

CAZY family	Environmental sequence	Nearest neighbour ^a	Species No. ^b	% Identity / % Similarity
GH61	AAA2YG01FL	<i>Thielavia terrestris</i>	62	66 / 77
GH61	AAA12YM05FL	<i>Gloeophyllum trabeum</i>	84	59 / 71
GH61	AAA21YH11FL	AAA15Y110FL	NA	45 / 57
GH61	AAA21YH11FL	<i>Sporotrichum termophile</i>	60	38 / 53
GH61	AAA15Y110FL	<i>Sporotrichum termophile</i>	60	35 / 44
GH61	ABA3YP05FL	<i>Punctularia strigosozonata</i>	98	40 / 54
GH7	AAA16YO07FL	<i>Thielavia terrestris</i>	62	67 / 75
GH7	AAA16YO07FL	<i>Dictyostellium discoideum</i>	4	57 / 70
GH7	AAA16YJ11FL	<i>Daphnia pulex</i>	115	57 / 68
GH45	AAA18YO03FL	<i>Lottia gigantea</i>	121	42 ^c / 54 ^c
GH11	AAA3YM21FL	<i>Talaromyces stipitatus</i>	61	50 / 67
GH5	AAA17YJ10FL	<i>Aspergillus nidulans</i>	33	55 / 68
PL1	AAA2YN20FL	<i>Schizophyllum commune</i>	100	38 / 58
CE1	ABA10YA10FL	<i>Cochliobolus heterostrophus</i>	38	70 / 85

Table S6: Percentages of conserved, identical and similar, amino acid positions between the 12 full length environmental CAZyme proteins and one or two of their closest phylogenetically related neighbours as presented in the phylogenetic analyses (Figures 2 and S4). Percentages were deduced from global amino acid sequence alignments obtained using the Needleman-Wunsch algorithm as implemented at the ncbi web site (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&SHOW_DEFAULTS=on&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink).

^a Species with one of the “most related” protein sequence (Figures 2 and S4).

^b As listed in Table S5

^c alignment limited to 175 amino acid positions representing the conserved GH45 domain (see Figure 2).