Database, Vol. 2011, Article ID bar020, doi:10.1093/database/bar020

Original article

Curation of characterized glycoside hydrolases of Fungal origin

Caitlin Murphy[1,2], Justin Powlowski[1,3], Min Wu[1], Greg Butler[1,4] and Adrian Tsang[1,2,*]

[1]Centre for Structural and Functional Genomics, [2]Department of Biology, [3]Department of Chemistry and Biochemistry and [4]Department of Computer Science and Software Engineering, Concordia University, Montreal QC H4B 1R6, Canada

***Corresponding author:** Tel: +514 848 2424 (Ext 3405); Fax: +514 848 4504; Email: tsang@alcor.concordia.ca

Submitted 9 February 2011; Revised 1 April 2011; Accepted 29 April 2011

Fungi produce a wide range of extracellular enzymes to break down plant cell walls, which are composed mainly of cellulose, lignin and hemicellulose. Among them are the glycoside hydrolases (GH), the largest and most diverse family of enzymes active on these substrates. To facilitate research and development of enzymes for the conversion of cell-wall polysaccharides into fermentable sugars, we have manually curated a comprehensive set of characterized fungal glycoside hydrolases. Characterized glycoside hydrolases were retrieved from protein and enzyme databases, as well as literature repositories. A total of 453 characterized glycoside hydrolases have been cataloged. They come from 131 different fungal species, most of which belong to the phylum Ascomycota. These enzymes represent 46 different GH activities and cover 44 of the 115 CAZy GH families. In addition to enzyme source and enzyme family, available biochemical properties such as temperature and pH optima, specific activity, kinetic parameters and substrate specificities were recorded. To simplify comparative studies, enzyme and species abbreviations have been standardized, Gene Ontology terms assigned and reference to supporting evidence provided. The annotated genes have been organized in a searchable, online database called *myco*CLAP (Characterized Lignocellulose-Active Proteins of fungal origin). It is anticipated that this manually curated collection of biochemically characterized fungal proteins will be used to enhance functional annotation of novel GH genes.

Database URL: http://mycoCLAP.fungalgenomics.ca/

Introduction

Plant cell walls are composed mainly of cellulose, lignin and hemicellulose. This composite is often referred to as lignocellulose, and is the most abundant renewable resource that has the potential of replacing fossil fuels in the production of a wide spectrum of fuels, chemicals and materials. One of the key challenges facing the widespread use of lignocellulose for fuel and chemical production is in finding economically and environmentally sustainable solutions to the conversion of lignocellulose into sugar building blocks. The fungal kingdom encompasses tremendous genetic diversity, and by virtue of secreted enzymes many of its members are potent decomposers of plant cell walls. Glycoside hydrolases (GH) are the most diverse group of enzymes used by microbes in the degradation of biomass. Over a hundred GH families have been classified to date (1–5). Many of them are responsible for the hydrolysis of the carbon–oxygen–carbon bonds that link the sugar residues in cellulose and hemicelluloses (6,7). Although aided by other enzymes, it is the glycoside hydrolases that degrade the main chains of these polysaccharides, thus potentially having the greatest impact on the conversion of lignocellulose. The discovery of efficient glycoside hydrolases and the development of optimal combinations of these enzymes are two important approaches in reducing the cost of bioconversion.

To support the discovery of novel biomass-degrading enzymes, an increasing number of genomes of lignocellulolytic fungi are being sequenced (8–19). This has resulted in numerous sequences, which are mostly annotated electronically or are without annotation. Current databases do not distinguish biochemically characterized data from electronically annotated data. Running a query sequence against one of these databases results in a long list of hits ranked

Database, Vol. 2011, Article ID bar020, doi:10.1093/database/bar020

according to highest percentage identity and coverage. Results must be sorted and individually evaluated to determine those electronically annotated from those whose function had been determined experimentally. To make the annotation process accurate and efficient, it is important to be able to easily link sequence information with the biochemically characterized properties of closely related sequences.

In this study, we have curated and annotated a comprehensive set of fungal genes encoding characterized GH family enzymes. This data set forms the basis of a searchable database of genes and their gene products, along with experimentally characterized biochemical properties, which is meant to be an ongoing, collaborative tool for fungal genome annotation and enzyme discovery.

Methods

Defining characterized glycoside hydrolases

For the purpose of this study, the term 'characterized glycoside hydrolase' refers to a protein that has satisfied the following criteria: (i) the gene sequence has been deposited in a public repository; (ii) the gene product has been assayed for a specific GH activity; and (iii) biochemical properties of the gene product have been reported in a peer-reviewed journal.

Literature survey

The EC Explorer on BRENDA [The Comprehensive Enzyme Information System (20)], http://www.brenda-enzymes.org/, was used as a guide for the different types of GH activities. The EC number 3.2.1, representing GH family enzymes, was selected on the Explorer. Under each 3.2.1.X, the table with the first column entitled 'Organism' was used as the starting point for collecting literature. Only literature associated with organisms of fungal origin were investigated further.

BRENDA provides either a direct link to the article on PubMed or cites the original publication. In the latter case, the Google Scholar 'Advanced Search' <http://scholar.google.com/advanced_scholar_search> was used to obtain the article of interest from another online resource. If an article was unobtainable through either PubMed or Google Scholar, a hard copy was ordered through an interlibrary loan system using the citation provided by BRENDA.

Once BRENDA was exhausted as a resource, PubMed was used. 'MyNCBI' was used to filter searches, keep track of the results and email to the curator any new additions that met the saved search criteria. Keyword searches were used to find articles of interest on PubMed. Each GH activity type listed on BRENDA was used as a keyword. Filters and limits were used to narrow the search results down to characterized enzymes of fungal origin.

Finding the sequence associated with the literature

If the sequences were available on GenBank (21), PubMed provided links to the gene and protein pages associated with the article. For articles from other sources and those PubMed articles without links to GenBank, the full text was searched for a sequence accession number and the associated database. For example, articles about glycoside hydrolases from the fungus *Rhizopus oryzae* usually refer to gene or protein identifiers from the Fungal Genome Initiative at the Broad Institute.

In some articles, the whole amino acid sequence was published but without an accession number. In these cases, the sequence was entered into BLASTp to search for a sequence ID in the appropriate database. UniProt (22) and GenBank were used as the default databases to search unless the species was known to have been sequenced by one of the major sequencing centers. A hit from the same organism having 100% identity and coverage with the query was considered a match.

On occasion, sequences were found by keyword search. Using the enzyme activity and name of the organism on GenBank or UniProt returned a list of hits. If the hit cited a published article of interest, the match was considered successful.

Cataloging characterized glycoside hydrolases of Fungal Origin

Data from published articles meeting our criteria of characterized glycoside hydrolases were organized in a spreadsheet format. The genes encoding the glycoside hydrolases were assigned unique identifiers. They were listed along the vertical rows and the data associated with the genes were recorded on the horizontal columns. Table 1 lists the titles of the columns included in the spreadsheet and the types of information described under each column. For the 'Literature' column, PubMed identification numbers (PMIDs) identified articles that described the characterization while literature that was not available on PubMed was identified by its DOI number or similar reference ID.

Assignment of standardized features

The Enzyme Commission (EC) and the Gene Ontology Project (GO) (23) <http://www.geneontology.org/index.shtml> developed EC numbers and GO terms, respectively. They are meant to standardize the functionality and characteristics of enzymes across all species. EC numbers were assigned based on the type of activity and substrate the enzyme acts on. GO terms are assigned based on the molecular function of the enzyme, the biological process that it acts in, and the cellular compartment where the enzyme is located.

Standardizing identifiers for genes and enzymes

Three-letter code for enzyme activity. In most of the articles, authors named genes using two- to three-letter

Table 1. The types of information extracted from the characterization literature

Column Number	Title	Type of information
1	Entry name	The unique identifier representing the enzyme. It incorporates the enzyme activity, the GH family it belongs to, and the phylogenetic origin of the enzyme.
2	Gene name	The assigned gene name based on the standardized naming convention adopted for this study (see 'Methods' section).
3	Species	The genus and species of the enzyme's natural host.
4	Strain	The strain of fungus used to obtain the gene and/or enzyme.
5	Gene name	The assigned gene name based on the standardized naming convention adopted for this study (see 'Methods' section).
6	Gene alias	Any other names the gene is referred to in the literature or sequence databases.
7	Enzyme name	The name most commonly used to identify an enzyme of a specific activity type.
8	Enzyme alias	Any other names the gene product (enzyme) is referred to in literature or public databases.
9	Systematic name	The systematic enzyme name according to the EC. <http://www.brenda-enzymes.org/>
10	The EC number	A numerical classification of enzymes based on the reactions they catalyze. <http://www.chem.qmul.ac.uk/iubmb/>
11	Gene ID (GenBank)	The nucleotide sequence ID issued by the GenBank database. <http://www.ncbi.nlm.nih.gov/>
12	UniProt ID	The ID issued to each protein in the UniProt database. <http://www.uniprot.org/>
13	Protein ID (GenBank)	The protein ID issued by the GenBank database. <http://www.ncbi.nlm.nih.gov/>
14	Characterization literature	The ID of the literature describing the enzyme's characterization and properties recorded as PMID (PubMed ID) or CSFGID (Centre for Structural and Functional Genomics).
15	Structure literature	The PMID or CSFGID of any literature describing the structure of the enzyme if available.
16	GH family	The GH family the enzyme belongs to. <http://www.cazy.org/>
17	Assay	The experiment used to determine the function and/or properties of the enzyme.
18	Activity assay conditions	The buffer, pH and temperature used in the assay to determine the enzyme activity
19	Kinetic assay conditions	The buffer, pH and temperature in which the K_m, k_{cat} and/or V_{max} were determined.
20	Substrates	The chemical substrates used to assay that the enzyme was assayed on.
21	Host (recombinant expression)	The organism used to produce the recombinant enzyme for the experimental assay.
22	Specific activity	The activity of the purified enzyme on the given substrate. Recorded in U/mg where 1 U (unit) = 1 μmol/min/mg = 16.67 nkat/mg.
23	Substrate specificity	The activity of the enzyme on a given substrate compared to other substrates tested. Expressed as a percentage with the highest activity usually equal to 100%.
24	K_m	The Michaelis–Menten constant (K_m) reflects the concentration of substrate at which initial velocity is one-half V_{max}. Recorded in millimolar (mM) or milligrams/milliliter (mg/ml).
25	k_{cat}	The maximum number of reactions the enzyme catalyzes in one second (s^{-1}).
26	V_{max}	The maximum velocity measured in U/mg at which an enzyme catalyzes a reaction. Reported in different ways, often as U/mg.
27	pH optimum	The pH at which enzyme activity is maximal highest.
28	pH stability	The pH range over which the enzyme is able to remain active. retains maximal activity (usually ≥80%) under the conditions defined in the paper
29	Temperature optimum	The temperature (°C) at which enzyme activity is maximal.
30	Temperature stability	The temperature (°C) beyond which the enzyme activity (usually ≥20%) is lost under the conditions defined in the study.
31	Isoelectric point (theoretical)	The pI of the enzyme calculated from its amino acid composition.
32	Isoelectric point (experimental)	The pI of the enzyme determined by isoelectric focusing.
33	Molecular weight (theoretical)	The molecular weight (kDa) of the enzyme calculated from its amino acid composition

(Continued)

Database, Vol. 2011, Article ID bar020, doi:10.1093/database/bar020

Table 1. Continued.

Column Number	Title	Type of information
34	Molecular weight (experimental)	The molecular weight (kDa) of the enzyme estimated using SDS–PAGE, gel filtration, etc.
35	Protein length	The number of amino acids in the enzyme before cleavage of the signal peptide (unless stated otherwise).
36	Signal peptide	The number of amino acids comprising the signal peptide, which targets the enzyme for secretion.
37	CBD	Carbohydrate binding domain if present as part of the enzyme.
38	Glycosylation	Type of glycosylation (only if experimentally determined)
39	Other features	Any other information regarding the enzyme's activity.
40	GO (molecular)	The GO term defining the molecular function of the enzyme
41	Evidence (molecular)	The type of information supporting the annotation of the molecular function of the enzyme.
42	GO (process)	The GO term defining the biological process the enzyme participates in.
43	Evidence (process)	The type of information supporting the annotation of the biological process. The biological process is only assigned the evidence code 'Inferred by Direct Assay (IDA)' when assayed on its natural substrate
44	GO (component)	The GO term defining the cellular compartment in which the enzyme acts.
45	Evidence (component)	The type of information supporting the enzyme's component annotation.

This table lists the names of the columns used to organize the collected data in a spreadsheet. The type of data each heading encompasses is explained on the right.

codes representing the activity of the encoded protein followed by an assigned number or letter to distinguish each one from others of the same function or from the same species. Sometimes, several different letter codes have been used for the same enzyme activity. For example, '*xyn*' (24), '*xyl*' (25) and '*xln*' (26) have all been used to describe xylanases. In other cases, the same letter code was used for enzymes with different activities. For example, '*cel*' has been used for endoglucanase (27), xyloglucanase (28), β-glucosidase (29) and cellobiohydrolase (30) activities. To avoid confusion, we have adopted a single three-letter code for each enzyme activity; for example, *xyn* for endo-1,4-β-xylanase (xylanase). The most commonly used codes for GH family enzymes in the literature were adopted as the unique codes (Table 2). In the case of bifunctional enzymes, where two functional domains can clearly be discerned by sequence analysis, the three-letter code would start with 'z' followed by two letters representing the functional domains of the protein. An enzyme carrying an α-arabinofuranosidase domain and a xylosidase domain, for example, would be called 'zax'.

Gene name. The following format was used to standardize the assignment of gene names. The three-letter code of the enzyme activity is followed by a number, which represents the GH family to which the enzyme belongs. Finally, a letter is added to distinguish the different genes of the same species encoding the same enzyme function from the same family. If the gene name given in the literature included a letter, that letter was kept in the standardized name. If the given gene name included a number, it was converted to the corresponding letter. For example, *xyn2* from GH family 11 would become *xyn11B*, while *bgl5* from GH family 3 would have become *bgl3E* and so on. When the same gene name had been given to multiple genes from the same species and family, their sequences were aligned to make sure they were the same sequence. If the genes were found to encode different enzymes, the letter component of the gene name was assigned according to the publication date of the literature. Thus, the letter 'A' (or the first available letter if 'A' was taken) represents the enzyme with the earliest published characterization data.

Entry identifier. To make each gene entry unique, a naming method similar to that of UniProt was used. A five-letter code representing the natural host of the enzyme was added onto the end preceded by an underscore. The first three letters were used to represent the genus of the fungus, followed by two letters representing the species (Table 3). For example, XYN11A_TRIRE would represent the GH11 xylanase gene, '*xynA*', from *Trichoderma reesei*. If the letters were the same for different species, *Penicillium janthinellum* and *Penicillium janczewskii,* for example (PENJA), another unique letter from the species name was used. In the case of *Penicillium*

Table 2. Activities of the characterized GH

Enzyme name	Code	Activity
α-1,2-mannosidase (2-α-mannosyl-oligosaccharide α-D-mannohydrolase)	MSD	Catalyzes the hydrolysis of terminal, non-reducing-end glucose in mannosyl-oligosaccharides
α-amylase (4-α- D-glucan glucanohydrolase)	AMY	Catalyzes the hydrolysis of internal α-1,4-glucosidic linkages in polysaccharides and releases products in α-configuration
α-arabinofuranosidase (α-L-arabinofuranoside arabinofuranohydrolase)	ABF	Catalyzes terminal, non-reducing-end hydrolysis of α-L-arabinofuranoside residues
α-galactosidase (α-D-galactoside galactohydrolase)	MEL	Catalyzes the hydrolysis of non-reducing-end α-D-galactose residues
α-glucosidase (4-α-D-glucohydrolase)	AGL	Releases glucose by catalyzing the hydrolysis of non-reducing-end α-D-glycosidic links
α-glucuronidase (α-D-glucosiduronate glucuronohydrolase)	AGU	Catalyzes the hydrolysis of glucuronic acid branches from hemicellulose
α-L-rhamnosidase (α-L-rhamnoside rhamnohydrolase)	RHA	Catalyzes the hydrolysis of non-reducing-end α-L-rhamnoside residues
α-xylosidase	AGD	Catalyzes the hydrolysis of terminal α-linked xylosides
Arabinogalactanase (arabinogalactan 4-β-D-galactanohydrolase)	GAN	Catalyzes the hydrolysis of internal β-1,4-linked galactosidic linkages
Arabinoxylan–arabinofuranosidase	AXH	Catalyzes the removal of arabinosides from xylan main chains
β-galactosidase (β-D-galactoside galactohydrolase)	LAC	Catalyzes the hydrolysis of terminal, non-reducing-end β-D-galactose residues
β-glucosidase (β-D-glucoside glucohydrolase)	BGL	Releases glucose by acting on terminal, non-reducing-end β-D-glucosidic links
Beta-mannanase (4-β-D-mannan mannanohydrolase)	MAN	Catalyzes the hydrolysis of β-1,4-mannosidic linkages in mannans, galactomannans and glucomannans
β-mannosidase (β-D-mannoside mannohydrolase)	MND	Catalyzes the hydrolysis of terminal, non-reducing-end β-D-mannose from β-D-mannosides
β-xylosidase (4-β-D-xylan-xylohydrolase)	XYL	Catalyzes the hydrolysis of the bond joinholding xylose sugars together in xylobiose
Cellobiohydrolase (4-β-D-glucan cellobiohydrolase)	CBH	Acts on non-reducing-end 1,4-β-D-glucosidic linkages to release cellobiose
Cellulase-enhancing protein	CEP	Exact function unknown but enhances hydrolysis of cellulose by cellulases
Chitinase ((1-4)-2-acetamido-2-deoxy-β-D-glucan glucanohydrolase)	CHI	Catalyzes the random hydrolysis of *N*-acetyl-β-D-1,4-glucoaminide
Chitosanase (chitosan *N*-acetylglucosaminohydrolase)	CSN	Catalyzes the hydrolysis of β-1,4 linkages in acetylated chitosans
Dextranase (6-α-D-glucan 6-glucanohydrolase)	DEX	Acts on 1,6-α-glucosidic linkages in dextrins
Endo-arabinanase (5-α-L-arabinan 5-α-L-arabinanohydrolase)	ABN	Catalyzes the hydrolysis of internal α-1,5-arabinofuranosidic linkages in arabinans
Endo-β-1,6-glucanase (6-β-D-glucan glucanohydrolase)	BGN	Catalyzes the random hydrolysis of β-1,6 linkages in β-1,6-linked glucans
Endo-β-*N*-acetylglucosaminidase	END	Catalyzes the removal of acetylated glycoprotein branches forming mannosyl-oligosaccharides
Endo-inulinase (1-β-D-fructan fructanohydrolase)	INU	Catalyzes the hydrolysis of internal fructosidic linkages in inulin
Endo-polygalacturonase (1,4-α-D-galacturonan glycanohydrolase)	PGA	Catalyzes the random hydrolysis of 1,4-α-galactosiduronic linkages in pectate and galacturonans
Endo-rhamnogalacturonase	RHG	Catalyzes the hydrolysis of links between galacturonic acid and rhamnopyranosyl residues in pectins
Endoglucanase (4-β-D-glucan 4-glucanohydrolase)	EGL	Catalyzes the hydrolysis of β-1,4-glucosidic linkages in cellulose
Exo-1,3-β-glucanase (3-β-D-glucan glucohydrolase)	EXG	Catalyzes the hydrolysis of glucose from the non-reducing-ends of β-1,3-glucans
Exo-arabinanase	ARB	Catalyzes the hydrolysis of α-1,5-arabinofuranosidic linkages from the ends of arabinans

(Continued)

Database, Vol. 2011, Article ID bar020, doi:10.1093/database/bar020

Table 2. Continued.

Enzyme name	Code	Activity
Exo-glucosaminidase (Chitosan exo-1,4-β-D-glucoaminidase)	GLS	Catalyzes the hydrolysis of glucosamine residues from the non-reducing ends of chitosans
Exo-inulinase (β-D-fructan fructohydrolase)	INX	Catalyzes the hydrolysis of terminal, non-reducing 2,1- and 2,6-linked fructofuranose in fructans
Exo-polygalacturonase (poly{1,4-α-D-galacturonide} galacturonohydrolase)	PGX	Catalyzes the hydrolysis of D-galacturonate from the ends of galacturonides
Exo-rhamnogalacturonase	RGX	Catalyzes the hydrolysis of rhamnoside residues from the ends of pectin
Galactanase (galactan endo-1,6-β-galactosidase)	GAL	Catalyzes the hydrolysis of internal β-1,6-galactosidic linkages in arabinopgalactans and the hydrolysis of β-1,3- and β-1,6-galactosidic linkages in mixed galactans
Hexosaminidase (β-N-acetyl-D-hexosaminide *N*-acetylhexosaminohydrolase)	HEX	Catalyzes the hydrolysis of terminal, non-reducing-end *N*-acetyl-D-hexosamine residues
Invertase (β-D-fructofuranoside fructohydrolase)	SUC	Catalyzes the hydrolysis of β-D-fructofuranoside from the non-reducing ends of fructofuranosides
Isopullulanase (pullulan 4-glucanohydrolase)	IPU	Catalyzes the hydrolysis of pullulan to isopanose
Laminarinase (3-β-D-glucan glucanohydrolase)	LAM	Catalyzes the hydrolysis of β-1,3-glucosidic linkages in β-1,3-glucans
Licheninase (1,3-, 1,4-β-D-glucan 4-glucanohydrolase)	LIC	Catalyzes the hydrolysis of β-1,4-glucosidic linkages in mixed-link glucans
Mixed-link glucanase (3(or 4)-β-D-glucan 3(4)-glucanohydrolase)	MLG	Catalyzes the hydrolysis of β-1,3 or β-1,4 linkages in mixed glucans when the glucose involved in the linkage is substituted at the 1,3 position
Mutanase (3-α-D-glucan 3-glucanohydrolase)	MUT	Catalyzes the internal hydrolysis of α-1,3-glycosidic linkages
Oligo-1,6-glucosidase (oligosaccharide 6-α-glucohydrolase)	OGL	Catalyzes the hydrolysis of 1,6-glycosidic linkages in oligosaccharides
Oligoxyloglucan cellobiohydrolase (oligoxyloglucan reducing-end cellobiohydrolase)	XBH	Catalyzes the hydrolysis of cellobiose from the reducing ends of xyloglucans with *O*-6 xylosyl substitutions on the second residue
Trehalase (α, α-trehalose glucohydrolase)	TRE	Catalyzes the hydrolysis of trehalose to release two D-glucose residues
Xylanase (4-β-D-xylan xylanohydrolase)	XYN	Acts on 1,4-β-xylosidic linkages in xylan
Xylogalacturonase	XGH	Catalyzes the hydrolysis of xylosyl substitutions on pectins
Xyloglucanase ([(1-6)-α-D-xylo]-(1-4)-β-D-glucan glucanohydrolase)	XEG	Catalyzes the hydrolysis of bonds involved in xyloglucan chains

This table lists the different enzyme activities collected in the literature survey. A combination of BRENDA <http://www.brenda-enzymes.org/> and The GO <http://www.geneontology.org/> were used to give a definition of each activity type and alternate names. The common, simpler enzyme name is used followed by the systematic name used by BRENDA. A three-letter code was used to represent the activity of the enzyme in the gene name and entry name. Codes were selected based on the most commonly used code for a particular activity in the literature. These codes were used in the standardized naming process.

janthinellum and *Penicillium janczewskii*, the entries would be PENJA and PENJZ, respectively.

Results

Using the procedures described in the 'Methods' section, we have collected a total of 453 characterized GH enzymes of fungal origin. They come from 131 different fungal species (Table 3), most of which are from the phylum Ascomycota. The genus *Aspergillus* encompasses the largest number of characterized GH proteins with *Aspergillus niger* in the lead accounting for 47 enzymes. The collected enzymes represent 49 different GH activities and cover 44 of the GH families described in CAZy (1–5) (Table 4). All of these enzymes were extracellular, with 443 enzymes as soluble extracellular proteins and only 6 that were shown to attach to the external cell wall.

Distribution of enzyme activities

Cellulases comprise ~27% of the characterized GH in this database. Collectively, they cover nine different GH families. The 63 endoglucanases represent activity type that has the most published characterization data. The majority belongs to GH5. The cellobiohydrolases

Table 3. Fungal species having characterized glycoside hydrolases

	Code	Number of enzymes characterized	Alternate names
Ascomycota species			
Acremonium blochii	ACRBL	1	
Acrophialophora nainiana	ACRNA	1	
Aphanocladium album	APHAL	1	
Arxula adeninivorans	ARXAD	2	
Aspergillus aculeatus	ASPAC	9	
Aspergillus awamori	ASPAW	11	
Aspergillus flavus	ASPFL	2	
Aspergillus fumigatus	ASPFU	5	*Sartorya fumigata*
Aspergillus kawachii	ASPKA	11	*Aspergillus awamori var. kawachii*
Aspergillus niger	ASPNG	47	
Aspergillus oryzae	ASPOR	16	
Aspergillus phoenicis	ASPPH	1	*Aspergillus saitoi*
Aspergillus shirousami	ASPSH	2	
Aspergillus sojae	ASPSO	1	
Aspergillus species	ASPSP	1	
Aspergillus sulphureus	ASPSU	2	
Aspergillus terreus	ASPTE	2	
Aspergillus tubingensis	ASPTU	6	
Aureobasidium pullulans	AURPU	4	
Bionectria ochroleuca	BIOOC	4	*Gliocladium roseum*
Bispora sp. MEY-1	BISSP	1	
Botryotinia fuckeliana	BOTFU	7	*Botrytis cinerea, Noble-rot fungus*
Candida albicans	CANAL	6	
Candida oleophila	CANOL	1	
Candida tsukubaensis	CANTS	1	
Candida wickerhamii	CANWI	1	
Chaetomium brasiliense	CHABR	1	
Chaetomium gracile	CHAGR	2	
Chaetomium thermophilum	CHATH	1	
Claviceps purpurea	CLAPU	2	
Coccidioides immitis	COCIM	2	*Valley Fever Fungus*
Cochliobolus carbonum	COCCA	8	*Bipolaris zeicola*
Cochliobolus sativus	COCSA	1	*Bipolaris sorokinia*
Cryphonectria parasitica	CRYPA	1	*Endothia parasitica, Chestnut Blight Fungus*
Daldinia eschscholzii	DALES	1	
Debaryomyces occidentalis	DEBOC	3	
Emericella desertorum	EMEDE	1	
Emericella nidulans	EMENI	34	*Aspergillus nidulans*
Fusarium equiseti	FUSEQ	1	*Fusarium scirpi*
Fusarium oxysporum	FUSOX	2	*Panama Disease Fungus*
Fusarium solanii	FUSSO	3	*Nectria ipomoeae*
Geotrichum species	GEOSP	2	*Fermentotrichon, Oosporoidea, Polymorphomyces*
Gibberella species 75	GIBSP	1	
Gibberella zeae	GIBZE	3	*Fusarium graminearum, Wheat Head Blight Fungus*

(Continued)

Database, Vol. 2011, Article ID bar020, doi:10.1093/database/bar020

Table 3. Continued.

	Code	Number of enzymes characterized	Alternate names
Hansenula anomala	HANAN	1	*Candida pelliculosa*
Hormoconis resinae	HORRE	1	*Creosote fungus, Amorphotheca resinae*
Humicola grisea var. thermoidea	HUMGT	4	
Humicola insolens	HUMIN	6	
Hypocrea schweintzii	HYPSC	1	
Isaria javanicus	ISAJA	1	*Paecilomyces javanicus*
Kluyveromyces lactis	KLULA	3	*Candida sphaerica*
Kluyveromyces marxianus	KLUMA	2	*Candida kefyr*
Kuraishia molischiana	KURMO	1	*Pichia capsulata*
Lipomyces konoenkoe	LIPKO	2	
Lipomyces starkeyi	LIPST	1	*Oleaginous yeast*
Magnaporthe grisea	MAGGR	5	*Pyricularia grisea, Rice Blast Fungus*
Melanocarpus albomyces	MELAO	3	
Metarhizium anisopliae	METAN	1	
Neotyphodium species	NEOSP	1	
Neurospora crassa	NEUCR	1	
Paecilomyces thermophila	PAETH	1	
Penicillium brasilianum	PENBR	1	
Penicillium canescens	PENCA	1	
Penicillium chrysogenum	PENCH	3	*Penicillium notatum*
Penicillium citrinum	PENCI	3	
Penicillium enchinulatum	PENEN	1	
Penicillium funiculosum	PENFN	6	
Penicillium janthinellum	PENJA	2	*Penicillium vitale*
Penicillium minioluteum	PENMI	2	
Penicillium olsonii	PENOL	2	
Penicillium purpurogenum	PENPU	6	
Penicillium simplicissimum	PENSI	1	
Penicillium species	PENSQ	8	
Periconia species	PERSP	1	
Pichia angusta	PICAN	2	*Hansenula polymorpha*
Pichia jadinii	PICJA	1	*Candida utilis*
Pichia pastoris	PICPA	1	
Robillarda species	ROBSP	1	
Saccharomyces cerevisiae	YEAST	9	*Baker's Yeast*
Saccharomycopsis fibuligera	SACFI	3	
Schizosaccharomyces pombe	SCHPO	5	
Stachybotrys echinata	STAEC	1	
Talaromyces emersonii	TALEM	2	
Thermoascus aurantiacus	THEAU	5	
Thermomyces lanuginosus	THELA	2	*Humicola lanuginosa*
Thielavia heterothallica	THIHE	1	
Thielavia terrestris	THITE	1	*Acremonium alabamense*
Trichoderma asperellum	TRIAS	3	
Trichoderma harzianum	TRIHA	11	*Hypocrea lisii*
Trichoderma koningii	TRIKO	2	*Hypocrea koningii*

(Continued)

Table 3. Continued.

	Code	Number of enzymes characterized	Alternate names
Trichoderma longibrachiatum	TRILO	1	
Trichoderma reesei	TRIRE	23	*Hypocrea jecorina*
Trichoderma species	TRISP	1	
Trichoderma viride	TRIVI	2	
Verticillium dahliae	VERDA	1	*Verticillium Wilt Fungus*
Yarrowia lipolytica	YARLI	1	*Candida lipolytica*
Basidiomycota species			
Agaricus bisporus	AGABI	4	*Common Mushroom*
Athelia rolfsii	ATHRO	2	*Sclerotinia rolfsii, Corticium rolfsii*
Chondrostereum purpureum	CHOPU	1	*Stereum purpureum*
Coprinopsis cinerea	COPCI	3	*Hormographiella aspergillata, Inky Cap Fungus*
Cryptococcus albidus	CRYAL	1	*Fiobasidium floriforma*
Cryptococcus flavus	CRYFL	1	
Cryptococcus species	CRYSP	2	
Fomitopsis palustris	FOMPA	2	
Fomitopsis pinicola	FOMPI	1	
Irpex lacteus	IRPLA	6	*Polyporus Tulipoferae, Milk-white Toothed Polypore*
Meripilus giganteus	MERGI	1	
Phaffia rhdozyma	PHARA	2	*Xanthophyllomyces dendrohous*
Phanerochaete chrysosporium	PHACH	11	*Sporotrichum prunosum*
Schizophyllum commune	SCHCO	1	*Bracket Fungus*
Sporobolomyces singularis	SPOSI	1	
Trametes hirsuta	TRAHI	1	
Uromyces fabae	UROFA	1	*Rust Fungus*
Mucoromycotinia species			
Gongronella species	GONSP	1	
Mortierella alliacea	MORAL	1	
Mucor circinelloides	MUCCI	2	*Mucor griseo-roseus*
Mucor hiemalis	MUCHI	1	
Mucor javanicus	MUCJA	1	
Mycocladus corymbiferus	MYCCO	1	*Absidia corymbiferus*
Phycomyces nitens	PHYNI	1	
Rhizopus oligosporus	RHIOL	1	
Rhizopus oryzae	RHIOR	17	*Rhizopus delemar*
Rhizopus species	RHISP	1	
Syncephalastrum racemosum	SYNRA	1	
Neocallimastigomycota species			
Neocallimastix frontalis	NEOFR	1	
Neocallimastix patriciarum	NEOPA	4	
Orpinomyces joyonii	ORPJO	1	
Orpinomyces species	ORPSP	3	
Piromyces equi	PIREQ	2	
Piromyces species	PIRSP	7	

This table lists the species of fungi and the number of characterized glycoside hydrolases collected from each. They are listed according to phylum. The species name used in this research is listed on the left, while any other names used for the same species are listed under 'Alternative Names'. The five-letter codes used for the standardized naming of genes are also listed here. The codes follow a naming system used by UniProt. The first three letters represent the genus and the last two letters represent the species of the fungus.

Database, Vol. 2011, Article ID bar020, doi:10.1093/database/bar020

Table 4. GH families having characterized enzymes of fungal origin

Family	Number of Enzymes Characterized	Activity
GH1	7	β-glucosidase (7)
GH2	5	β-mannosidase (2), chitosanase (1), exo-glucosaminidase (1), β-galactosidase (1)
GH3	30	β-glucosidase (22), β-xylosidase (8)
GH5	45	Endoglucanase (22), exo-1,3-β-glucanase (12), β-mannanase (8), galactanase (2), endo-1,6-β-glucanase (1)
GH6	12	Cellobiohydrolase (11), endoglucanase (1)
GH7	29	Cellobiohydrolase (18), endoglucanase (10), xylanase (1)
GH9	1	Endoglucanase (1)
GH10	19	Xylanase (19)
GH11	44	Xylanase (44)
GH12	24	Endoglucanase (20), xyloglucanase (3), licheninase (1)
GH13	10	α-glucosidase (3), α-amylase (6), oligo-1,6-glucosidase (1)
GH15	14	Glucoamylase (14)
GH16	5	Mixed-link glucanase (3), laminarinase (1), licheninase (1)
GH17	2	Laminarinase (1), exo-1,3-β-glucanase (1)
GH18	13	Chitinase (13)
GH20	6	Hexosaminidase (6)
GH26	3	β-mannanase (3)
GH27	6	α-galactosidase (6)
GH28	54	Endo-polygalacturonase (40), exo-polygalacturonase (9), endo-rhamnogalacturonase (3), exo-rhamnogalacturonase (1), xylogalacturonase (1)
GH30	3	Endo-1,6-β-glucanase (3)
GH31	10	α-glucosidase (8), α-xylosidase (1), invertase (1)
GH32	22	Invertase (10), exo-inulinase (7), endo-inulinase (4)
GH35	1	β-galactosidase (1)
GH36	7	α-galactosidase (7)
GH37	4	Trehalase (4)
GH43	6	Endo-1,5-α-arabinanase (3), α-L-arabinofuranosidase (2), β-xylosidase (1)
GH45	8	Endoglucanase (8)
GH47	5	α-1,2-mannosidase (5)
GH49	4	Dextranase (3), isopullulanase (1)
GH51	5	α-L-arabinofuranosidase (5)
GH53	6	Arabinogalactanase (6)
GH54	9	α-L-arabinofuranosidase (9)
GH55	4	Exo-1,3-β-glucanase (3), laminarinase (1)
GH61	3	Cellulase-enhancing protein (3)
GH62	2	Arabinoxylan arabinofuranosidase (2)
GH65	2	Trehalase (2)
GH67	4	α-glucuronidase (4)
GH71	3	Mutanase (3)
GH74	6	Xyloglucanase (3), oligoxyloglucan cellobiohydrolase (2), endoglucanase (1)
GH75	2	Chitosanase (2)
GH78	3	α-rhamnosidase (3)
GH81	2	Laminarinase (2)
GH85	1	*N*-acetylglucosaminidase (1)
GH93	2	Exo-arabinanase (2)
Total	**453**	

The GH families and the total number of biochemically characterized enzymes collected for each are listed here. The GH families that did not have any characterized fungal enzymes were not included. The column titled 'Activity' shows the different types of activities the collected enzymes from each GH have. The numbers in brackets show the distribution of activity types in each family.

Table 5. Some properties of characterized fungal cellulases

Activity	GH family	Total	Optimal pH	Optimal temperature (°C)	Temperature stability (°C)	Mass (kDa)
β-glucosidase	GH1	7	3.5–6.3	40–55	40–55	52–94
	GH3	22	3.5–8.0	37–72	30–70	74–145
Cellobiohydrolase	GH6	11	4.8–9.0	40–50	30–60	40–60
	GH7	18	5.0–6.0	35–65	50–55	47–90
Endoglucanase	GH5	22	3.5–8.5	40–75	37–80	35–56
	GH6	1	5.5	NA	NA	38
	GH7	10	4.0–5.5	452–57	40–50	46–56
	GH9	1	NA	NA	NA	90
	GH12	20	2.0–5.0	55–70	40–55	25–32
	GH45	8	5.0–7.0	30–65	60	20–47
	GH74	1	4.5	55	30	90
Cellulase-enhancing proteins	GH61	3	4.0–6.0	NA	NA	36–56

The three different types of cellulase activities; β-glucosidase, endoglucanase and cellobiohydrolase are listed here. Cellulose-enhancing proteins have been included as well. A range of the biochemical properties for each activity type is presented according to GH family. 'NA' indicates that the information was not available. The summarized data presented in the table were compiled from this study.

come from GH6 and GH7 with 11 and 18 proteins, respectively. The β-glucosidases also cover two families with 7 from GH1 and 22 from GH3.

With 64 entries, xylanases have the highest number characterized of any activity type. The xylanases come from GH families 10 and 11 for the most part. A single xylanase, XYN7A_PENFN, contains a GH7 domain (31,32), which is unusual. No other literature in the collection described a xylanase from GH7. Other xylan-active enzymes collected included β-xylosidases, arabinoxylan arabinofuranosidases and α-glucuronidases.

Several different types of arabinan-active enzymes are represented in the collection. There are three endo-arabinanases and two exo-arabinanases, belonging to GH43 and GH93, respectively. We also collected 17 enzymes having α-arabinofuranosidase activity. These enzymes act on arabinans as well as arabinosyl side chains attached to other polysaccharides. They are divided among GH families 43, 51 and 54.

The mannan-active enzymes collected in the greatest number are the β-mannanases. There are 11 enzymes with 8 from GH5 and 3 from GH26. Some showed high thermostability having optimal temperatures as high as 79°C. β-Mannosidases also act on the main chain of mannans. Two β-mannosidases, both from GH2, were collected.

Mannans may have other sugars such as glucose incorporated into the backbone or galactose present in side chains. Enzymes with different specificities are required to hydrolyze these residues. One example of these types of enzymes is α-galactosidase, which is represented in the database by six enzymes from GH27 and seven from GH36.

Pectin, another common lignocellulose polymer, can occur in a variety of forms. It is mostly composed of galacturonic acids, which can alternate with rhamnogalactuonans in the main chain, or have branches composed of a variety of different residues. All of the glycoside hydrolases active on the main chains of pectins are from GH 28. There are 54 in total with 40 having endo-polygalacturonase activity, 9 having exo-polygalacturonase activity, 3 having endo-rhamnogalacturonase activity and 1 each of exo-rhamnogalacturonase and xylogalacturonase.

Chitin, inulin and starch are also components of biomass. The two major enzymes active on chitin are chitinases and chitosanases. Characterized chitinases are the more numerous of the 2 with 13 in the database, all of which are from GH18. Three chitosanases were collected; one from GH2 and two from GH75. The characterized inulin-active enzymes mostly come from GH32 except for one invertase with a GH31 domain. This invertase along with the GH32 invertases and endo- and exo-inulinases add up to a total of 21 inulin-active enzymes. There is a greater variety of glycoside hydrolases active on starch compared to chitin and inulin. Some of these include α-amylase, glucoamylase, oligo-1,6-glucosidase, dextranase and trehalase.

A variety of enzymes active on non-cellulosic β-glucans were collected as well. The most abundant are the 16 *exo*-1,3-β-glucanases from GH families 5, 17 and 55.

We have cataloged a limited number of characterized enzymes for other GH activity types including α-rhamnosidase, oligoxyloglucan cellobiohydrolase, mutanase, β-galactosidase, α-1,2-mannosidase, endo-*N*-acetylglucosaminidase and galactanase to name a few. For more details

Database, Vol. 2011, Article ID bar020, doi:10.1093/database/bar020

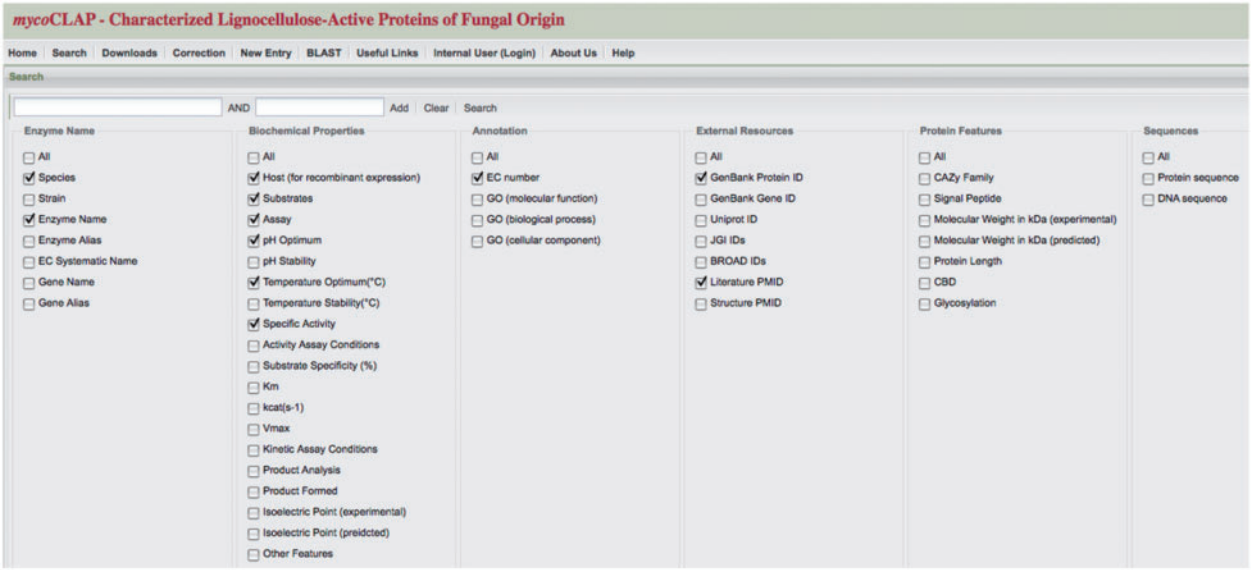


Figure 1. *myco*CLAP search page. The main search page from *myco*CLAP is shown here. Keywords such as enzyme activity, glycoside hydrolase family or a substrate name can be entered as search terms. Leaving the field blank and clicking on 'search' allows a user to browse the database. The information recorded during the curation process has been divided into six categories shown here; Enzyme Name, Biochemical Properties, Annotation, External Resources, Protein Features and Sequence. By checking boxes under these categories a user can determine which types of information will be displayed on the results page. The default settings are shown here. The tabs along the top allow quick and easy navigation through all of *myco*CLAP's features.

and properties of the individual enzymes, and the data and literature on these 453 characterized GH, see *myco*CLAP <http://mycoCLAP.fungalgenomics.ca/>.

The mycoCLAP database

The *myco*CLAP collection of fungal enzymes is searchable by BLAST alignment using a query sequence or by keywords. A BLAST search will display entries most similar to the query while a keyword search displays results in a tabular format (Figure 1). The results table lists enzymes by their unique entry name followed by the corresponding data. Results can be filtered by selecting or deselecting specific entries in the search page. A browsing option is also possible by leaving the keyword fields blank and clicking on the 'Search' button. Selecting an entry name leads to the gene page containing all the data, sequences and literature related to the enzyme (Figure 2).

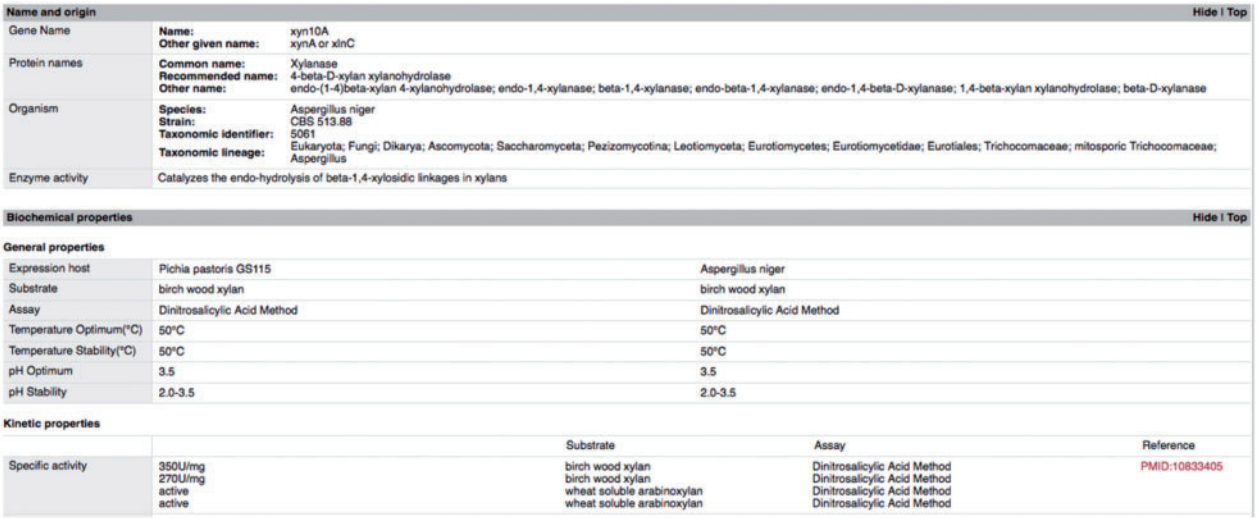
The download option allows a user to download data text and/or sequences in fasta format. The types of data to be downloaded are selected in the same way as the keyword search. Individual enzymes, or a subset of them, can be selected for download by using the check boxes on the left of the results table. *myco*CLAP also provides a list of resources for various annotation tools and ongoing sequencing projects. They are listed under the tab 'Useful Links' along with a short description.

Users are encouraged to add new entries and make corrections to existing entries using the 'New Entry' and 'Correction' forms. A curator will review each submission before any changes or additions are made to the database.

The database CAZy is a well-maintained and comprehensive resource for carbohydrate-active enzymes while *myco*CLAP in its current stage of development with the focus on fungal glycoside hydrolases is far less comprehensive. The major difference between these two databases is that *myco*CLAP contains only sequences whose products have been biochemically characterized, whereas CAZy includes sequences with predicted function. *myco*CLAP also provides a BLAST resource, an important tool in the annotation of novel sequences, and this feature is not available in CAZy. Using BLAST to search only the characterized enzyme database allows the closest related sequence with experimentally documented characteristics to be rapidly located. It also provides relevant structural and biochemical data extracted from published literature in an easy-to-view format. It should be noted that *myco*CLAP is focused on natural diversity and does not yet include engineered or evolved variants of naturally occuring fungal enzymes.

Conclusion

Characterized glycoside hydrolases were identified from literature obtained through BRENDA, PubMed, Google Scholar and myNCBI. Their properties were collected in a spreadsheet and the corresponding gene and protein sequences were collected from GenBank and UniProt.



Name and origin Hide | Top

Gene Name	**Name:**	xyn10A
	Other given name:	xynA or xlnC
Protein names	**Common name:**	Xylanase
	Recommended name:	4-beta-D-xylan xylanohydrolase
	Other name:	endo-(1-4)beta-xylan 4-xylanohydrolase; endo-1,4-xylanase; beta-1,4-xylanase; endo-beta-1,4-xylanase; endo-1,4-beta-D-xylanase; 1,4-beta-xylan xylanohydrolase; beta-D-xylanase
Organism	**Species:**	Aspergillus niger
	Strain:	CBS 513.88
	Taxonomic identifier:	5061
	Taxonomic lineage:	Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomyceta; Pezizomycotina; Leotiomyceta; Eurotiomycetes; Eurotiomycetidae; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus
Enzyme activity	Catalyzes the endo-hydrolysis of beta-1,4-xylosidic linkages in xylans	

Biochemical properties Hide | Top

General properties

Expression host	Pichia pastoris GS115	Aspergillus niger
Substrate	birch wood xylan	birch wood xylan
Assay	Dinitrosalicylic Acid Method	Dinitrosalicylic Acid Method
Temperature Optimum(°C)	50°C	50°C
Temperature Stability(°C)	50°C	50°C
pH Optimum	3.5	3.5
pH Stability	2.0-3.5	2.0-3.5

Kinetic properties

		Substrate	Assay	Reference
Specific activity	350U/mg	birch wood xylan	Dinitrosalicylic Acid Method	PMID:10833405
	270U/mg	birch wood xylan	Dinitrosalicylic Acid Method	
	active	wheat soluble arabinoxylan	Dinitrosalicylic Acid Method	
	active	wheat soluble arabinoxylan	Dinitrosalicylic Acid Method	

Figure 2. Gene page example. This screen shot illustrates the set-up and types of data available on *myco*CLAP. This is part of the gene page for the glycoside hydrolase XYN10A_ASPNG (a family 10 xylanase from *Aspergillus niger*). The 'Names and Origin' section includes any names or abbreviations used to identify the enzyme in the literature or on other databases. A search by any of those names will deliver this enzyme as a hit. The next section contains biochemical properties extracted from the literature. This entry has the enzyme's specific activity, pH optima and temperature optima on birch wood xylan when expressed from two different hosts. Other information recorded on gene pages includes nucleotide and amino acid sequences, protein domains, assay conditions, enzyme family, literature citations and other features recorded in the literature that make the entry unique.

Standardized functional annotations from The GO and The EC were assigned based on findings from the literature. The collected data and assigned annotations were then deposited in *myco*CLAP.

The *myco*CLAP database is intended to facilitate the annotation of glycoside hydrolases active in the decomposition of plant biomass by providing a mechanism for comparison of novel sequences to a set of sequences whose gene products have been characterized. Such comparisons should result in decreased occurrence of false positives in searching for homologs, shortened times for the sorting process and expedition of the identification of targets to guide experimental analysis.

The curation of characterized GH data is an ongoing project that will be continually updated and expanded. Currently, characterized carbohydrate esterases, polysaccharide lyases and lipases involved in biomass degradation are being curated for incorporation into *myco*CLAP. Future work will include the collection of other characterized lignocellulose-active enzymes such as peroxidases, cellobiose dehydrogenases, proteases as well as engineered versions of these characterized enzymes. Using the methods outlined, we believe that we have exhausted the literature describing fungal glycoside hydrolysis. However, there is no way to be sure that all the relevant literature has been collected. With the efficient breakdown of biomass for the production of biofuels and bioproducts becoming more and more important, new information will constantly become available. With the continuing contribution from our curators and the help of submissions from other researchers in the field, the database will be regularly updated and thus provide the fungal research community with the latest and most comprehensive collection of knowledge and data.

While *myco*CLAP offers a detailed, searchable database that can be used to survey and rapidly locate information on characterized, sequenced, lignocellulose-active enzymes, it is important to recognize the limitations it has as a comparative tool. Some parameters such as temperature and pH optima for different enzymes can, with some care, be compared with one another. Others, such as V_{max} and K_m, which are dependent on parameters such as temperature and pH, were not always performed under optimal conditions. For these parameters, the user should refer to the original articles for additional details. Finally, it is not often appreciated that reducing sugar assays can give quite different results depending on the method used (33), as can protein assays. Hence, it is very difficult to compare the specific activities of different enzyme preparations unless the same activity and protein assays were used.

Acknowledgement

The authors would like to thank Vineet Dua for his contribution to editing the collected data.

Funding

Cellulosic Biofuel Network of the Agricultural Bioproducts Innovation Program of Agriculture and Agri-Food Canada;

Genome Canada and Génome Québec. Funding for open access charge: Genome Canada.

Conflict of interest. None declared.

References

1. Henrissat,B. and Davies,G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.*, **7**, 637–644.
2. Davies,G. and Henrissat,B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853–859.
3. Henrissat,B. and Bairoch,A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.*, **316**(Pt 2), 695–696.
4. Henrissat,B. and Bairoch,A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, **293**(Pt 3), 781–788.
5. Henrissat,B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, **280**(Pt 2), 309–316.
6. Lundell,T.K., Makela,M.R. and Hilden,K. (2010) Lignin-modifying enzymes in filamentous basidiomycetes–ecological, functional and phylogenetic review. *J. Basic Microbiol.*, **50**, 5–20.
7. Sanchez,C. (2009) Lignocellulosic residues: biodegradation and bioconversion by fungi. *Biotechnol. Adv.*, **27**, 185–194.
8. Coleman,J.J., Rounsley,S.D., Rodriguez-Carres,M. *et al*. (2009) The genome of Nectria haematococca: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet.*, **5**, e1000618.
9. Ellwood,S.R., Liu,Z., Syme,R.A. *et al*. (2010) A first genome assembly of the barley fungal pathogen Pyrenophora teres f. teres. *Genome Biol.*, **11**, R109.
10. Magrini,V., Warren,W.C., Wallis,J. *et al*. (2004) Fosmid-based physical mapping of the Histoplasma capsulatum genome. *Genome Res.*, **14**, 1603–1609.
11. Martin,F., Aerts,A., Ahren,D. *et al*. (2008) The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature*, **452**, 88–92.
12. Martinez,D., Berka,R.M., Henrissat,B. *et al*. (2008) Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). *Nat. Biotechnol.*, **26**, 553–560.
13. Martinez,D., Challacombe,J., Morgenstern,I. *et al*. (2009) Genome, transcriptome, and secretome analysis of wood decay fungus Postia placenta supports unique mechanisms of lignocellulose conversion. *Proc. Natl Acad. Sci. USA*, **106**, 1954–1959.
14. Martinez,D., Larrondo,L.F., Putnam,N. *et al*. (2004) Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78. *Nat. Biotechnol.*, **22**, 695–700.
15. Nierman,W.C., Pain,A., Anderson,M.J. *et al*. (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature*, **438**, 1151–1156.
16. Dean,R.A., Talbot,N.J., Ebbole,D.J. *et al*. (2005) The genome sequence of the rice blast fungus Magnaporthe grisea. *Nature*, **434**, 980–986.
17. Galagan,J.E., Calvo,S.E., Borkovich,K.A. *et al*. (2003) The genome sequence of the filamentous fungus Neurospora crassa. *Nature*, **422**, 859–868.
18. Espagne,E., Lespinet,O., Malagnac,F. *et al*. (2008) The genome sequence of the model ascomycete fungus Podospora anserina. *Genome Biol.*, **9**, R77.
19. Kamper,J., Kahmann,R., Bolker,M. *et al*. (2006) Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. *Nature*, **444**, 97–101.
20. Chang,A., Scheer,M., Grote,A. *et al*. (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.*, **37**, D588–D592.
21. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J. *et al*. (2008) GenBank. *Nucleic Acids Res.*, **36**, D25–D30.
22. Consortium,U. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
23. Consortium,T.G.O. (2000) Gene Ontology: tool for the unification of biology. *Nature Genet.*, **25**, 25–29.
24. Hessing,J.G., van Rotterdam,C., Verbakel,J.M. *et al*. (1994) Isolation and characterization of a 1,4-beta-endoxylanase gene of A. awamori. *Curr. Genet.*, **26**, 228–232.
25. Giesbert,S., Lepping,H.B., Tenberge,K.B. *et al*. (1998) The xylanolytic system of claviceps purpurea: cytological evidence for secretion of xylanases in infected rye tissue and molecular characterization of two xylanase genes. *Phytopathology*, **88**, 1020–1030.
26. de Graaff,L.H., van den Broeck,H.C., van Ooijen,A.J. *et al*. (1994) Regulation of the xylanase-encoding xlnA gene of Aspergillus tubigensis. *Mol. Microbiol.*, **12**, 479–490.
27. Steenbakkers,P.J., Ubhayasekera,W., Goossen,H.J. *et al*. (2002) An intron-containing glycoside hydrolase family 9 cellulase gene encodes the dominant 90kDa component of the cellulosome of the anaerobic fungus Piromyces sp. strain E2. *Biochem. J.*, **365**(Pt 1), 193–204.
28. Desmet,T., Cantaert,T., Gualfetti,P. *et al*. (2007) An investigation of the substrate specificity of the xyloglucanase Cel74A from Hypocrea jecorina. *FEBS J.*, **274**, 356–363.
29. Murray,P., Aro,N., Collins,C. *et al*. (2004) Expression in Trichoderma reesei and characterisation of a thermostable family 3 beta-glucosidase from the moderately thermophilic fungus Talaromyces emersonii. *Protein Expr. Purif.*, **38**, 248–257.
30. Ohnishi,Y., Nagase,M., Ichiyanagi,T. *et al*. (2007) Transcriptional regulation of two cellobiohydrolase encoding genes (cel1 and cel2) from the wood-degrading basidiomycete Polyporus arcularius. *Appl. Microbiol. Biotechnol.*, **76**, 1069–1078.
31. Alcocer,M.J., Furniss,C.S., Kroon,P.A. *et al*. (2003) Comparison of modular and non-modular xylanases as carrier proteins for the efficient secretion of heterologous proteins from Penicillium funiculosum. *Appl. Microbiol. Biotechnol.*, **60**, 726–732.
32. Furniss,C.S.M., Williamson,G. and Kroon,P.A. (2005) The substrate specificity and susceptibility to wheat inhibitor proteins of *Penicillium funiculosum* xylanase from a commercial enzyme preparation. *J. Sci. Food Agriculture*, **85**, 574–582.
33. Kongruang,S., Han,M.J., Breton,C.I. *et al*. (2004) Quantitative analysis of cellulose-reducing ends. *Appl. Biochem. Biotechnol.*, **113–116**, 213–231.