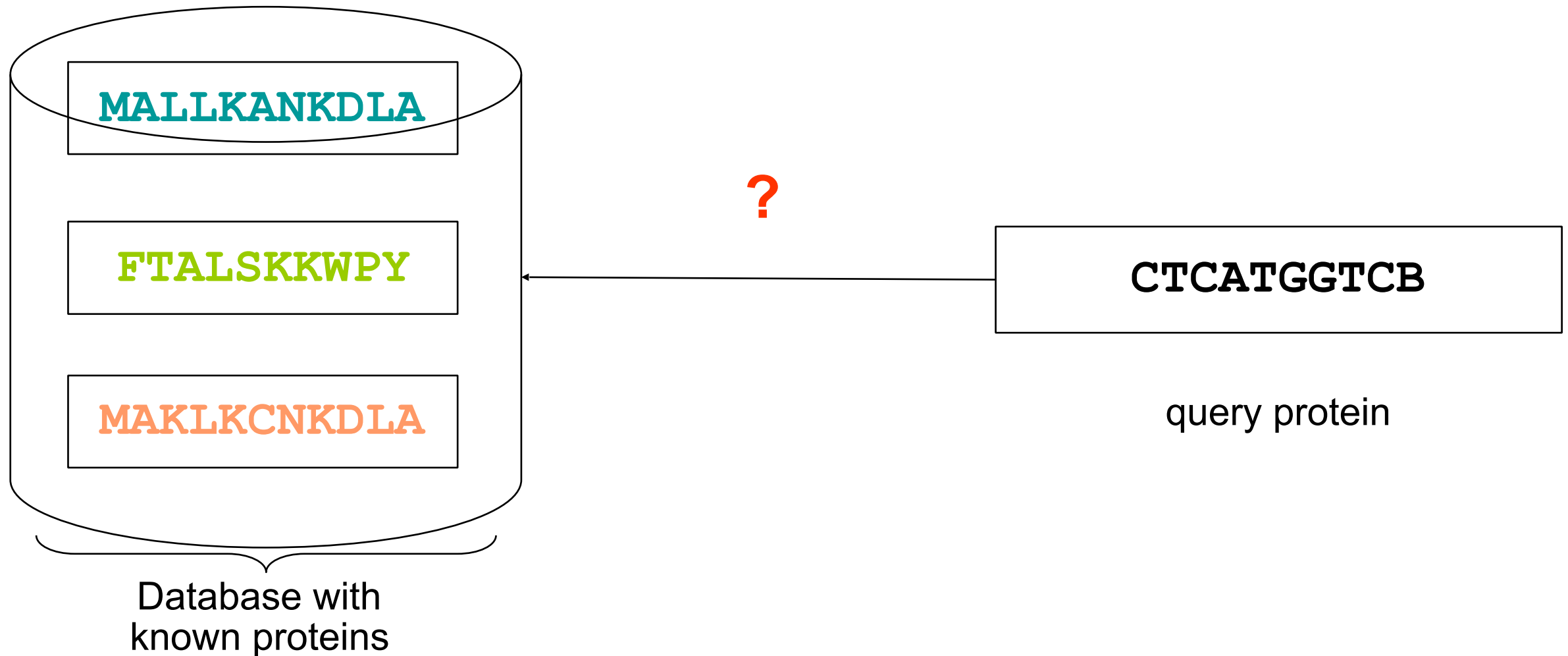


# Profile Hidden Markov Models

## HMMer

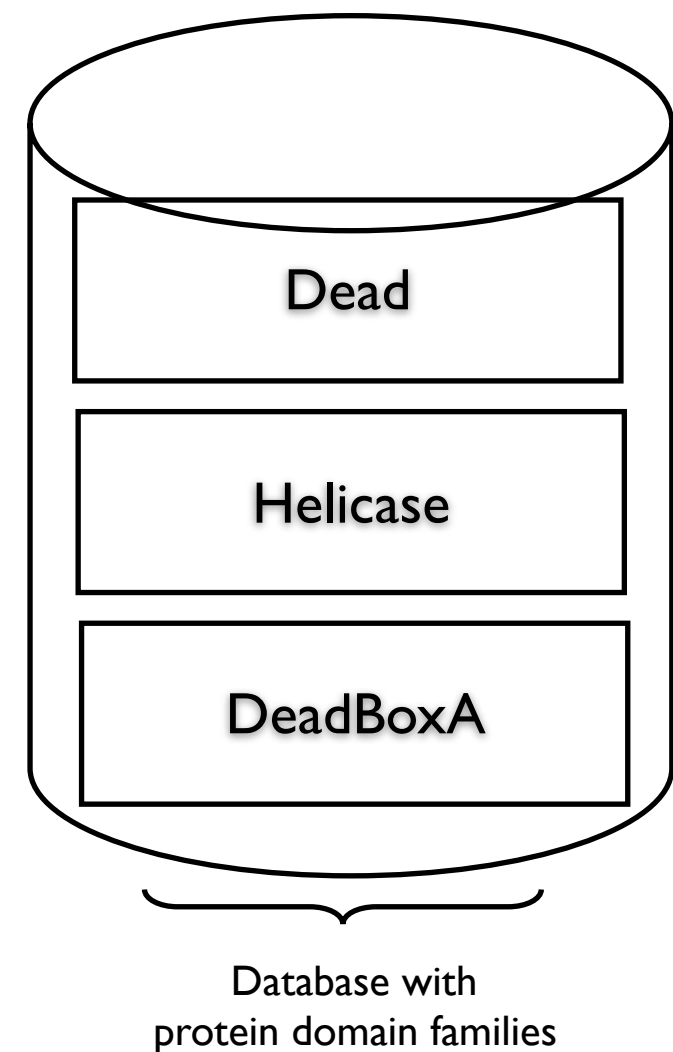
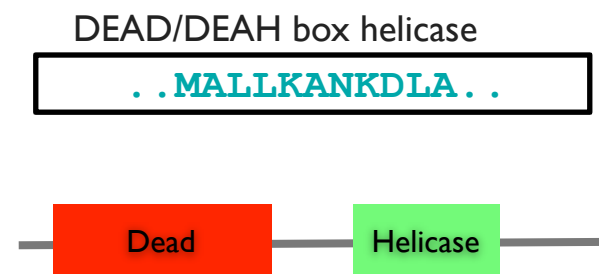
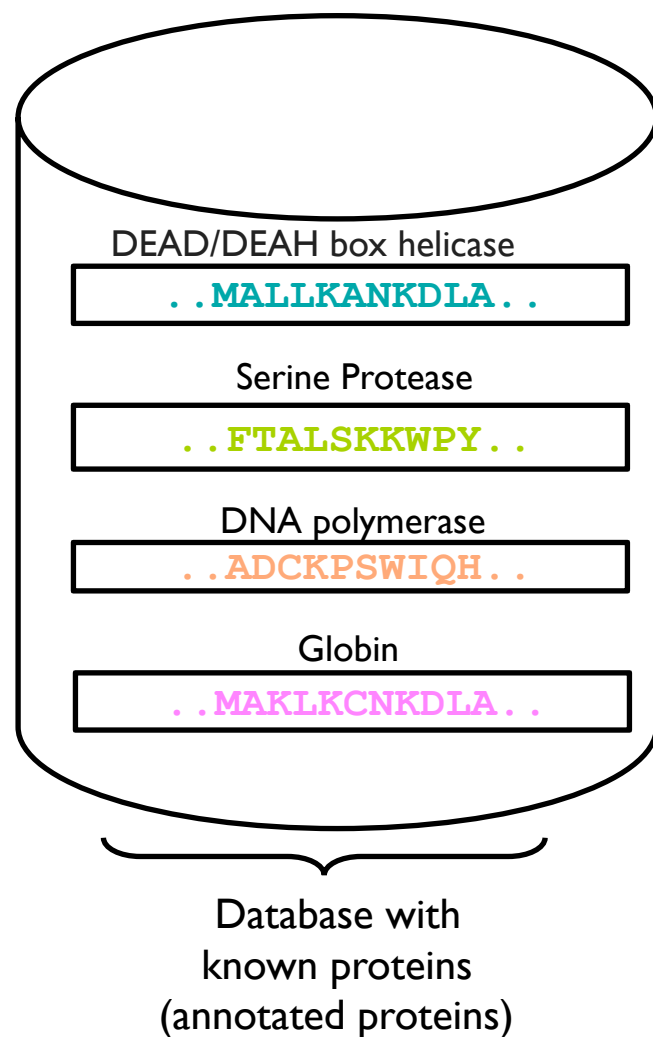
# Homology Detection



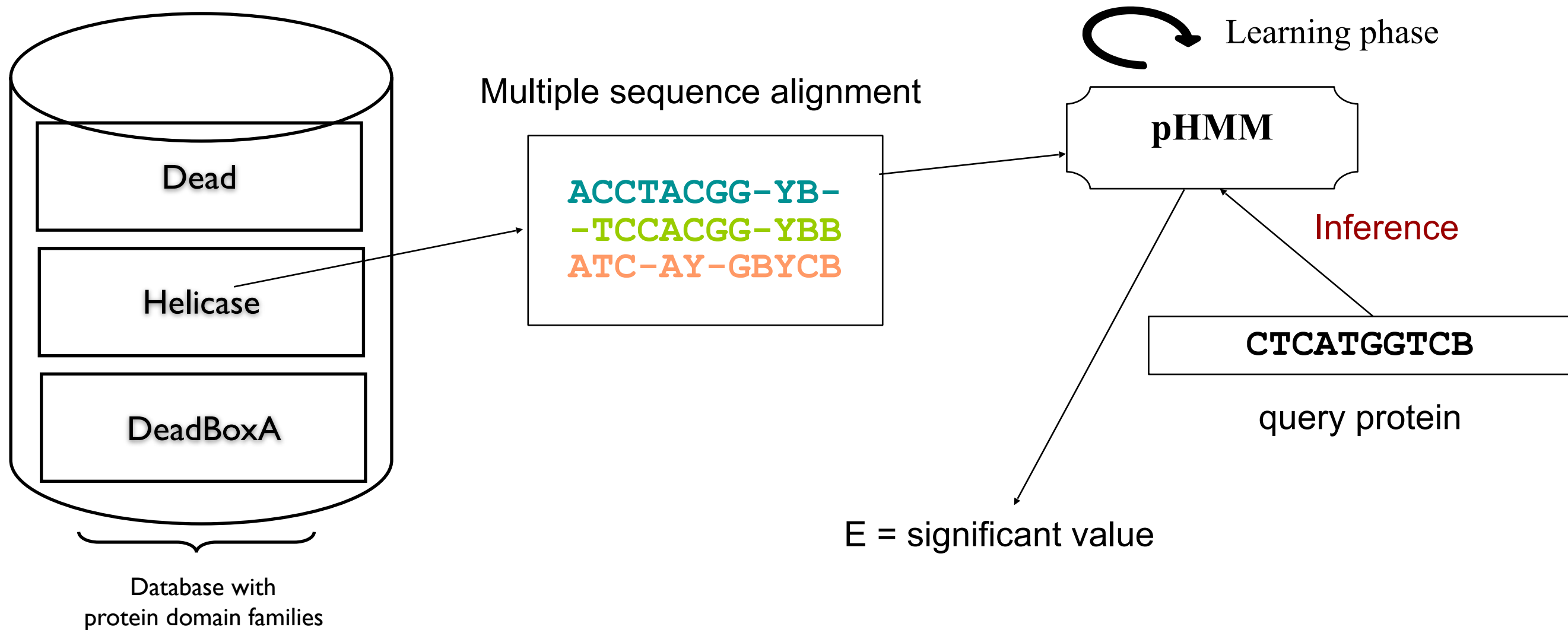
# Homology Detection

➡ To improve homology detection, we can classify known protein sequences according to their functional regions (**domains**).

➡ Proteins are generally comprised of one or more **domains**.




# profile HMMs



hmmer.janelia.org/software

Most Visited Getting Started Apple iCloud Facebook Twitter Wikipedia Yahoo!



# HMNER

biosequence analysis using profile hidden Markov models

Home Search Results **Software** Help About

Current Archive

## The current version of HMNER

### Download

The **current version** is HMNER **3.1b2** (05 March 2015).

|                                   |                       |                        |         |
|-----------------------------------|-----------------------|------------------------|---------|
| Source:                           | <a href="#">[FTP]</a> | <a href="#">[HTTP]</a> | 5.8 MB  |
| with Linux/Intel ia32 binaries:   | <a href="#">[FTP]</a> | <a href="#">[HTTP]</a> | 18.1 MB |
| with Linux/Intel x86_64 binaries: | <a href="#">[FTP]</a> | <a href="#">[HTTP]</a> | 20.2 MB |
| with MacOSX/Intel binaries:       | <a href="#">[FTP]</a> | <a href="#">[HTTP]</a> | 13.5 MB |

If you are looking for **older versions** of the software, try the [archive](#) link at the top of the page.

### Documentation

[Release notes](#) and [User's Guide](#): [\[PDF, 116 pages\]](#).

### Briefly, to compile from source:

```
% tar xzf hmmer-3.1b2.tar.gz
% cd hmmer-3.1b2
% ./configure
% make
% make check
```

# Building HMMER models

**hmmbuild [-options] <hmmfile output> <alignment file input>**

Options for selecting alphabet rather than guessing it:

- amino : input alignment is protein sequence data
- dna : input alignment is DNA sequence data
- rna : input alignment is RNA sequence data

Alternative model construction strategies:

- fast : assign cols w/  $\geq$  symfrac residues as consensus [default]
- hand : manual construction (requires reference annotation)
- symfrac <x> : sets sym fraction controlling --fast construction [0.5]
- fragthresh <x> : if  $L < x < L$ , tag sequence as a fragment [0.5]

Alternative relative sequence weighting strategies:

- wpb : Henikoff position-based weights [default]
- wgsc : Gerstein/Sonnhammer/Chothia tree weights
- wblosum : Henikoff simple filter weights
- wnone : don't do any relative weighting; set all to 1
- wgiven : use weights as given in MSA file
- wid <x> : for --wblosum: set identity cutoff [0.62] ( $0 \leq x \leq 1$ )

# Searching with HMMER

**hmmsearch [options] <hmmfile> <target seqfile>**

options directing output:

- o <f> : direct output to file <f>, not stdout
- A <f> : save multiple alignment of all hits to file <s>
- tblout <f> : save parseable table of per-sequence hits to file <s>
- domtblout <f> : save parseable table of per-domain hits to file <s>
- acc : prefer accessions over names in output
- noali : don't output alignments, so output is smaller
- notextw : unlimit ASCII text output line width
- textw <n> : set max width of ASCII text output lines [120] (n>=120)

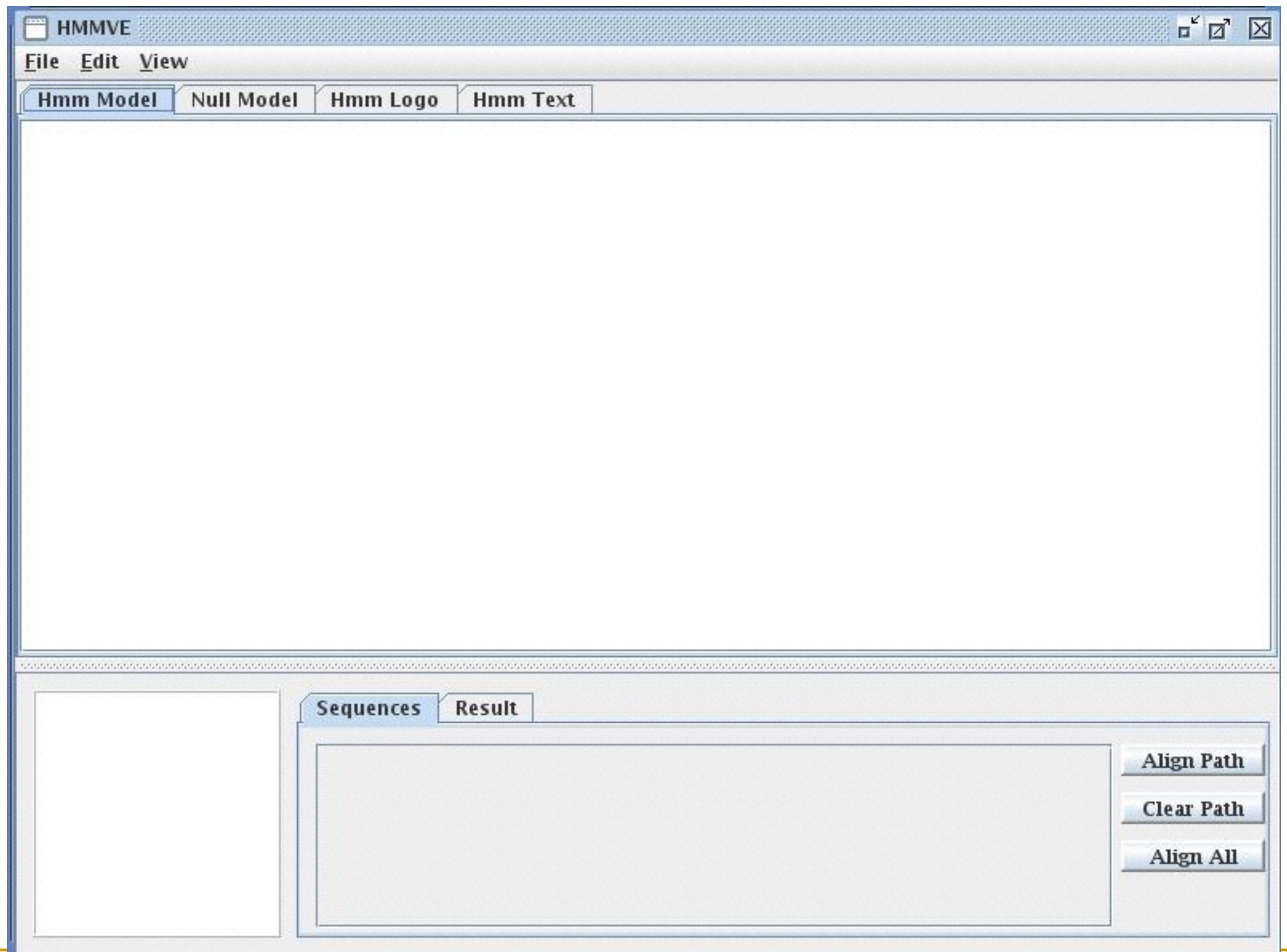
options controlling reporting thresholds:

- E <x> : report sequences <= this E-value threshold in output [10.0] (x>0)
- T <x> : report sequences >= this score threshold in output
- domE <x> : report domains <= this E-value threshold in output [10.0] (x>0)
- domT <x> : report domains >= this score cutoff in output

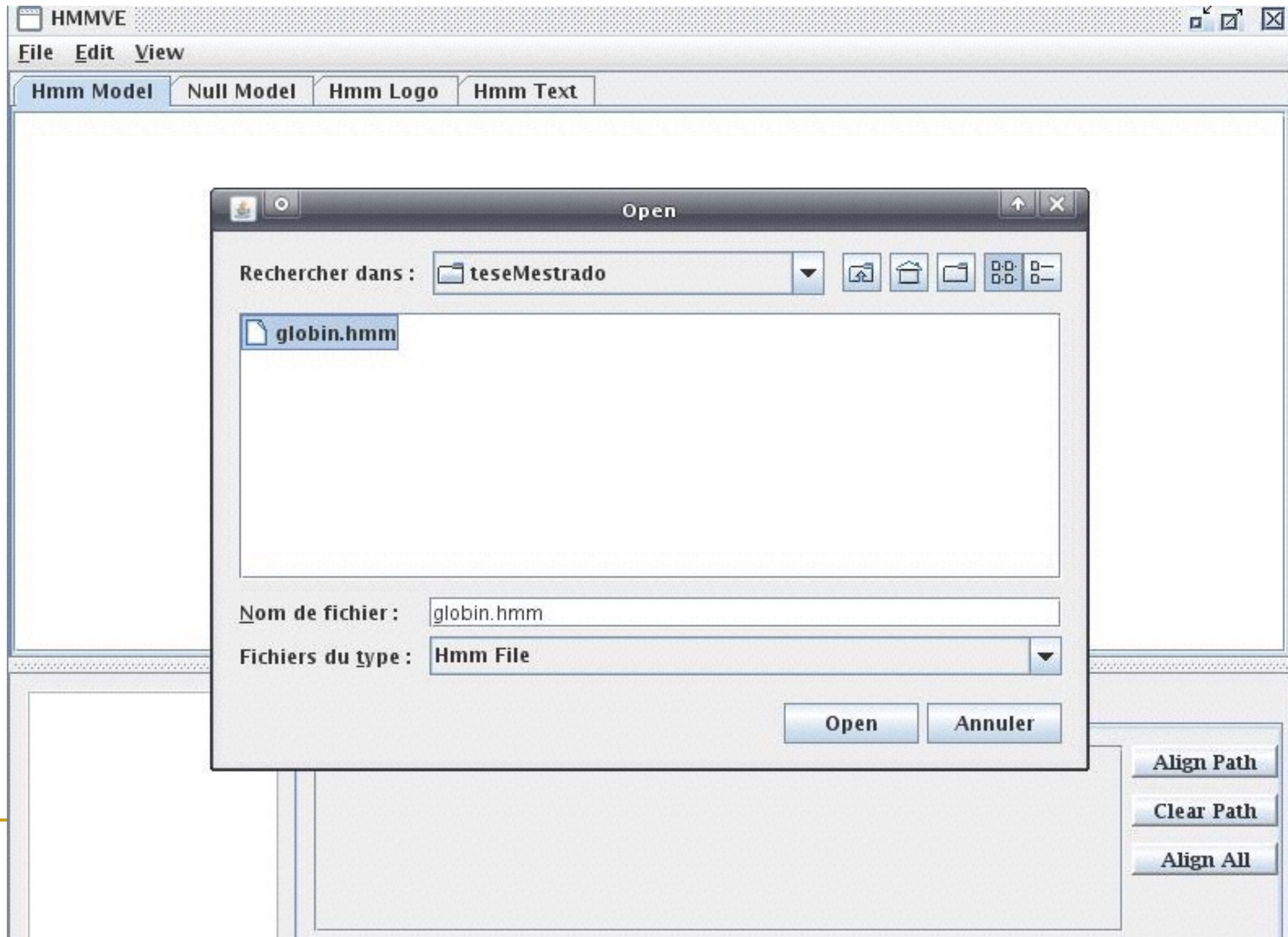
## Visualising profile HMM

```
jsilva@pretinha:~$ cd HMMEditor/  
jsilva@pretinha:~/HMMEditor$ ls  
HMMVE_1.1.jar  
jsilva@pretinha:~/HMMEditor$ pwd  
/home/jsilva/HMMEditor  
jsilva@pretinha:~/HMMEditor$ ls  
HMMVE_1.1.jar  
jsilva@pretinha:~/HMMEditor$ java -jar HMMVE_1.1.jar
```

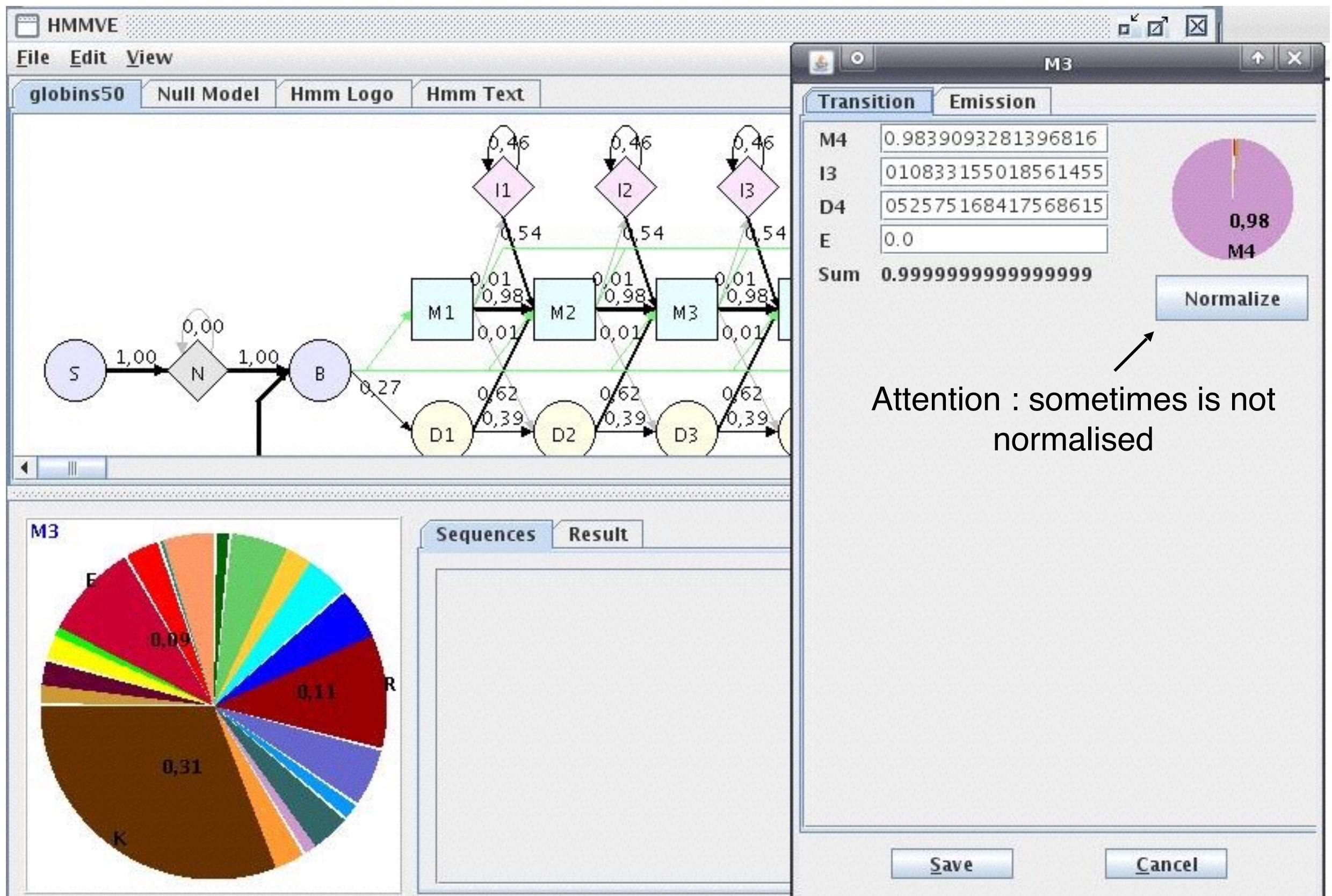




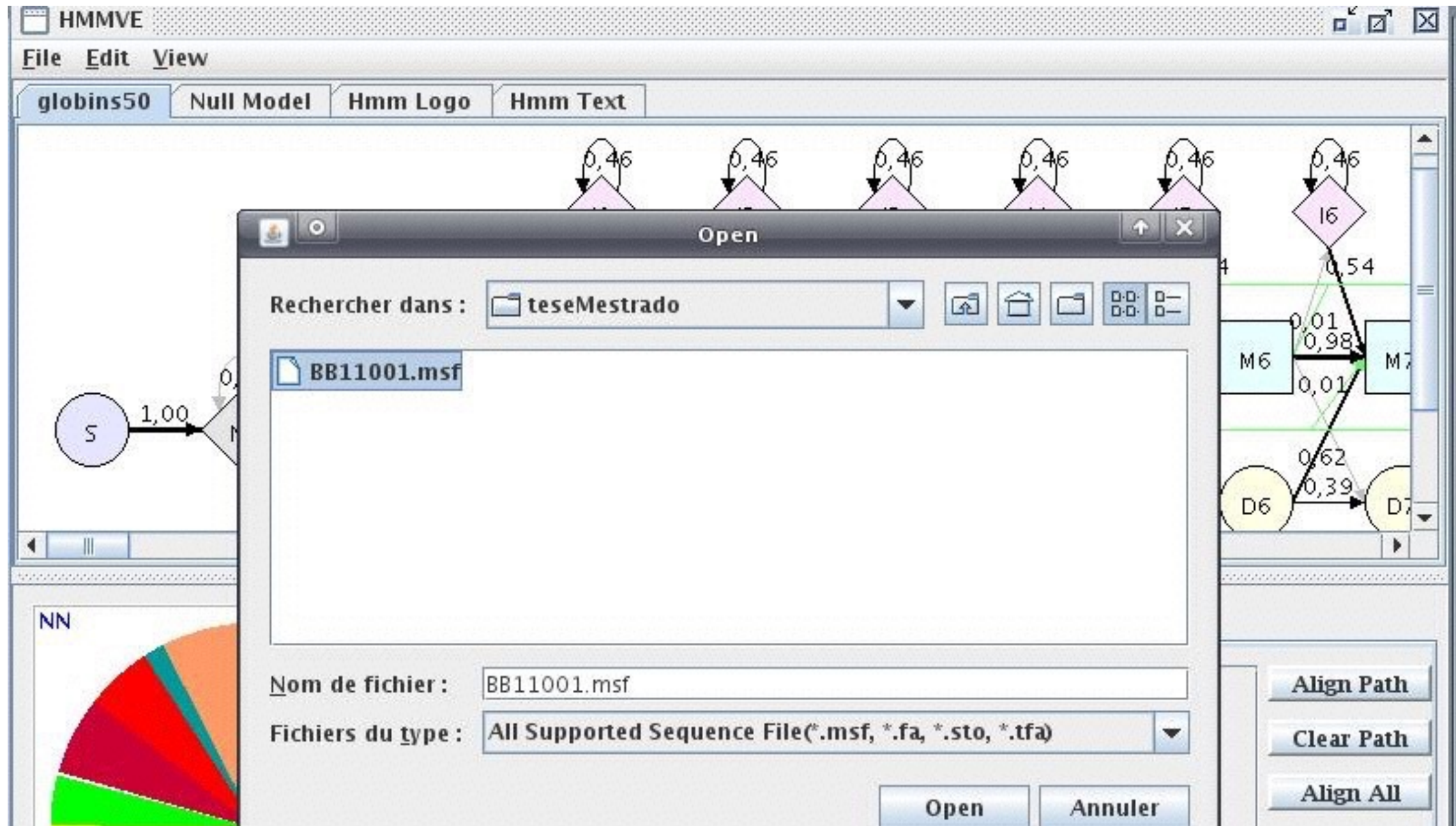
- File → open HMM...
- Select a pHMM model.





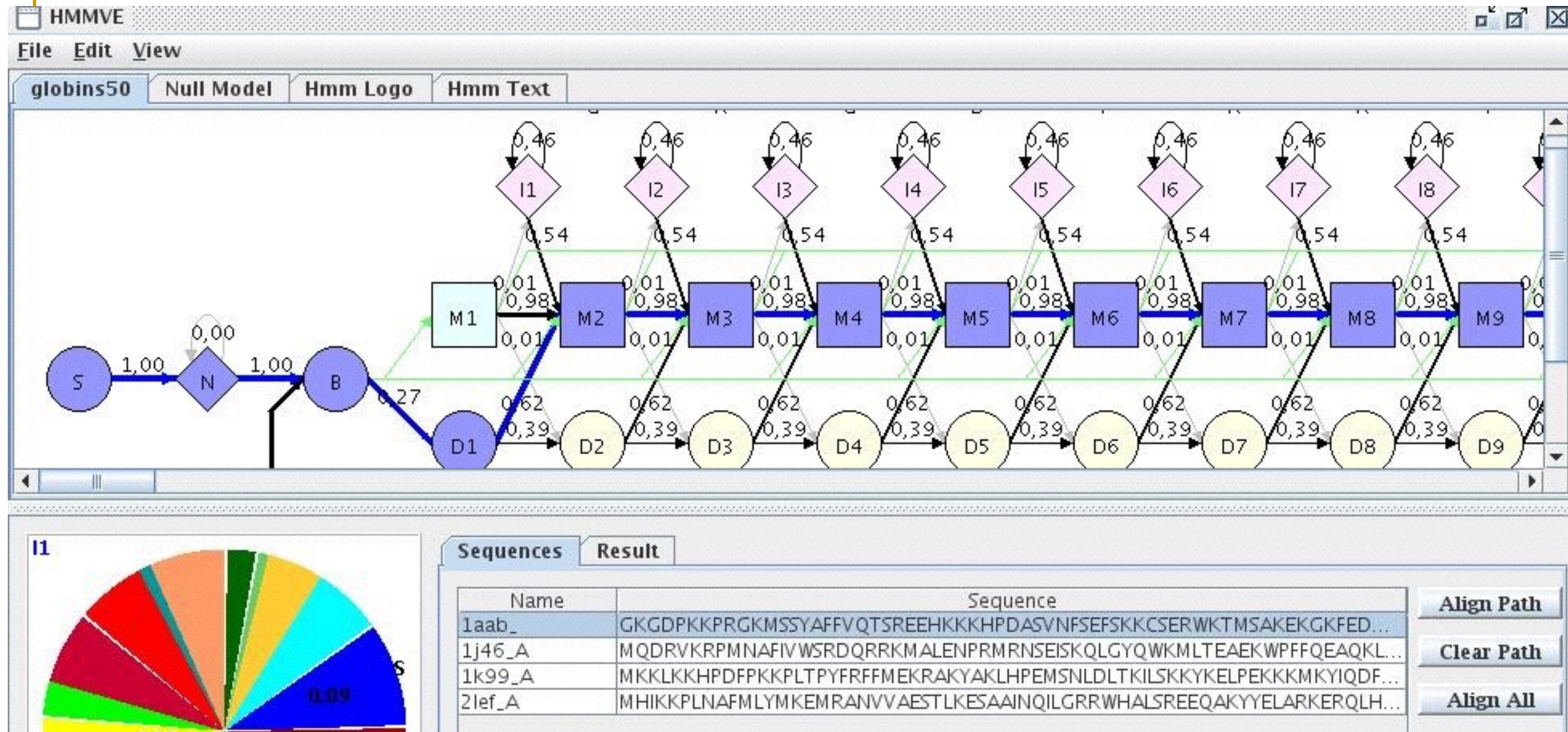


- File → open Sequence...
- Select the query sequence in fasta format.





# - Viterbi pathway



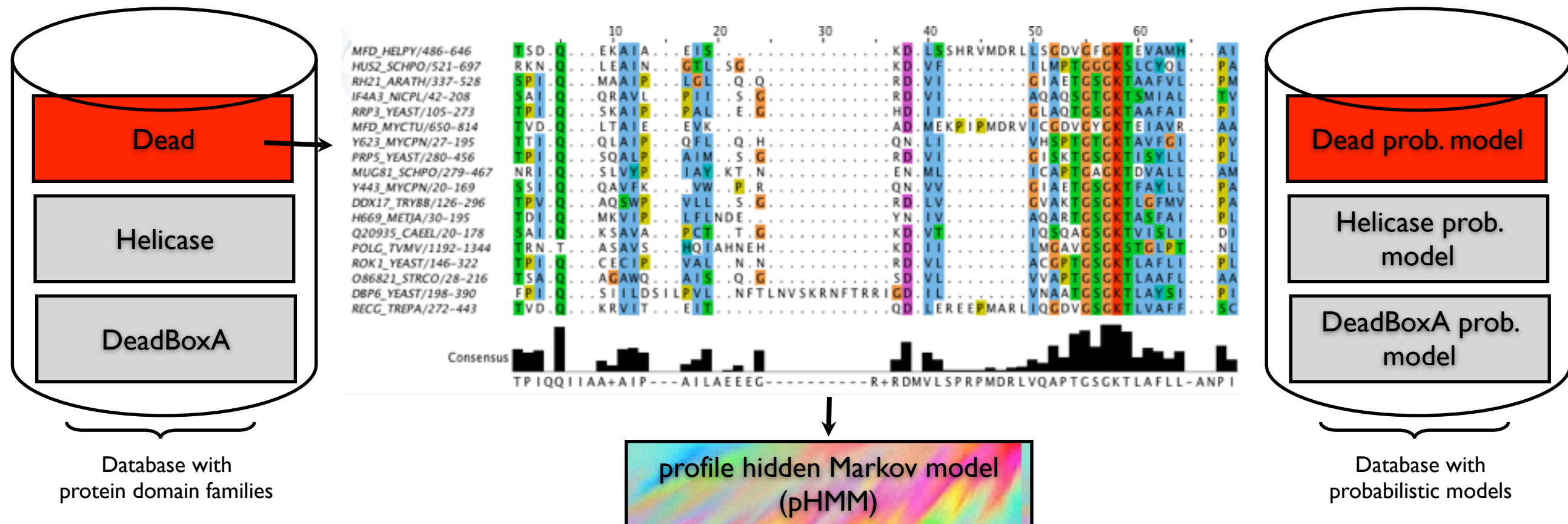
# pHMMs packages

- Softwares
  - HMMER → <http://hmmer.janelia.org/>
  - SAM → <http://compbio.soe.ucsc.edu/sam.html>
- Databases of pHMMs



# Pfam 31.0 (March 2017, 16712 entries)

- ➔ Known domains are described with probabilistic models representing the consensus among domain sequences





# Species

## Family: DEAD (PF00270)

1271 architectures

81109 sequences

21 interactions

5319 species

298 structures

### Summary

### Domain organisation

### Clan

### Alignments

### HMM logo

### Trees

### Curation & model

### Species

### Interactions

### Structures

### Jump to...

enter ID/acc

Go

### Summary: DEAD/DEAH box helicase

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

Wikipedia: DEAD/DEAH box helicase

Pfam

InterPro

This is the Wikipedia entry entitled "DEAD/DEAH box helicase". [More...](#)

### DEAD/DEAH box helicase

[Edit Wikipedia article](#)

The **DEAD/DEAH box helicases** are a family of [proteins](#) whose purpose is to unwind [nucleic acids](#). The [DEAD box helicases](#) are involved in various aspects of [RNA metabolism](#), including nuclear [transcription](#), pre mRNA splicing, [ribosome biogenesis](#), nucleocytoplasmic transport, translation, RNA decay and organellar gene expression. <sup>[2]</sup> <sup>[3]</sup>

### References

- <sup>^</sup> Johnson ER, McKay DB (December 1999). "Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase". *RNA*. **5** (12): 1526–34. doi:10.1017/S1355838299991410. PMC 1369875. PMID 10606264.
- <sup>^</sup> de la Cruz J, Kressler D, Linder P (May 1999). "Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families". *Trends Biochem. Sci.* **24** (5): 192–8. doi:10.1016/S0968-0004(99)01376-6. PMID 10322435.
- <sup>^</sup> Aubourg S, Kreis M, Lecharny A (January 1999). "The DEAD box RNA helicase family in *Arabidopsis thaliana*". *Nucleic Acids Res.* **27** (2): 628–36. doi:10.1093/nar/27.2.628. PMC 148225. PMID 9862990.

This article incorporates text from the [public domain](#) Pfam and InterPro IPR011545.

This page is based on a [Wikipedia article](#). The text is available under the [Creative Commons Attribution/Share-Alike License](#).

### DEAD/DEAH box helicase



Structure of the amino terminal domain of yeast initiation factor 4A. <sup>[1]</sup>

#### Identifiers

|                    |                             |
|--------------------|-----------------------------|
| <b>Symbol</b>      | DEAD                        |
| <b>Pfam</b>        | PF00270 <a href="#">↗</a>   |
| <b>Pfam clan</b>   | CL0023 <a href="#">↗</a>    |
| <b>InterPro</b>    | IPR011545 <a href="#">↗</a> |
| <b>PROSITE</b>     | PDOC00039 <a href="#">↗</a> |
| <b>SCOP</b>        | 1qva <a href="#">↗</a>      |
| <b>SUPERFAMILY</b> | 1qva <a href="#">↗</a>      |
| <b>CDD</b>         | cd00046 <a href="#">↗</a>   |

Available protein structures: [\[show\]](#)



## Species distribution

Sunburst Tree

This visualisation provides a simple graphical representation of the distribution of this family across species. You can find the original interactive tree in the [adjacent tab](#). [More...](#)



# Family: *DEAD* (PF00270)

1271 architectures

81109 sequences

## Summary

## Domain organisation

## Clan

## Alignments

## HMM logo

## Trees

## Curation & model

## Species

## Interactions

## Structures

## Jump to...

enter ID/acc

Go

## Alignments

We store a range of different sequence alignments for families. As well as the seed alignment from which the family is built, we provide the full alignment ([proteomes](#)) using the family HMM. We also generate alignments using four [representative proteomes](#) (RP) sets, the UniProtKB sequence database. [More...](#)

## View options

We make a range of alignments for each Pfam-A family. You can see a description of each [above](#). You can view these alignments in various ways others may not be available for all families, most commonly because the alignments are too large to handle.

|            | Seed<br>(181) | Full<br>(81109) | Representative proteomes |                 |                 |                  | UniProt<br>(180618) | NCBI<br>(362201) | Meta<br>(12814) |
|------------|---------------|-----------------|--------------------------|-----------------|-----------------|------------------|---------------------|------------------|-----------------|
|            |               |                 | RP15<br>(18454)          | RP35<br>(45242) | RP55<br>(72823) | RP75<br>(103504) |                     |                  |                 |
| Jalview    | ✓             | ✓               | ✓                        | ✓               | ✓               | ✓                | ✓                   | ✓                | ✓               |
| HTML       | ✓             | —               | ×                        | ×               | ×               | ×                | ×                   | ×                | ×               |
| PP/heatmap | ×             | —               | ×                        | ×               | ×               | ×                | ×                   | ×                | ×               |

<sup>1</sup>Cannot generate PP/Heatmap alignments for seeds; no PP data available

**Key:** ✓ available, × not generated, — not available.

## Format an alignment

|                | Seed<br>(181)  | Full<br>(81109)       | Representative proteomes |                       |                       |                       | UniProt<br>(180618)   | NCBI<br>(362201)      | Meta<br>(12814)       |
|----------------|--|-----------------------|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|                |  |                       | RP15<br>(18454)          | RP35<br>(45242)       | RP55<br>(72823)       | RP75<br>(103504)      |                       |                       |                       |
| Alignment:     | <input checked="" type="radio"/>   | <input type="radio"/> | <input type="radio"/>    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Format:        | Stockholm  |                       |                          |                       |                       |                       |                       |                       |                       |
| Order:         | <input checked="" type="radio"/> Tree <input type="radio"/> Alphabetical                 |                       |                          |                       |                       |                       |                       |                       |                       |
| Sequence:      | <input checked="" type="radio"/> Inserts lower case <input type="radio"/> All upper case |                       |                          |                       |                       |                       |                       |                       |                       |
| Gaps:          | Gaps as "." or "-" (mixed)   |                       |                          |                       |                       |                       |                       |                       |                       |
| Download/view: | <input checked="" type="radio"/> Download <input type="radio"/> View                     |                       |                          |                       |                       |                       |                       |                       |                       |

Generate

## Download options

We make all of our alignments available in Stockholm format. You can download them here as raw, plain text files or as [gzip](#)-compressed files.

|               | Seed<br>(181) | Full<br>(81109) | Representative proteomes |                 |                 |                  | UniProt<br>(180618) | NCBI<br>(362201) | Meta<br>(12814) |
|---------------|---------------|-----------------|--------------------------|-----------------|-----------------|------------------|---------------------|------------------|-----------------|
|               |               |                 | RP15<br>(18454)          | RP35<br>(45242) | RP55<br>(72823) | RP75<br>(103504) |                     |                  |                 |
| Raw Stockholm | ✓             | ✓               | ✓                        | ✓               | ✓               | ✓                | —                   | —                | ✓               |
| Gzipped       | ✓             | ✓               | ✓                        | ✓               | ✓               | ✓                | —                   | —                | ✓               |

You can also [download](#) a FASTA format file containing the **full-length sequences** for all sequences in the full alignment.



- Summary
- Domain organisation
- Clan

Alignments

- HMM logo
- Trees
- Curation & model
- Species
- Interactions
- Structures

Jump to...

enter ID/acc

Go

Alignments

We store a range of different sequence alignments for families. As well as the seed alignment from which the family is built, we provide the full alignment ([proteomes](#)) using the family HMM. We also generate alignments using four [representative proteomes](#) (RP) sets, the UniProtKB sequence database, database. [More...](#)

View options

We make a range of alignments for each Pfam-A family. You can see a description of each [above](#). You can view these alignments in various ways but others may not be available for all families, most commonly because the alignments are too large to handle.

|            | Seed<br>(181)  | Full<br>(81109) | Representative proteomes |                 |                 |                  | UniProt<br>(180618) | NCBI<br>(362201) | Meta<br>(12814) |
|------------|----------------|-----------------|--------------------------|-----------------|-----------------|------------------|---------------------|------------------|-----------------|
|            |                |                 | RP15<br>(18454)          | RP35<br>(45242) | RP55<br>(72823) | RP75<br>(103504) |                     |                  |                 |
| Jalview    | ✓              | ✓               | ✓                        | ✓               | ✓               | ✓                | ✓                   | ✓                | ✓               |
| HTML       | ✓              | —               | ✗                        | ✗               | ✗               | ✗                | ✗                   | ✗                | ✗               |
| PP/heatmap | ✗ <sup>1</sup> | —               | ✗                        | ✗               | ✗               | ✗                | ✗                   | ✗                | ✗               |

<sup>1</sup>Cannot generate PP/Heatmap alignments for seeds; no PP data available

Key: ✓ available, ✗ not generated, — not available.

Format an alignment

|                | Seed<br>(181)   | Full<br>(81109)       | Representative proteomes |                       |                       |                       | UniProt<br>(180618)   | NCBI<br>(362201)      | Meta<br>(12814)       |
|----------------|---|-----------------------|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|                | RP15<br>(18454)   | RP35<br>(45242)       | RP55<br>(72823)          | RP75<br>(103504)      |                       |                       |                       |                       |                       |
| Alignment:     | <input checked="" type="radio"/>  | <input type="radio"/> | <input type="radio"/>    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Format:        | <div>FASTA</div>  |                       |                          |                       |                       |                       |                       |                       |                       |
| Order:         | <div><input checked="" type="radio"/> Tree <input type="radio"/> Alphabetical</div>                 |                       |                          |                       |                       |                       |                       |                       |                       |
| Sequence:      | <div><input checked="" type="radio"/> Inserts lower case <input type="radio"/> All upper case</div> |                       |                          |                       |                       |                       |                       |                       |                       |
| Gaps:          | <div>No gaps (unaligned)</div>  |                       |                          |                       |                       |                       |                       |                       |                       |
| Download/view: | <div><input checked="" type="radio"/> Download <input type="radio"/> View</div>                     |                       |                          |                       |                       |                       |                       |                       |                       |

Generate

Download options

We make all of our alignments available in Stockholm format. You can download them here as raw, plain text files or as [gzip](#)-compressed files.

|               | Seed<br>(181) | Full<br>(81109) | Representative proteomes |                 |                 |                  | UniProt<br>(180618) | NCBI<br>(362201) | Meta<br>(12814) |
|---------------|---------------|-----------------|--------------------------|-----------------|-----------------|------------------|---------------------|------------------|-----------------|
|               |               |                 | RP15<br>(18454)          | RP35<br>(45242) | RP55<br>(72823) | RP75<br>(103504) |                     |                  |                 |
| Raw Stockholm | ✓             | ✓               | ✓                        | ✓               | ✓               | ✓                | —                   | —                | ✓               |
| Gzipped       | ✓             | ✓               | ✓                        | ✓               | ✓               | ✓                | —                   | —                | ✓               |

You can also [download](#) a FASTA format file containing the **full-length sequences** for all sequences in the full alignment.

# Logo

## Family: *DEAD* (PF00270)

1271 architectures

81109 sequences

21 interactions

5319 species

298 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

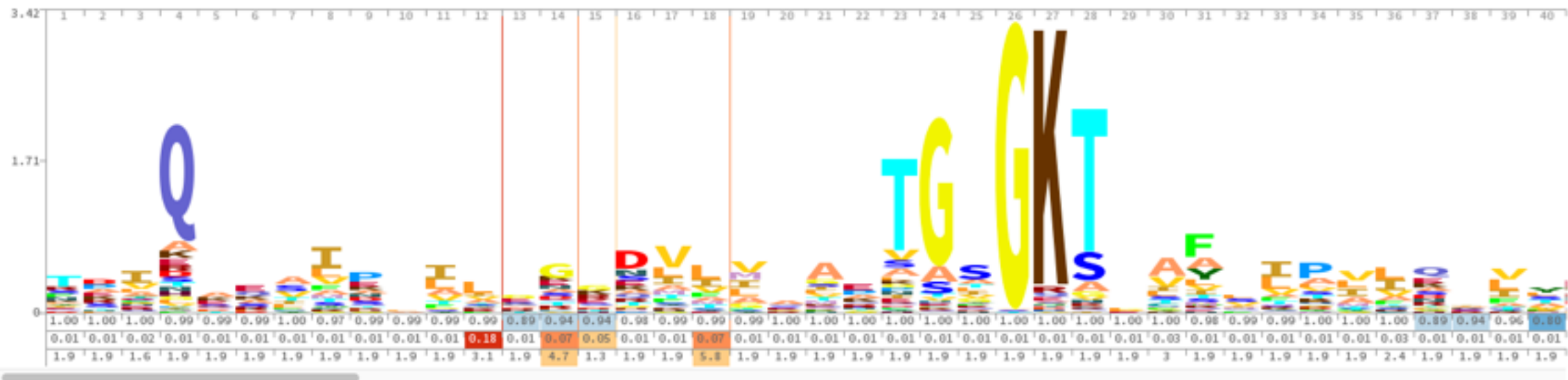
Jump to...

enter ID/acc

Go

### HMM logo

HMM logos is one way of visualising profile HMMs. Logos provide a quick overview of the properties of an HMM in a graphical form. You can see a more detailed description of HMM logos and find out how you can interpret them [here](#). [More...](#)



# Family: *DEAD* (PF00270)

Summary

Domain organisation

Clan

Alignments

HMM logo


Trees

Curation & model

Species

Interactions

Structures

Jump to... 

enter ID/acc

Go

## Curation and family details

This section shows the detailed information about the Pfam family. You can see the definitions of the various sections of the help pages.

### Curation

|   |  |
|---|--|
| Seed source:                                    | Published_alignment                      |
| Previous IDs:                                   | none                                     |
| Type:   | Domain                                   |
| Author:   | Bateman A, Bruskiewich R, Sonnhammer ELL |
| Number in seed:                                 | 181                                      |
| Number in full:                                 | 81109                                    |
| Average length of the domain:                   | 169.90 aa                                |
| Average identity of full alignment:             | 22 %                                     |
| Average coverage of the sequence by the domain: | 22.25 %                                  |

### HMM information

|                       |  |          |        |
|-----------------------|--|----------|--------|
| HMM build commands:   | <i>build method:</i> hmmbuild -o /dev/null --hand HMM SEED<br><i>search method:</i> hmmsearch -Z 26740544 -E 1000 --cpu 4 HMM pfam |          |        |
| Model details:        | Parameter  | Sequence | Domain |
|                       | Gathering cut-off  | 26.0     | 24.1   |
|                       | Trusted cut-off  | 26.0     | 24.1   |
|                       | Noise cut-off  | 25.9     | 24.0   |
| Model length:         | 176  |          |        |
| Family (HMM) version: | 28   |          |        |
| Download:             | <a href="#">download</a> the raw HMM for this family   |          |        |

# Clans

## Family: DEAD (PF00270)

1271 architectures

81109 sequences

21 interactions

5319 species

298 structures

Summary

Domain organisation

Clan

Alignments

HMM logo

Trees

Curation & model

Species

Interactions

Structures

Jump to...

enter ID/acc

Go

### Pfam Clan

This family is a member of clan [P-loop\\_NTPase](#) (CL0023), which has the following description:

AAA family proteins often perform chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes [2].

The clan contains the following 217 members:

|                                 |                                 |                                 |                              |                                 |                                 |                                 |                                 |
|---------------------------------|---------------------------------|---------------------------------|------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| <a href="#">6PF2K</a>           | <a href="#">AAA</a>             | <a href="#">AAA-ATPase_like</a> | <a href="#">AAA_10</a>       | <a href="#">AAA_11</a>          | <a href="#">AAA_12</a>          | <a href="#">AAA_13</a>          | <a href="#">AAA_14</a>          |
| <a href="#">AAA_15</a>          | <a href="#">AAA_16</a>          | <a href="#">AAA_17</a>          | <a href="#">AAA_18</a>       | <a href="#">AAA_19</a>          | <a href="#">AAA_2</a>           | <a href="#">AAA_21</a>          | <a href="#">AAA_22</a>          |
| <a href="#">AAA_23</a>          | <a href="#">AAA_24</a>          | <a href="#">AAA_25</a>          | <a href="#">AAA_26</a>       | <a href="#">AAA_27</a>          | <a href="#">AAA_28</a>          | <a href="#">AAA_29</a>          | <a href="#">AAA_3</a>           |
| <a href="#">AAA_30</a>          | <a href="#">AAA_31</a>          | <a href="#">AAA_32</a>          | <a href="#">AAA_33</a>       | <a href="#">AAA_34</a>          | <a href="#">AAA_35</a>          | <a href="#">AAA_5</a>           | <a href="#">AAA_6</a>           |
| <a href="#">AAA_7</a>           | <a href="#">AAA_8</a>           | <a href="#">AAA_9</a>           | <a href="#">AAA_PrkA</a>     | <a href="#">ABC_ATPase</a>      | <a href="#">ABC_tran</a>        | <a href="#">ABC_tran_Xtn</a>    | <a href="#">Adeno_IVa2</a>      |
| <a href="#">Adenylsucc_synt</a> | <a href="#">ADK</a>             | <a href="#">AFG1_ATPase</a>     | <a href="#">AIG1</a>         | <a href="#">APS_kinase</a>      | <a href="#">Arf</a>             | <a href="#">ArgK</a>            | <a href="#">AraA_ATPase</a>     |
| <a href="#">ATP-synt_ab</a>     | <a href="#">ATP_bind_1</a>      | <a href="#">ATP_bind_2</a>      | <a href="#">ATPase</a>       | <a href="#">ATPase_2</a>        | <a href="#">Bac_DnaA</a>        | <a href="#">BCA_ABC_TP_C</a>    | <a href="#">Beta-Casp</a>       |
| <a href="#">Cas_Csn2</a>        | <a href="#">Cas_St_Csn2</a>     | <a href="#">CblA</a>            | <a href="#">CBP_BcsQ</a>     | <a href="#">CDC73_C</a>         | <a href="#">CENP-M</a>          | <a href="#">CFTR_R</a>          | <a href="#">CLP1_P</a>          |
| <a href="#">CMS1</a>            | <a href="#">CoaE</a>            | <a href="#">CobA_CobO_BtuR</a>  | <a href="#">CobU</a>         | <a href="#">cobW</a>            | <a href="#">CPT</a>             | <a href="#">CSM2</a>            | <a href="#">CTP_synt_N</a>      |
| <a href="#">Cytidylate_kin</a>  | <a href="#">Cytidylate_kin2</a> | <a href="#">DAP3</a>            | <a href="#">DEAD</a>         | <a href="#">DEAD_2</a>          | <a href="#">DLIC</a>            | <a href="#">DNA_pack_C</a>      | <a href="#">DNA_pack_N</a>      |
| <a href="#">DNA_pol3_delta</a>  | <a href="#">DNA_pol3_delta2</a> | <a href="#">DnaB_C</a>          | <a href="#">dNK</a>          | <a href="#">DUF1611</a>         | <a href="#">DUF1726</a>         | <a href="#">DUF2075</a>         | <a href="#">DUF2326</a>         |
| <a href="#">DUF2478</a>         | <a href="#">DUF257</a>          | <a href="#">DUF2791</a>         | <a href="#">DUF2813</a>      | <a href="#">DUF3584</a>         | <a href="#">DUF463</a>          | <a href="#">DUF815</a>          | <a href="#">DUF853</a>          |
| <a href="#">DUF87</a>           | <a href="#">DUF927</a>          | <a href="#">Dynammin_N</a>      | <a href="#">Dynein_heavy</a> | <a href="#">ERCC3_RAD25_C</a>   | <a href="#">Exonuc_V_gamma</a>  | <a href="#">FeoB_N</a>          | <a href="#">Fer4_NifH</a>       |
| <a href="#">Flavi_DEAD</a>      | <a href="#">FTHFS</a>           | <a href="#">FtsK_SpoIIIE</a>    | <a href="#">G-alpha</a>      | <a href="#">Gal-3-0_sulfotr</a> | <a href="#">GBP</a>             | <a href="#">GBP_C</a>           | <a href="#">GTP_EFTU</a>        |
| <a href="#">Gtr1_RagA</a>       | <a href="#">Guanylate_kin</a>   | <a href="#">GvpD</a>            | <a href="#">HDA2-3</a>       | <a href="#">Helicase_C</a>      | <a href="#">Helicase_C_2</a>    | <a href="#">Helicase_C_4</a>    | <a href="#">Helicase_RecD</a>   |
| <a href="#">Herpes_Helicase</a> | <a href="#">Herpes_ori_bp</a>   | <a href="#">Herpes_TX</a>       | <a href="#">Hydin_ADK</a>    | <a href="#">IIGP</a>            | <a href="#">IPPT</a>            | <a href="#">IPT</a>             | <a href="#">IstB_IS21</a>       |
| <a href="#">KAP_NTPase</a>      | <a href="#">KdpD</a>            | <a href="#">Kinesin</a>         | <a href="#">KTI12</a>        | <a href="#">LAP1C</a>           | <a href="#">Lon_2</a>           | <a href="#">LpxK</a>            | <a href="#">MCM</a>             |
| <a href="#">MEDS</a>            | <a href="#">Mq_chelatase</a>    | <a href="#">Microtub_bd</a>     | <a href="#">MlpZ</a>         | <a href="#">MMR_HSR1</a>        | <a href="#">MMR_HSR1_C</a>      | <a href="#">MobB</a>            | <a href="#">MukB</a>            |
| <a href="#">MutS_V</a>          | <a href="#">Myosin_head</a>     | <a href="#">NACHT</a>           | <a href="#">NB-ARC</a>       | <a href="#">NOG1</a>            | <a href="#">NTPase_1</a>        | <a href="#">NTPase_P4</a>       | <a href="#">ORC3_N</a>          |
| <a href="#">ParA</a>            | <a href="#">Parvo_NS1</a>       | <a href="#">PAXNEB</a>          | <a href="#">PduV-EutP</a>    | <a href="#">PhoH</a>            | <a href="#">PIF1</a>            | <a href="#">Podovirus_Gp16</a>  | <a href="#">Polyoma_Ig_T_C</a>  |
| <a href="#">Pox_A32</a>         | <a href="#">PPK2</a>            | <a href="#">PPV_E1_C</a>        | <a href="#">PRK</a>          | <a href="#">PSY3</a>            | <a href="#">Rad17</a>           | <a href="#">Rad51</a>           | <a href="#">Ras</a>             |
| <a href="#">RecA</a>            | <a href="#">ResIII</a>          | <a href="#">RHD3</a>            | <a href="#">RHSP</a>         | <a href="#">RNA12</a>           | <a href="#">RNA_helicase</a>    | <a href="#">Roc</a>             | <a href="#">RsqA_GTPase</a>     |
| <a href="#">RuvB_N</a>          | <a href="#">SbcCD_C</a>         | <a href="#">SecA_DEAD</a>       | <a href="#">Septin</a>       | <a href="#">Sigma54_activ_2</a> | <a href="#">Sigma54_activat</a> | <a href="#">SKI</a>             | <a href="#">SMC_N</a>           |
| <a href="#">SNF2_N</a>          | <a href="#">Spore_IV_A</a>      | <a href="#">SRP54</a>           | <a href="#">SRPRB</a>        | <a href="#">SulA</a>            | <a href="#">Sulfotransfer_1</a> | <a href="#">Sulfotransfer_2</a> | <a href="#">Sulfotransfer_3</a> |
| <a href="#">Sulphotransf</a>    | <a href="#">T2SSE</a>           | <a href="#">T4SS-DNA_transf</a> | <a href="#">Terminase_1</a>  | <a href="#">Terminase_3</a>     | <a href="#">Terminase_6</a>     | <a href="#">Terminase_GpA</a>   | <a href="#">Thymidylate_kin</a> |
| <a href="#">TIP49</a>           | <a href="#">TK</a>              | <a href="#">TniB</a>            | <a href="#">Torsin</a>       | <a href="#">TraG-D_C</a>        | <a href="#">tRNA_lig_kinase</a> | <a href="#">TrwB_AAD_bind</a>   | <a href="#">TsaE</a>            |
| <a href="#">UvrB</a>            | <a href="#">UvrD-helicase</a>   | <a href="#">UvrD_C</a>          | <a href="#">UvrD_C_2</a>     | <a href="#">Viral_helicase1</a> | <a href="#">VirC1</a>           | <a href="#">VirE</a>            | <a href="#">Zeta_toxin</a>      |
| <a href="#">Zot</a>             |                                 |                                 |                              |                                 |                                 |                                 |                                 |



## Summary

## Domain organisation

### Clan

### Alignments

### HMM logo

### Trees

### Curation & model

### Species

### Interactions

### Structures

## Jump to...



## Domain organisation

Below is a listing of the unique domain organisations or architectures in which t

**There are 33324 sequences with the following architecture: DEAD**

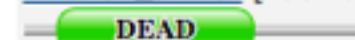
[X7F901\\_9RHOB](#) [Roseivivax isopora LMG 25204] DEAD/DEAH box helicase {E



[Show](#) all sequences with this architecture.

**There are 4751 sequences with the following architecture: DEAD**

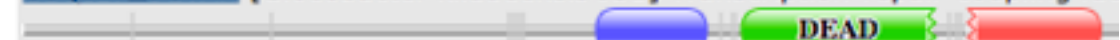
[W7YL34\\_9BACL](#) [Paenibacillus pini JCM 16418] ATP-dependent RNA helicase Y



[Show](#) all sequences with this architecture.

**There are 3666 sequences with the following architecture: Card**

[U5QIE7\\_9CYAN](#) [Gloeobacter kilaueensis JS1] Transcription-repair-coupling fac



[Show](#) all sequences with this architecture.

**There are 3518 sequences with the following architecture: DEAD, Helicase\_C, HA2, OB\_NTP\_bind**

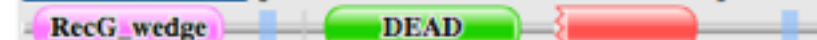
[W9YGA6\\_9EURO](#) [Capronia coronata CBS 617.96] Adenosinetriphosphatase {ECO:0000313|EMBL:EXJ88710.1} (980 residues)



[Show](#) all sequences with this architecture.

**There are 3224 sequences with the following architecture: RecG\_wedge, DEAD, Helicase\_C**

[W7UXIO\\_RUMFL](#) [Ruminococcus flavefaciens 007c] ATP-dependent DNA helicase RecG {ECO:0000256|RuleBase:RU363016} (678 residues)



[Show](#) all sequences with this architecture.

**There are 2988 sequences with the following architecture: DEAD, Helicase\_C, DUF4217**

[W1QJU9\\_OGAPD](#) [Ogataea parapolymorpha (strain ATCC 26012 / BCRC 20466 / JCM 22074 / NRRL Y-7560 / DL-1) (Yeast) (Hansenula polymorpha)] {ECO:0000313|EMBL:ESX02922.1} (742 residues)



[Show](#) all sequences with this architecture.

**There are 2919 sequences with the following architecture: DEAD, Helicase\_C, DbpA**

[A0A140L5S6\\_9CLOT](#) [Thermotalea metallivorans] DEAD-box ATP-dependent RNA helicase CshA {ECO:0000313|EMBL:KXG75901.1} (528 residues)



[Show](#) all sequences with this architecture.

**There are 2443 sequences with the following architecture: DEAD, Helicase\_C, RecQ\_Zn\_bind, RQC, HRDC**

[G5GBG5\\_9BACT](#) [Alloprevotella rava F0323] ATP-dependent DNA helicase RecQ {ECO:0000313|EMBL:EHG23164.1} (608 residues)



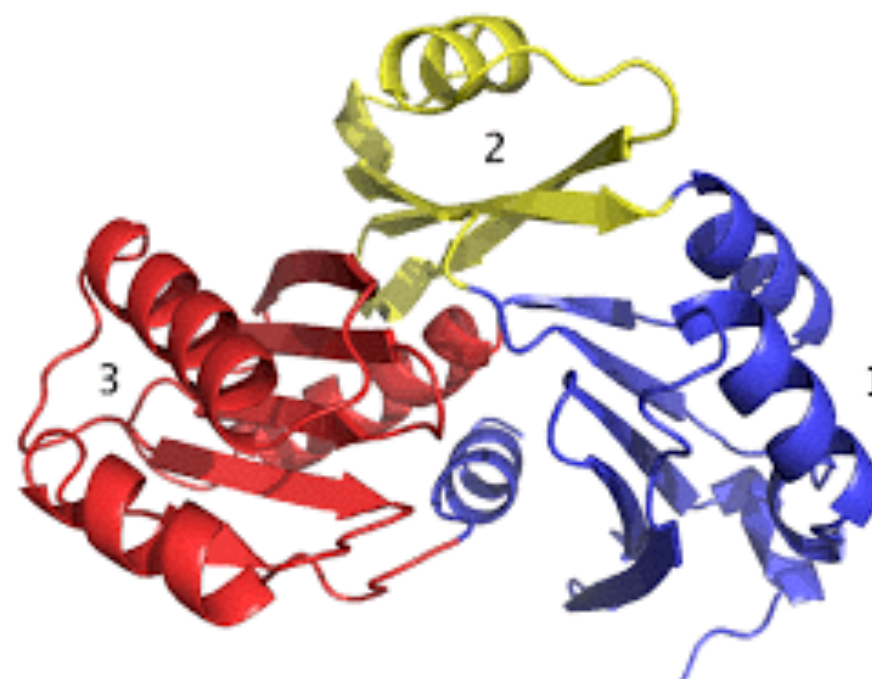
[Show](#) all sequences with this architecture.

**There are 1925 sequences with the following architecture: DEAD, Helicase\_C, RecQ\_Zn\_bind**

[I3C0K8\\_9FLAO](#) [Joostella marina DSM 19592] ATP-dependent DNA helicase, RecQ family {ECO:0000313|EMBL:EIJ37151.1} (635 residues)

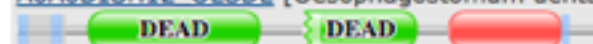


[Show](#) all sequences with this architecture.

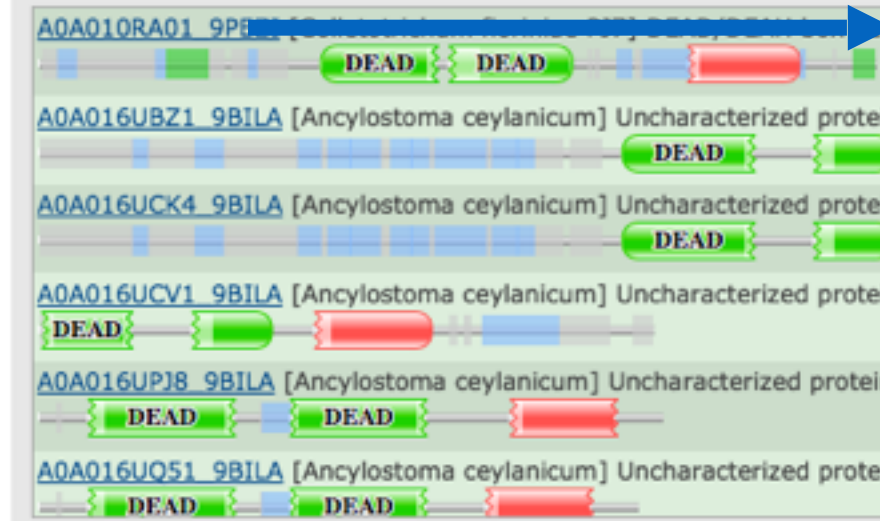


There are 390 sequences with the following architecture:

[A0A0B1SH12\\_OESDE](#) [Oesophagostomum dentatum (Nodular worm)] D



[Hide](#) all sequences with this architecture.



## Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will show the domains. [More...](#)



[Download](#) the data used to generate the domain graphic in JSON format.

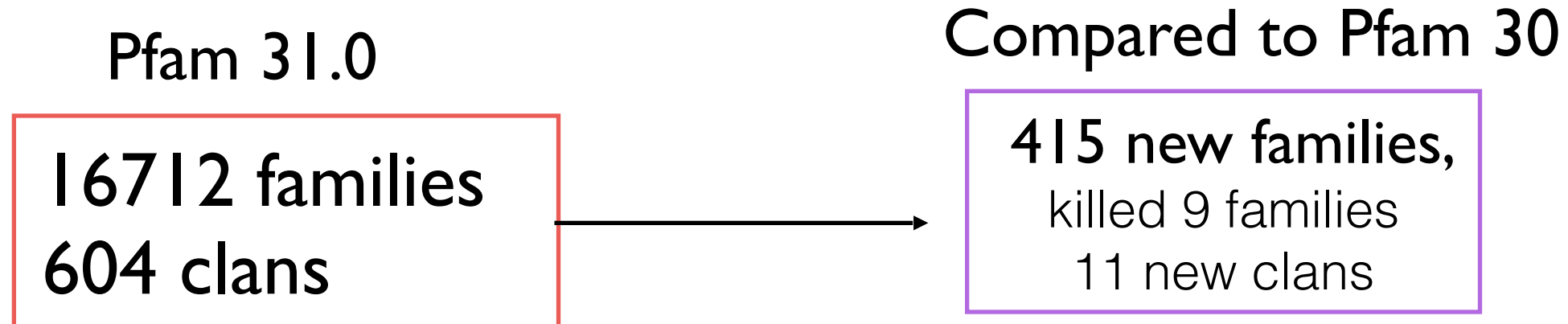
| Source         | Domain                     | Start | End | Gathering threshold (bits) |        | Score (bits) |        | E-value  |         |
|----------------|----------------------------|-------|-----|----------------------------|--------|--------------|--------|----------|---------|
|                |                            |       |     | Sequence                   | Domain | Sequence     | Domain | Sequence | Domain  |
| disorder       | n/a                        | 3     | 181 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 17    | 37  | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 114   | 137 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| coiled_coil    | n/a                        | 125   | 164 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 142   | 163 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 190   | 234 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 205   | 212 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 238   | 243 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| Pfam           | <a href="#">DEAD</a>       | 277   | 391 | 26.00                      | 24.10  | 112.30       | 61.20  | 3.1e-29  | 1.6e-13 |
| disorder       | n/a                        | 379   | 380 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| Pfam           | <a href="#">DEAD</a>       | 403   | 523 | 26.00                      | 24.10  | 112.30       | 48.90  | 3.1e-29  | 9.4e-10 |
| low_complexity | n/a                        | 404   | 417 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 405   | 406 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 411   | 419 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 539   | 540 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 546   | 549 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 567   | 581 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 585   | 640 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 591   | 644 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| Pfam           | <a href="#">Helicase_C</a> | 637   | 747 | 20.90                      | 20.90  | 70.60        | 69.00  | 1.9e-16  | 6.2e-16 |
| low_complexity | n/a                        | 659   | 674 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 681   | 683 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 688   | 713 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| low_complexity | n/a                        | 740   | 751 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 779   | 781 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| coiled_coil    | n/a                        | 799   | 819 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |
| disorder       | n/a                        | 813   | 820 | n/a                        | n/a    | n/a          | n/a    | n/a      | n/a     |

[Show](#) or [hide](#) domain scores.



# Pfam 31.0

- Based on Uniprot reference proteomes contains 26.7 million sequences,
- 73% of Uniprot reference proteomes have a match to at least one Pfam
- 48% of all residues fall within a Pfam family.



# Manually curated gathering thresholds

## cut\_ga

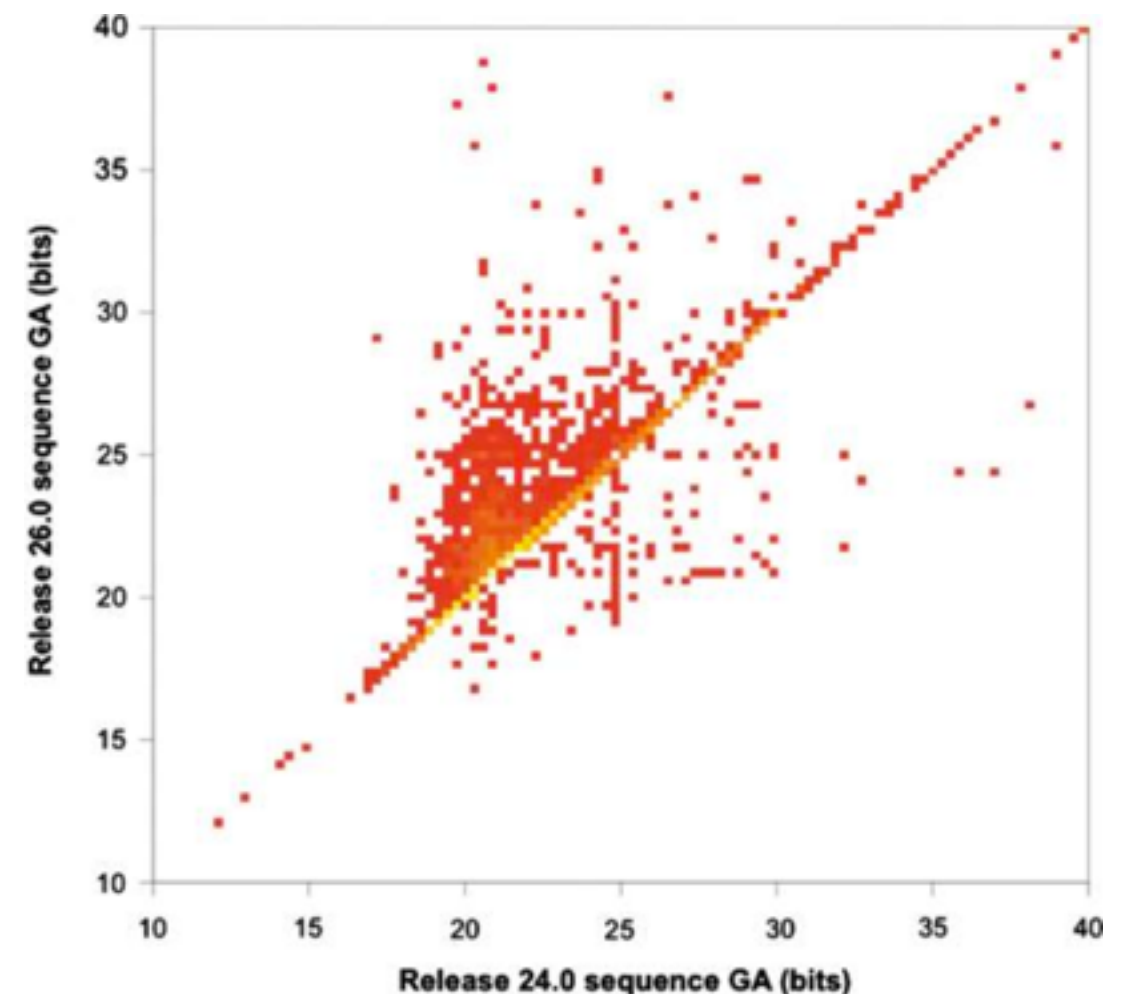
- Each Pfam entry has gathering thresholds (GAs)
- All sequence regions that score above the GAs are included in the full alignment for the family.
- GAs, are **manually curated**, family-specific, bit score thresholds that are chosen by Pfam curators when a family is built.
- Every family has two GAs, a 'sequence' threshold, and a 'domain' threshold.

# Criteria for gathering threshold assignment

- GAs are chosen to maximise coverage and exclude any false positive matches.
- The number of false positives is generally unknown, but we can check for overlaps between one Pfam family and another.
- If the same region of a sequence matches two Pfam families, it should be considered a false positive in one of them.
- Another way is to generate artificial false positive sequences.
  - hmmemit and shuffling
  - Domain sequences (same family or not) and shuffling

# Gathering threshold for new families

- When building a new family, the GA choice is often influenced by overlaps with other families.
- Overlap-resolution between old and new families leads to GAs modifications.
- GAs for families in Pfam release 24.0 with GAs of the same families in release 26.0. Overall, 13% of GAs have changed,



# Gathering threshold x E-values

