایگاه داده UniProtKB استخراج و با استفاده از ابزارهای InterProScan، COBALT، MEME Suite و بازسازی فیلوژنتیکی در نرم‌افزار MEGA11 مورد تحلیل قرارگرفتند. تحلیل دامنه‌ها حضور پنج خانواده اصلی از آنزیم‌های CAZy شامل خانواده‌های گلیکوزید هیدرولاز GH5، GH7، GH10، GH11 و خانواده فعالیت کمکی AA9 را آشکار ساخت که بیانگر تنوع گسترده آنزیم‌های سلولاز در آسکومیست‌ها است. تحلیل موتیف‌ها وجود موتیف‌های حفاظت‌شده و اختصاصی هر خانواده را در کنار تنوع ساختاری در نواحی جانبی پروتئین‌ها نشان داد که حاکی از حفاظت هسته‌های کاتالیتیکی و تغییرپذیری بخش‌های پیرامونی است. هم‌ترازی چندگانه توالی‌ها و تحلیل‌های فیلوژنتیکی خوشه‌بندی مشخص آنزیم‌های سلولاز را بر اساس خانواده‌های CAZy، با پشتیبانی بوت‌استرپ بالا، نشان داد که بیانگر وجود سیگنال فیلوژنتیکی قوی مبتنی بر ترکیب دامنه‌ها و معماری موتیف‌ها است. در مجموع، این مطالعه تنوع ساختاری و تکاملی آنزیم‌های سلولاز در آسکومیست‌ها را روشن ساخته و چارچوبی ارزشمند برای مطالعات آینده در زمینه‌های عملکردی، بوم‌شناختی و زیست‌فناورانه سلولازهای قارچی فراهم می‌آورد.

واژگان کلیدی: COBALT، InterProScan، MEME suite، NCBI، Uniprot.

مکاتبه‌کننده: محمد تاجیک قنبری، m.tajick@sanru.ac.ir

تاریخ دریافت: ۱۴۰۴/۰۹/۲۶، تاریخ بازنگری: ۱۴۰۵/۰۴/۰۴، تاریخ پذیرش: ۱۴۰۵/۰۴/۰۶، تاریخ انتشار برخط: ۱۴۰۵/۰۴/۰۹