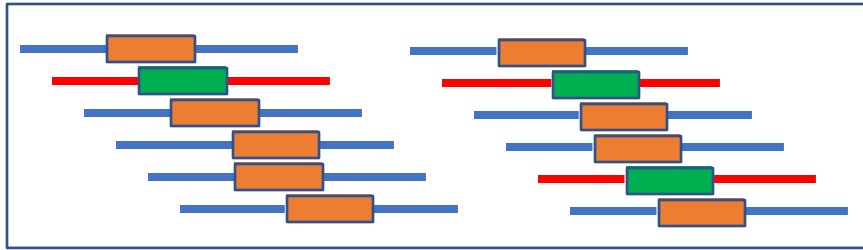


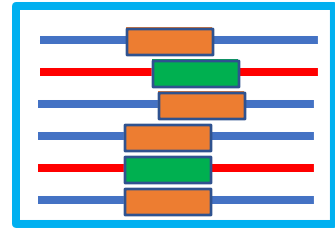
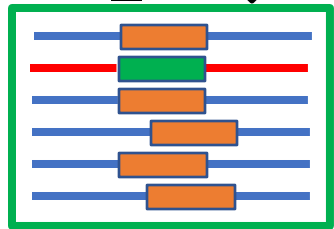
Proteins of CAZy family (e.g., GH1:  
<http://www.cazy.org/GH1.htm> )

## workflow of dbCAN-sub HMM construction

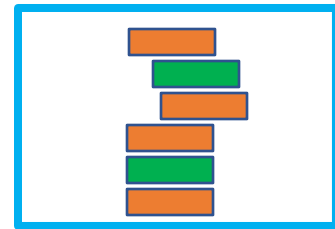
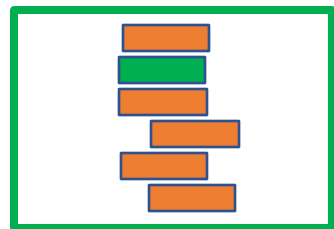


proteins of a CAZy family (e.g., GH1)

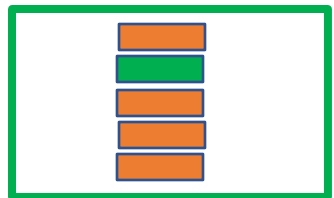
GH1\_e0  **eCAMI** *Bioinformatics 2020*  GH1\_e1



proteins of eCAMI subfamilies (e.g., GH1\_e0)



dbCAN GH1 domains of proteins of GH1\_e0 (hmmscan)





dereplicated dbCAN GH1 domains (cd-hit 95% sequence identity)

HMM-GH1\_e0

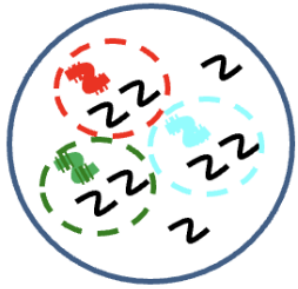
HMM-GH1\_e1

multiple sequence alignment and HMM (mafft & hmmbuild)

 CAZy proteins with EC numbers  
 CAZy proteins without EC numbers

### Notes:

1. some proteins do not contain the dbCAN domain (rare but possible)
2. some domains are removed after cd-hit
3. not all subfamilies have CAZy proteins with EC numbers
4. only subfamilies with sequence count  $\geq 4$  after cd-hit are used for HMM construction



# dbCAN-sub

**Subfamily → EC → substrates**

**Database of CAZyme subfamilies for substrate annotation**

[Home](#)

[Download](#)

[Help](#)

[About us](#)

**AA**

[CBM](#)

[CE](#)

[GH](#)

[GT](#)

[PL](#)

**Cite us: [NAR/gky328](#), [gky418](#) and [gks479](#)**

[AA1](#)

[AA2](#)

[AA3](#)

[AA4](#)

[AA5](#)

[AA6](#)

[AA7](#)

[AA8](#)

[AA9](#)

[AA10](#)

[AA11](#)

[AA12](#)

[AA13](#)

[AA14](#)

[AA15](#)

[AA16](#)

[AA17](#)

[AA18](#)

click each tab to access the six CAZyme classes  
all families of the selected class are shown  
click on each family will expand to show the subfamilies



## AA1

Cite us: NAR/gky328, gky418 and gks479

[<- Back to dbCAN-sub](#)

[AA1\\_e0\(0\)](#)   [AA1\\_e1\(1\)](#)   [AA1\\_e2\(0\)](#)   [AA1\\_e3\(2\)](#)   [AA1\\_e4\(0\)](#)   [AA1\\_e5\(0\)](#)   [AA1\\_e6\(0\)](#)   [AA1\\_e7\(0\)](#)   [AA1\\_e8\(1\)](#)   [AA1\\_e9\(0\)](#)   [AA1\\_e10\(0\)](#)   [AA1\\_e11\(0\)](#)  
[AA1\\_e12\(0\)](#)   [AA1\\_e13\(0\)](#)   [AA1\\_e14\(0\)](#)   [AA1\\_e15\(0\)](#)   [AA1\\_e16\(0\)](#)   [AA1\\_e17\(0\)](#)   [AA1\\_e18\(2\)](#)   [AA1\\_e19\(0\)](#)   [AA1\\_e20\(0\)](#)   [AA1\\_e21\(1\)](#)   [AA1\\_e22\(0\)](#)  
[AA1\\_e23\(0\)](#)   [AA1\\_e24\(0\)](#)   [AA1\\_e25\(0\)](#)   [AA1\\_e26\(1\)](#)   [AA1\\_e27\(0\)](#)   [AA1\\_e28\(0\)](#)   [AA1\\_e29\(0\)](#)   [AA1\\_e30\(0\)](#)   [AA1\\_e31\(0\)](#)   [AA1\\_e32\(0\)](#)   [AA1\\_e33\(77\)](#)  
[AA1\\_e34\(0\)](#)   [AA1\\_e35\(0\)](#)   [AA1\\_e36\(0\)](#)   [AA1\\_e37\(0\)](#)   [AA1\\_e38\(0\)](#)   [AA1\\_e39\(0\)](#)   [AA1\\_e40\(0\)](#)   [AA1\\_e41\(0\)](#)   [AA1\\_e42\(0\)](#)   [AA1\\_e43\(1\)](#)   [AA1\\_e44\(0\)](#)  
[AA1\\_e45\(0\)](#)   [AA1\\_e46\(5\)](#)   [AA1\\_e47\(0\)](#)   [AA1\\_e48\(0\)](#)   [AA1\\_e49\(0\)](#)   [AA1\\_e50\(0\)](#)   [AA1\\_e51\(0\)](#)   [AA1\\_e52\(0\)](#)   [AA1\\_e53\(0\)](#)   [AA1\\_e54\(0\)](#)   [AA1\\_e55\(0\)](#)  
[AA1\\_e56\(0\)](#)   [AA1\\_e57\(0\)](#)   [AA1\\_e58\(0\)](#)   [AA1\\_e59\(0\)](#)   [AA1\\_e60\(0\)](#)   [AA1\\_e61\(0\)](#)   [AA1\\_e62\(0\)](#)   [AA1\\_e63\(0\)](#)   [AA1\\_e64\(0\)](#)   [AA1\\_e65\(0\)](#)   [AA1\\_e66\(0\)](#)  
[AA1\\_e67\(0\)](#)   [AA1\\_e68\(0\)](#)   [AA1\\_e69\(0\)](#)   [AA1\\_e70\(0\)](#)   [AA1\\_e71\(0\)](#)   [AA1\\_e72\(0\)](#)   [AA1\\_e73\(0\)](#)   [AA1\\_e74\(0\)](#)   [AA1\\_e75\(0\)](#)   [AA1\\_e76\(0\)](#)   [AA1\\_e77\(0\)](#)  
[AA1\\_e78\(0\)](#)   [AA1\\_e79\(0\)](#)   [AA1\\_e80\(0\)](#)   [AA1\\_e81\(0\)](#)   [AA1\\_e82\(0\)](#)   [AA1\\_e83\(0\)](#)   [AA1\\_e84\(0\)](#)   [AA1\\_e85\(0\)](#)   [AA1\\_e86\(0\)](#)   [AA1\\_e87\(0\)](#)   [AA1\\_e88\(0\)](#)  
[AA1\\_e89\(0\)](#)   [AA1\\_e90\(0\)](#)   [AA1\\_e91\(0\)](#)   [AA1\\_e92\(0\)](#)   [AA1\\_e93\(0\)](#)   [AA1\\_e94\(0\)](#)   [AA1\\_e95\(0\)](#)   [AA1\\_e96\(0\)](#)   [AA1\\_e97\(0\)](#)   [AA1\\_e98\(0\)](#)   [AA1\\_e99\(0\)](#)  
[AA1\\_e100\(0\)](#)   [AA1\\_e101\(0\)](#)   [AA1\\_e102\(0\)](#)   [AA1\\_e103\(0\)](#)   [AA1\\_e104\(0\)](#)   [AA1\\_e105\(0\)](#)   [AA1\\_e106\(0\)](#)   [AA1\\_e107\(0\)](#)   [AA1\\_e108\(0\)](#)   [AA1\\_e109\(0\)](#)  
[AA1\\_e110\(0\)](#)   [AA1\\_e111\(0\)](#)   [AA1\\_e112\(0\)](#)   [AA1\\_e113\(0\)](#)   [AA1\\_e114\(0\)](#)   [AA1\\_e115\(0\)](#)   [AA1\\_e116\(0\)](#)   [AA1\\_e117\(0\)](#)   [AA1\\_e118\(0\)](#)   [AA1\\_e119\(0\)](#)  
[AA1\\_e120\(0\)](#)   [AA1\\_e121\(0\)](#)   [AA1\\_e122\(0\)](#)   [AA1\\_e123\(0\)](#)   [AA1\\_e124\(0\)](#)   [AA1\\_e125\(0\)](#)   [AA1\\_e126\(0\)](#)   [AA1\\_e127\(0\)](#)   [AA1\\_e128\(0\)](#)   [AA1\\_e129\(0\)](#)  
[AA1\\_e130\(0\)](#)   [AA1\\_e131\(0\)](#)   [AA1\\_e132\(0\)](#)   [AA1\\_e133\(0\)](#)   [AA1\\_e134\(0\)](#)   [AA1\\_e135\(0\)](#)   [AA1\\_e136\(0\)](#)   [AA1\\_e137\(0\)](#)   [AA1\\_e138\(0\)](#)   [AA1\\_e139\(0\)](#)

click on AA1 to show all subfamilies of AA1

the subfamilies were classified by eCAMI: <https://github.com/yinlabniu/eCAMI>

these subfamilies are named with an “e” in them, e.g., AA1\_e0 (e means eCAMI)

click on each subfamily will open a new page

## AA1\_e1

[https://bcb.unl.edu/dbCAN\\_sub/dbsub.php?contig=AA1\\_e1](https://bcb.unl.edu/dbCAN_sub/dbsub.php?contig=AA1_e1)[<- Back to dbCAN-sub](#)

## summary of CAZy proteins in the subfamily (different steps lead to different counts)

### Summary

Number of CAZy proteins	109
Number of CAZy proteins with ECs	1
Number of CAZy proteins with the corresponding HMM domains	109
Number of HMM domains	109
Number of HMM domains after cd-hit	32
Number of CAZy proteins with the corresponding HMM domains and with ECs	1

#### Notes:

1. some proteins do not contain the dbCAN domain (rare but possible)
2. some domains are removed after cd-hit
3. not all subfamilies have CAZy proteins with EC numbers
4. only subfamilies with sequence count  $\geq 4$  after cd-hit are used for HMM construction

[Download HMM domains after cd-hit](#) [Download AA1\\_e1](#) [Download CAZy proteins with the corresponding HMM domains and with ECs](#)

### Substrate Table

## CAZy proteins with EC numbers in this subfamily

Show  entries Search:

EC	Count	CAZy protein ID	Substrates
<a href="#">1.10.3.2</a>	1	<a href="#">AFC76164.1</a>	<a href="#">lignin</a>

Showing 1 to 1 of 1 entries First Previous 1 Next Last

[Download Substrate Table](#)

manually curated family-EC-substrate mapping table

# Search dbCAN-sub @ the dbCAN meta server

<https://bcb.unl.edu/dbCAN2/blast.php>

check here to search dbCAN-sub  
for substrate prediction

## Choose Sequence Type:

Protein sequence (example) ?  Nucleotide sequence (example) ?

## Select Tools To Run:

HMMER: dbCAN (E-Value < 1e-15, coverage > 0.35)  DIAMOND: CAZy (E-Value < 1e-102)  HMMER: dbCAN-sub (E-Value < 1e-15, coverage > 0.35)   
CGCFinder (Distance <= 2, signature genes = CAZyme+TC)?

**Just paste some sequences here (note: only FASTA format please!!!)**

Try [example](#) sequences

**Or upload a fasta file including many sequences (Max File Size: 10MB)**

Fasta file:  No file chosen

# Search dbCAN-sub @ the dbCAN meta server

<https://bcb.unl.edu/dbCAN2/blastation.php?jobid=20251219173214>

dbCAN-sub result tab  
for details

this col has dbCAN-sub result

link to dbCAN-sub page

Result of job: 20251219173214

Overview

HMMER: dbCAN

DIAMOND: CAZy

HMMER: dbCAN\_sub

[Download SignalP output](#) [Download input file](#) [Download CAZyme sequences](#) [Download this table](#) (keep those with # of Tools >=2 will give you best result; and use dbCAN domain assignment is recommended) ?

Show 15 entries

Search:

Gene ID	EC#	HMMER	DIAMOND	dbCAN_sub	Signal Peptide	# of Tools
<a href="#">AT1G11720.1</a>	<a href="#">2.4.1.21</a>	<a href="#">CBM53</a> (154-237)+ <a href="#">CBM53</a> (329-423)+ <a href="#">CBM53</a> (496-584)+ <a href="#">GT5</a> (595-1038)	<a href="#">CBM53+GT5</a>	<a href="#">CBM53_e1+CBM53_e0+CBM53_e0+GT5_e38</a>	N	3
<a href="#">gj 222529846 ref YP_002573728.1 </a>	<a href="#">3.2.1.4 3.2.1.176 3.2.1.73 3.2.1.- 3.2.1.78 3.2.1.151 3.2.1.8 3.2.1.55 3.2.1.74 3.2.1.91 3.2.1.14</a>	<a href="#">GH9</a> (36-466)+ <a href="#">CBM3</a> (491-576)+ <a href="#">CBM3</a> (724-804)+ <a href="#">CBM3</a> (923-1003)+ <a href="#">GH48</a> (1134-1755)	<a href="#">CBM3+GH48+GH9</a>	<a href="#">GH9_e22+CBM3_e0+CBM3_e16+CBM3_e16+GH48_e1</a>	N	3

dbCAN-sub result tab for details

link to dbCAN-sub page

these are the seq composition of CBM53\_e1 HMM

CBM53:91|GT5:87|GT31:2 means the CBM53\_e1 subfam contains 91 CAZy proteins from CBM53 fam, 87 from GT5, and 2 from GT31 (CAZymes are often multi-domain proteins)

## Result of job: 20251219173214

Overview

HMMER: dbCAN

DIAMOND: CAZy

HMMER: dbCAN\_sub

Download dbCAN\_sub output

Show 15 entries

Search:

Query ID	dbCAN Subfam Length	dbCAN Subfam	Subfam Composition	Subfam EC	Substrate	E Value	Coverage
AT1G11720.1	89	<a href="#">CBM53_e0</a>	CBM53:62 GT5:59	" <a href="#">2.4.1.21:13</a> "	starch	5.8e-28	0.955056179775
AT1G11720.1	89	<a href="#">CBM53_e1</a>	CBM53:91 GT5:87 GT31:2	" <a href="#">2.4.1.21:27</a> "	starch	6.3e-28	0.966292134831
AT1G11720.1	89	<a href="#">CBM53_e0</a>	CBM53:62 GT5:59	" <a href="#">2.4.1.21:13</a> "	starch	1.5e-28	0.977528089888
AT1G11720.1	445	<a href="#">GT5_e38</a>	GT5:214 CBM53:145 GT31:2	" <a href="#">2.4.1.21:40</a> "	-	8.6e-237	0.997752808989

# CBM53\_e1 sequence composition is described in the dbCAN-sub webpage

Cite us: [NAR/gky328](#), [gky418](#) and [gks479](#)

## CBM53\_e1

[<- Back to dbCAN-sub](#)

### Summary

Number of CAZy proteins	93
Number of CAZy proteins with ECs	26
Number of CAZy proteins with the corresponding HMM domains	263
Number of HMM domains	263
Number of HMM domains after cd-hit	99
Number of CAZy proteins with the corresponding HMM domains and with ECs	26

[Download HMM domains after cd-hit](#) [Download CBM53\\_e1](#) [Download CAZy proteins with the corresponding HMM domains and with ECs](#)

### Substrate Table

Show  entries

Search:

EC	Count	CAZy protein ID	Substrates
<a href="#">2.4.1.21</a>	26	<a href="#">AAD30251.1</a> <a href="#">AAE35726.1</a> <a href="#">AAN99623.1</a> <a href="#">AAR75241.1</a> <a href="#">AAY42381.1</a> <a href="#">ABB46391.1</a> <a href="#">AD531123.1</a>	<a href="#">alpha-glucan</a>