

Find Functions: Pfam

Pfam Browser

From the **Find Functions** top-level menu, the **Pfam** option on the second-level menu leads to the Pfam Browser, as shown in Figure 1(i). This is a two-level display with Pfam categories (which were derived from COG categories) shown on the first level, and Pfam pathways listed under the corresponding categories.

Top-level Pfam categories include:

- Amino acid transport and metabolism [E]
- Carbohydrate transport and metabolism [G]
- Cell cycle control, cell division, chromosome partitioning [D]
- Cell motility [N]
- Cell wall/membrane/envelope biogenesis [M]
- Chromatin structure and dynamics [B]
- Coenzyme transport and metabolism [H]
- Cytoskeleton [Z]
- Defense mechanisms [V]
- Energy production and conversion [C]
- Extracellular structures [W]
- Function unknown [S]
- General function prediction only [R]
- Inorganic ion transport and metabolism [P]
- Intracellular trafficking, secretion, and vesicular transport [U]
- Lipid transport and metabolism [I]
- Nuclear structure [Y]
- Nucleotide transport and metabolism [F]
- Posttranslational modification, protein turnover, chaperones [O]
- RNA processing and modification [A]
- Replication, recombination and repair [L]
- Secondary metabolites biosynthesis, transport and catabolism [Q]
- Signal transduction mechanisms [T]
- Transcription [K]
- Translation, ribosomal structure and biogenesis [J]

Clicking on any of Pfam category will lead to a page showing all Pfam IDs associated with this category and the corresponding isolate genome and metagenome counts (Figure 1(ii)). Clicking on any of the

Pfam pathway will lead to a page showing all Pfam IDs associated with this pathway and the corresponding isolate genome and metagenome counts (Figure 1(iii)).

The screenshot displays the Pfam Browser interface, divided into three main sections:

- (i) Pfam Browser:** A sidebar on the left containing a list of Pfam categories. The category "Amino acid transport and metabolism" is highlighted with a red box. Below it, "Leucine biosynthesis" is also highlighted with a red box. Other categories include Arginine biosynthesis, Histidine biosynthesis, Isoleucine biosynthesis, Methionine biosynthesis, Phenylalanine/tyrosine biosynthesis, Proline biosynthesis, Threonine biosynthesis, Tryptophan biosynthesis, Valine biosynthesis, Carbohydrate transport and metabolism [G], Entner-Doudoroff pathway, Gluconeogenesis, Glycolysis, Pentose phosphate pathway, Cell cycle control, cell division, chromosome parti..., and Cell motility [N] with a sub-item Flagellum structure and biogenesis.
- (ii) Pfam Category Details:** A panel titled "Pfam Category Details" for "Amino acid transport and metabolism". It features a table with columns for Pfam ID, Pfam Name, Isolate Genome Count, and Metagenome Count. The table lists several Pfam IDs, including pfam00060, pfam00171, pfam00208, pfam00217, and pfam00218. A red arrow points from the "Amino acid transport and metabolism" category in section (i) to this panel.
- (iii) Pfam Pathway Details:** A panel titled "Pfam Pathway Details" for "Leucine biosynthesis". It features a table with columns for Pfam ID, Pfam Name, Isolate Genome Count, and Metagenome Count. The table lists Pfam IDs pfam00920, pfam01063, pfam01450, and pfam07836. A red arrow points from the "Leucine biosynthesis" category in section (i) to this panel.

Figure 1. Pfam Browser.

Click on any of the Pfam IDs will lead to corresponding Pfam definition page at EMBL-EBI (see Figure 4(iii)); for example, <http://pfam.xfam.org/family/PF00060>

Users can select any subset of listed Pfam IDs to be saved to Function Cart or Workspace Function Set for further analysis.

Users can also click on a genome count to view all isolate genomes or metagenomes having genes associated with a particular Pfam ID. For example, Figure 2(ii) lists all isolate genomes having genes annotated with *pfam00060*. Clicking on a gene count will lead to a page listing the actual genes (see Figure 2(iii)). Genomes and genes can be selected to add to analysis carts or Workspace for further analysis.

(iii)

Select	Gene ID	Locus Tag	Gene Product Name	Genome ID	Genome Name
<input type="checkbox"/>	2623437816	Ga0070236_103262	amino acid ABC transporter substrate-binding protein, PAAT family	2622736589	Nautella italica DSM 26436

(ii)

Isolate Genomes with pfam00060

Genomes with Lig_chan - Ligand-gated ion channel

Phylogenetic Distribution

Domains(D): * = Microbiome,
 B = Bacteria, A = Archaea, E = Eukarya, P = Plasmids, G = GFragment, V = Viruses.
 Genome Completion(C): F = Finished, P = Permanent Draft, D = Draft.

Add Selected to Genome Cart Select All Clear All

Filter column: Domain Filter text Apply

Export Page 1 of 10 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >>

Column Selector Select Page Deselect Page

Select	Domain	Status	Genome	Gene Count
<input type="checkbox"/>	B	P	Nautella italica DSM 26436	1
<input type="checkbox"/>	B	P	Vibrio campbellii CAIM1500	1
<input type="checkbox"/>	B	F	Synechococcus sp. WH 8103 WH8103	1
<input type="checkbox"/>	B	F	Altererythrobacter atlanticus 26DY36	1

(i)

Pfam Category Details

Details for Pfam Category: *Amino acid transport and metabolism*

Add Selected to Function Cart Select All

Filter column: Pfam ID Filter text Apply

Export Page 1 of 2 << first < prev 1 2 next > last >>

Column Selector Select Page Deselect Page

Select	Pfam ID	Pfam Name	Isolate Genome Count	Metagenome Count
<input type="checkbox"/>	pfam00060	Ligand-gated ion channel	921	3767
<input type="checkbox"/>	pfam00171	Aldehyde dehydrogenase family	50085	7950
<input type="checkbox"/>	pfam00208	Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	40702	7490
<input type="checkbox"/>	pfam00217	ATP:guanido phosphotransferase, C-terminal catalytic domain	10184	6207

Figure 2. *pfam00060* Detail and Associated Genomes and Genes.

Pfam List

The **Pfam List** option shows all Pfam IDs in a table display (Figure 3(i)).

Users can select a subset of Pfam IDs to be added to Function Cart for further analysis.

Click on any of the Pfam IDs will lead to corresponding Pfam definition page at EMBL-EBI (see Figure 4(iii)).

(i)

Buttons: Add Selected to Function Cart, Select All, Clear All

Filter column: Pfam ID Filter text

Page 1 of 164

Select	Pfam ID	Pfam Name
<input type="checkbox"/>	pfam00001	7tm_1 - 7 transmembrane receptor (rhodopsin family)
<input type="checkbox"/>	pfam00002	7tm_2 - 7 transmembrane receptor (Secretin family)
<input type="checkbox"/>	pfam00003	7tm_3 - 7 transmembrane sweet-taste receptor of 3 GCPR
<input type="checkbox"/>	pfam00004	AAA - ATPase family a
<input type="checkbox"/>	pfam00005	ABC_tran - ABC transp
<input type="checkbox"/>	pfam00006	ATP-synt_ab - ATP synt
<input type="checkbox"/>	pfam00007	Cys_knot - Cystine-kn

(ii)

The statistics was last updated on: Thu, Jun 1 2017 3:07 am

Buttons: Add Selected to Function Cart, Select All, Clear All

Filter column: Pfam Name Filter text

Page 1 of 164

Select	Pfam ID	Pfam Name	Isolate Genome Count	Metagenome Count
<input type="checkbox"/>	pfam00001	7tm_1 - 7 transmembrane receptor (rhodopsin family)	104	3198
<input type="checkbox"/>	pfam00002	7tm_2 - 7 transmembrane receptor (Secretin family)	73	2647
<input type="checkbox"/>	pfam00003	7tm_3 - 7 transmembrane sweet-taste receptor of 3 GCPR	58	3206
<input type="checkbox"/>	pfam00004	AAA - ATPase family associated with various cellular activities (AAA)	53949	8014
<input type="checkbox"/>	pfam00005	ABC_tran - ABC transporter	54584	8120

Figure 3. Pfam List.

Pfam List w/ Stats

This option is similar to **Pfam List** except that there are additional genome counts (Figure 3(ii)). Users can click on a genome count to view the list of genomes or metagenomes having genes annotated with a particular Pfam ID (Figure 2(ii)). Please note that genome counts are pre-computed and may not be up-to-date.

Pfam Clans

The **Pfam Clans** option lists all Pfam clans and their corresponding Pfam IDs (Figure 4(i)).

Clicking on a clan name will lead to the corresponding clan definition page in EMBL-EBI (Figure 4(ii)).

Clicking on a Pfam ID will lead to the corresponding family definition page in EMBL-EBI (Figure 4(iii)).

The image shows three panels illustrating the Pfam website interface:

- (i) Pfam Clans:** A list of Pfam clans with checkboxes for selection. A red box highlights the entry for "2Fe-2S iron-sulfur cluster binding domain" with Pfam ID [pfam10418](#). Below it, the "2H phosphoesterase superfamily" is also listed with several Pfam IDs.
- (ii) Clan: *Fer2* (CL0486):** The clan definition page for the 2Fe-2S iron-sulfur cluster binding domain. It includes a summary, domain organization, and statistics: 375 architectures, 42533 sequences, 51 interactions, 4752 species, and 488 structures.
- (iii) Family: *DHODB_Fe-S_bind* (PF10418):** The family definition page for the Iron-sulfur cluster binding domain of dihydroorotate dehydrogenase B. It includes a summary, domain organization, and statistics: 32 architectures, 3006 sequences, 3 interactions, 1703 species, and 10 structures. A 3D protein structure is shown on the right.

Red arrows indicate the flow from the Pfam Clans list to the clan definition page (ii) and from the Pfam ID [pfam10418](#) to the family definition page (iii).

Figure 4. Pfam Clan

Users can select a subset of Pfam IDs to be added to Function Cart for further analysis.