

**Pfam** è una banca dati di famiglie proteiche creata dal Wellcome Trust Sanger Institute a partire dal database Pfamseq basato su Uniprot. Le famiglie proteiche di Pfam sono state ottenute dall'analisi di sequenze proteiche tramite multi allineamento e il programma HHMER3, che permette di tracciare i profili Hidden Markov Models relativi ai multi allineamenti.



HOME | SEARCH | BROWSE | FTP | HELP | ABOUT



Pfam 24.0 (October 2009, 11912 families)

The Pfam database is a large collection of protein families, each represented by **multiple sequence alignments** and **hidden Markov models (HMMs)**. [More...](#)

QUICK LINKS	YOU CAN FIND DATA IN PFAM IN VARIOUS WAYS...
<a href="#">SEQUENCE SEARCH</a>	Analyze your protein sequence for Pfam matches
<a href="#">VIEW A PFAM FAMILY</a>	View Pfam family annotation and alignments
<a href="#">VIEW A CLAN</a>	See groups of related families
<a href="#">VIEW A SEQUENCE</a>	Look at the domain organisation of a protein sequence
<a href="#">VIEW A STRUCTURE</a>	Find the domains on a PDB structure
<a href="#">KEYWORD SEARCH</a>	Query Pfam by keywords
<a href="#">JUMP TO</a>	<input type="text" value="enter any accession or ID"/> <input type="button" value="Go"/> <a href="#">Example</a>

Enter any type of accession or ID to jump to the page for a Pfam family or clan, UniProt sequence, PDB structure, etc.

Or view the [help](#) pages for more information

Il database è costituito da due componenti :

Pfam A , le cui entries sono curate manualmente, in modo da garantire un certo standar di qualità e una buona copertura delle sequenze durante l'allineamento;

Pfam B , le cui entries sono generate grazie ad ADDA database, ovvero Automatic Domain Decomposition Algorithm. Questo database permette di individuare e clasterizzare famiglie di domini proteici in base ad allineamenti di sequenze tutti-contro-tutti utilizzando il parametro di maximum likelihood.

Pfam permette, inoltre, di identificare gruppi di famiglie proteiche correlate, detti clan. Ogni clan è costituito da una raccolta di entris di Pfam A con sequenza, struttura o profilo HMM simili.

Le entries di Pfam sono classificate con 4 definizioni:

- Famiglia: raggruppamento di proteine correlate
- Dominio: unità strutturale che può essere riscontrata in diverse proteine
- Repeat: unità ripetuta, instabile se isolata singolarmente, ma generalmente stabile quando forma dei complessi proteici con altre unità uguali
- Motivo: unità strutturale non appartenente a domini proteici globulari

Dalla home page del sito possiamo accedere a:

Search Pfam: pagina da dove possiamo selezionare diverse modalità di ricerca a seconda dei dati di partenza in possesso. La struttura di questa pagina verrà analizzata in seguito.

Browser : da cui si può visualizzare liste di famiglie proteiche, clan o proteomi selezionando la prima lettera/cifra dell'ID relativo. Nel caso delle famiglie proteiche, è presente un interlink ([twenty largest families](#)) che apre una pagina dove è presente una tabella delle famiglie proteiche più grandi per numero di sequenze contenute.

ID	Accession	Type	Number of sequences		Average length	Average %id	Average coverage	Has 3D	Change status	Description
			Seed	Full						
<a href="#">ABC_tran</a>	<a href="#">PF00005</a>	Domain	60	127458	126.40	24	26.88	✓	Changed	ABC transporter
<a href="#">RVT_1</a>	<a href="#">PF00078</a>	Family	155	105557	170.10	67	42.89	✓	Changed	Reverse transcriptase (RNA-dependent)
<a href="#">GP120</a>	<a href="#">PF00516</a>	Family	24	94479	201.30	52	81.84	✓	Changed	Envelope glycoprotein GP120
<a href="#">WD40</a>	<a href="#">PF00400</a>	Repeat	1843	93618	37.90	22	5.47	✓	Changed	WD domain, G-beta repeat
<a href="#">COX1</a>	<a href="#">PF00115</a>	Family	94	91909	235.20	47	92.89	✓	Changed	Cytochrome C and Quinol oxidase polyp
<a href="#">RVP</a>	<a href="#">PF00077</a>	Domain	50	76107	94.20	85	29.40	✓	Changed	Retroviral aspartyl protease
<a href="#">Response_reg</a>	<a href="#">PF00072</a>	Domain	57	62074	111.60	26	27.71	✓	Changed	Response regulator receiver domain

**FTP:** permette di effettuare download del materiale contenuto nel database, come ad esempio i tool utilizzati o gli articoli correlati alle diverse famiglie proteiche.

**HELP:** da spiegazioni molto utili sulla terminologia di Pfam e su come utilizzare le informazioni date dal database. Una nota di rilievo va alle spiegazioni riguardo ai punteggi ottenuti con i diversi tool, in particolare ai profili HMMs.

**Scores**

[Citing Pfam](#)

[Linking to Pfam](#)

[Guide to Graphics](#)

[Tools & Services](#)

[RESTful interface](#)

[Pfam database](#)

[FTP site](#)

[Website installation](#)

**What Pfam HMM scores mean**

Pfam-A is based around hidden Markov model (HMM) searches, as provided by the [HMMER2](#) package. In HMMER2, like BLAST, E-values (expected number of hits) are calculated. The E-value is the number of hits that would be expected to have a score equal or better than this by chance alone. A good E-value is around 1. Around 1 is what we expect just by chance. In principle, all you need to decide on the significance of a match is the E-value.

However, there are a few complications.

The most serious complication is that there are no analytical results available for accurately determining E-values for gapped alignments, especially with HMMER. HMMER uses empirical methods to estimate E-values. These methods are generally rather accurate. However, when in doubt, HM is on the conservative side.

We use a second, and even more empirical, system in maintaining Pfam models. This system is implemented in the Pfam database rather than in the software. For each Pfam family, we record a "trusted cutoff" and a "noise cutoff", TC1 and NC1. TC1 is the lowest score for sequences we include in the Full alignment. NC1 is the highest score for sequences we did not include in the Full alignment. (Since Full alignments are produced with a trusted sequence cutoff is always greater than the noise sequence cutoff.)

**ABOUT :** alcune informazioni sulla costruzione del database e link dei mirror site.

### ***RICERCA NELLA BANCA DATI***

La ricerca in Pfam può essere effettuata direttamente dalla home avendo il codice Pfam o PDB della proteina, dominio o superfamiglia di proteine; oppure possiamo, come precedentemente accennato, utilizzare la pagina Pfam search a partire da diversi tipi di query.

## Search Pfam

0 architectures
0 sequences
0 interactions
0 species

**Sequence**

[Batch search](#)

[Keyword](#)

[Functional similarity](#)

[Domain architecture](#)

[DNA sequence](#)

[Taxonomy](#)

**Jump to...**

enter ID/accession

**Sequence search**

Find Pfam families within your sequence of interest. Paste your protein sequence into the box below, to have it searched for matching Pfam families.

Sequence

Cut-off  Gathering threshold  Use E-value

E-value

Search for PfamBs

1) *SEQUENCE*: è sufficiente copiare la nostra sequenza aminoacidica ( non vengono riconosciute quelle nucleotidiche ), non è necessariamente in formato FASTA.

Di default vengono usati valori soglia (threshold) per la ricerca HMM ed E-value =1.0; tuttavia si può impostare l'opzione gathering threshold per ogni HMM, ovvero il valore soglia con il quale si decide se aggiungere o meno (clusterizzare) una sequenza all'allineamento Pfam effettuato, specificando un valore di E-value a piacere ( dovrà essere sempre maggiore di zero).

2) *BATCH SEARCH*: si può caricare una serie di sequenze proteiche in formato multi FASTA per cercare la corrispondenza di queste con le famiglie proteiche presenti in Pfam. Il risultato viene inviato all'indirizzo mail specificato.

Per il formato FASTA ci sono alcune restrizioni:

- Il programma utilizzato per ottenere i profili HMM utilizza soltanto i primi 60 caratteri presenti su una linea; è perciò consigliabile che ogni sequenza sia composta da linee di al massimo 60 caratteri
- si possono inserire al massimo 5000 sequenze costituite da 6 a 2000 caratteri

Questo tipo di ricerca utilizza delle euristiche per individuare quelle sequenze che hanno un livello ragionevole di variabilità in modo da evitare che vengano sequenze ridondanti o con residui duplicati.

3) *KEYWORD*: effettua la ricerca di similarità con le famiglie proteiche presenti nel database a partire parole chiave diverse: Pfam entries, specie o descrizione della proteina , HEADER e TITLE del formato PDB inerente, ID di Gene Ontology oppure Abstract di InterPro

4) *FUNCTIONAL SIMILARITY*: ricerca di similarità funzionale tra famiglie proteiche a partire dall'ID oppure dall'accession di Pfam. Il calcolo della similarità è basato su Gene Ontology e assume come soglia arbitraria del punteggio di similarità di il valore 0,75 .

5) *DOMAIN ARCHITECTURE*: ricerca di domini in base all'architettura, ovvero il numero e la disposizione degli elementi di struttura secondaria nel dominio. Ci sono tre riquadri:

- Pfam domains: cliccando nella tab soprastante, Pfam inserirà automaticamente una lista di famiglie proteiche in relazione alla lettera selezionata( nel caso del tasto top twenty vengono visualizzate le più comuni). Tra queste possiamo scegliere le proteine di nostro interesse utilizzando i tasti "include" e " does not include"
- Architecture Must Include: sono presenti le famiglie proteiche che abbiamo selezionato come "include" nel riquadro precedente. Possiamo eliminarle dalla lista con il tasto "remove"

- Architecture Must Not Include: sono presenti le famiglie proteiche che abbiamo scartato con “does not include” nella lista Pfam domain . Possiamo eliminarle dalla lista con il tasto “remove”

L’output permette di visualizzare tutte le sequenze che hanno in comune quel particolare dominio proteico, mostrando anche il grafico in cui viene riportata la disposizione reciproca dei domini nella nostra sequenza. In particolare vengono fornite le seguenti informazioni:

- Numero totale di sequenze che hanno in comune il dominio proteico richiesto o un multi dominio che lo contenga.
- ID identificativo del dominio grazie al quale si può accedere alla pagina di Pfam correlata
- L’organismo dal quale è stato isolato
- La rappresentazione grafica del dominio all’interno della sequenza proteica
- Show: visualizzazione grafica delle sequenze contenenti il dominio

There are 21 sequences with the following architecture: **A-2\_8-polyST**

[Q8VRL9\\_NEIME](#) [Neisseria meningitidis] SiaD (492 residues)

**A-2\_8-polyST**

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

[Q8VRL9\\_NEIME](#) [Neisseria meningitidis] SiaD (492 residues)

**A-2\_8-polyST**

[Q06435\\_NEIME](#) [Neisseria meningitidis] SynE (492 residues)

**A-2\_8-polyST**

[Q51281\\_NEIME](#) [Neisseria meningitidis] SiaD protein (495 residues)

**A-2\_8-polyST**

[Q47404\\_ECOLX](#) [Escherichia coli] Glycosyl transferase (409 residues)

**A-2\_8-polyST**

[Q57269\\_ECOLX](#) [Escherichia coli] Polysialyltransferase protein (409 residues)

**A-2\_8-polyST**

[Q83TU5\\_NEIME](#) [Neisseria meningitidis] Sialyltransferase (Fragment) (467 residues)

**A-2\_8-polyST**

6) *DNA SEQUENCE*: ricerca di domini presenti in Pfam a partire da una sequenza nucleotidica utilizzando il tool Wise 2 di EBI; il risultato viene inviato all'indirizzo mail specificato.

### ESEMPIO DI RICERCA IN Pfam

Prendiamo come esempio il dominio PCMT. Inviando la nostra query apparirà in output la pagina inerente al nostro dominio da dove possiamo ricavare diverse informazioni. Possiamo accedere alle informazioni che ci interessano sia dalla tool bar, cliccando sulle icone, sia dal menù laterale. Nel menù laterale troviamo

**Family: PCMT (PF01135)**

18 architectures 1614 sequences 1 interaction 628 species 18 structures

**Summary**

**Protein-L-isoaspartate(D-aspartate) O-methyltransferase (PCMT)** [Add annotation](#)

No Pfam abstract.

InterPro entry [IPR000682](#)

Protein-L-isoaspartate(D-aspartate) O-methyltransferase () (PCMT) [PUBMED:9253175](#) (which is also known as L-isoaspartyl protein carboxyl methyltransferase) is an enzyme that catalyses the transfer of a methyl group from S-adenosylmethionine to the free carboxyl groups of D-aspartyl or L-isoaspartyl residues in a variety of peptides and proteins. The enzyme does not act on normal L-aspartyl residues L-isoaspartyl and D-aspartyl are the products of the spontaneous deamidation and/or isomerisation of normal L-aspartyl and L-asparaginyl residues in proteins. PCMT plays a role in the repair and/or degradation of these damaged proteins; the enzymatic methyl esterification of the abnormal residues can lead to their conversion to normal L-aspartyl residues. The SAM domain is present in most of these proteins.

**Clan**

This family is a member of clan [NADP\\_Rossmann \(CL0063\)](#), which contains the following 148 members:

<a href="#">2-Hacid_dh_C</a>	<a href="#">3Beta_HSD</a>	<a href="#">3HCDH_N</a>	<a href="#">adh_short</a>	<a href="#">ADH_zinc_N</a>	<a href="#">AdoHcyase_NAD</a>
<a href="#">AdoMet_MTase</a>	<a href="#">AlaDh_PNT_C</a>	<a href="#">Amino_oxidase</a>	<a href="#">ApbA</a>	<a href="#">Bac_GDH</a>	<a href="#">Bin3</a>
<a href="#">CheR</a>	<a href="#">CMAS</a>	<a href="#">CmcI</a>	<a href="#">CoA_binding</a>	<a href="#">Cons_hypoth95</a>	<a href="#">DAO</a>
<a href="#">DapB_N</a>	<a href="#">DFP</a>	<a href="#">DNA_circ_N</a>	<a href="#">DNA_methylase</a>	<a href="#">DOT1</a>	<a href="#">DREV</a>
<a href="#">DUF1442</a>	<a href="#">DUF166</a>	<a href="#">DUF1776</a>	<a href="#">DUF185</a>	<a href="#">DUF2431</a>	<a href="#">DUF248</a>

**Example structure**

[PDB entry 1kr5](#): Crystal structure of human L-isoaspartyl methyltransferase

[View a different structure:](#) [1kr5](#)

### 1) SUMMARY :

ci da informazioni relative a:

- La entry di InterPro ( data base proteico): permette di visualizzare l'abstract del contenuto della entry e il link correlato all'articolo di PubMed associato
- La famiglia di appartenenza; nel nostro caso è una L-isoaspartato D-aspartato O-metiltransferasi
- La funzione che svolge. Il dominio PCMT catalizza il trasferimento di un metile dalla S-adenosilmetionina a proteine residui di aspartato o isoaspartato che derivano dalla normale deaminazione di residui di asparagina e acido aspartico; questo processo è implicato nella riparazione e degradazione di queste proteine degradate. Può derivare anche dallo studio delle strutture cristallografiche delle proteine appartenenti alla famiglia d'interesse.
- descrizione della struttura ( nel nostro caso assente)
- Literature references ( non presente nel nostro caso), ovvero i titoli e i link correlati degli articoli inerenti al dominio presenti in PubMed

- Clan: indica a che clan/superfamiglia omologa appartiene il nostro dominio e gli altri componenti (PCMT appartiene al clan delle NADP Rossmann fold). Cliccando sul nome relativo al clan si visualizza la pagina correlata ad esso in Pfam.

- Gene Ontology: è l'interlink alla pagina dell'ontologia relativa al dominio

### Gene Ontology

<b>Molecular function</b>	<a href="#">protein-L-isoaspartate (D-aspartate) O-methyltransferase activity (GO:0004719)</a>
<b>Biological process</b>	<a href="#">protein modification process (GO:0006464)</a>

- Internal Database links e External Database links

### Internal database links

<b>SCOOP:</b>	<a href="#">AdoMet_MTase</a>
---------------	------------------------------

### External database links

<b>PANDIT:</b>	<a href="#">PF01135</a>
<b>PROSITE:</b>	<a href="#">PDOC00985</a>
<b>Pseudofam:</b>	<a href="#">PF01135</a>
<b>SCOP:</b>	<a href="#">1dl5</a>
<b>SYSTEMS:</b>	<a href="#">PCMT</a>

PANDIT: Protein and Associated Nucleotide Domains and Inferred Trees; visualizza gli alberi che sono stati creati dalle sequenze di aminoacidi Pfam A grazie a Seed Alignment ( porzioni di sequenze) , da sequenze nucleotidiche o da sequenze aminoacidiche ricavate dalle corrispondenti sequenze nucleotidiche.

PROSITE: pagina relativa nel data base di proteine del Swiss Institute of Bionformatics

Pseudofam: è una banca dati contenenti gli pseudogeni associati alle diverse famiglie proteiche

SCOP : classificazione della proteina

SYSTER: SYSTEMatic Resercing ; classifica in modo gerarchico le sequenze proteiche in famiglie omologhe e superfamiglie

- Struttura 3D di una proteina contenente il dominio con relativa PDB entry; cliccandoci sopra otteniamo una serie di informazioni su come e da chi è stato effettuato l'esperimento per isolare la struttura, in quale specie è stata isolata, gli articoli in PubMed correlati e interlink

correlati ad altri database proteici.

Experiment type:	X-ray, resolution 2.10Å
Deposition date:	08-JAN-02
Authors:	Ryttersgaard, C. , Griffith, S.C. , Sawaya, M.R. , MacLaren, D.C. , Clarke, S. , Yeates, T.O.
Species:	<i>Homo sapiens (Human)</i>
PubMed reference:	<a href="#">11792715</a>

External database links

RCSB PDB	<a href="#">1kr5</a>
PDBe	<a href="#">1kr5</a>
PDBsum	<a href="#">1kr5</a>
CATH	<a href="#">1kr5</a>
SCOP	<a href="#">1kr5</a>
Proteopedia	<a href="#">1kr5</a>

In particolare CATH e SCOP ci danno una classificazione della proteina in base alla struttura, mentre PDB sum permette di individuare ligandi della proteina, i siti di legame e molte altre informazioni .

## 2) DOMAIN ORGANISATION:

mostra una lista delle architetture o topologie contenenti il dominio di interesse; l'output contiene:

Summary

### Domain organisation

Below is a listing of the unique domain organisations or architectures in which this domain is found. [More...](#)

There are 1498 sequences with the following architecture: PCMT  
[PIMT1\\_ARCFU](#) [*Archaeoglobus fulgidus*] Protein-L-isoaspartate O-methyltransferase 1 EC=2.1.1.77 (216 residues)

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

There are 53 sequences with the following architecture: TP\_methylase, PCMT  
[O87694\\_BACME](#) [*Bacillus megaterium*] CbiET protein (398 residues)

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

There are 22 sequences with the following architecture: PCMT, Erythro\_esteras  
[Q92X74\\_RHIME](#) [*Rhizobium meliloti* (*Sinorhizobium meliloti*)] Putative uncharacterized protein (658 residues)

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

There are 14 sequences with the following architecture: PCMT, Mg-por\_mtran\_C  
[Q9RFD7\\_RHOSH](#) [*Rhodobacter sphaeroides* (*Rhodopseudomonas sphaeroides*)] BchM (222 residues)

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

There are 5 sequences with the following architecture: PCMT x 2  
[A7FQA4\\_SCLS1](#) [*Sclerotinia sclerotiorum* (strain ATCC 18683 / 1980 / Ss-1) (*White mold*) (*Whetzelinia sclerotiorum*)] Putative uncharacterized protein (214

- Il numero di sequenze che presentano una particolare architettura a cui appartiene il dominio ( ci posso essere diverse architetture che contengono il dominio, a cui appartengono diverse sequenze)
- Una descrizione testuale di tale architettura
- Il link alla pagina di Pfam contenente le informazioni riguardo la sequenza descritta nel grafico
- Il numero di residui che costituiscono la sequenza
- Il grafico prodotto da Pfam; cliccando sul grafico viene visualizzata la pagina inerente alla famiglia proteica inerente ad ognuno dei domini rappresentati

### 3) ALIGNMENTS:

permette di visualizzare o effettuare il download degli allineamenti di sequenza utilizzati nell'identificazione del dominio

- VIEW OPTIONS: per selezionare il programma con cui visualizzare l'allineamento, tra jalview, visualizzatore HTML, Heat map o Pfam viewer; le sequenze possono essere allineate in modo diverso sia per tutta la loro lunghezza (full), sia considerando porzioni di sequenze (seed), oppure utilizzando le sequenze presenti in NCBI o Metagenomics.

#### View options

<b>Alignment:</b>	<input checked="" type="radio"/> Seed (9)	<input type="radio"/> Full (1614)
	<input type="radio"/> NCBI (8776)	<input type="radio"/> Metagenomics (3824)
<b>Viewer:</b>	jalview <input type="button" value="v"/>	
<input type="button" value="View"/>		

- FORMATTING OPTIONS: oltre a selezionare la modalità di allineamento, ci permette di scegliere il formato con cui verrà visualizzato, l'ordine con il quale vengono clusterizzate le sequenze e come visualizzare i gap. Inoltre possiamo scegliere se visualizzare semplicemente l'output oppure se effettuarne il download; in questo caso può essere salvato in una cartella zip come file in Stockholm format, oppure come formato FASTA contenente tutte le sequenze full length dell'intero allineamento.

#### Formatting options

<b>Alignment:</b>	<input checked="" type="radio"/> Seed (9)	<input type="radio"/> Full (1614)
<b>Format:</b>	Selex <input type="button" value="v"/>	
<b>Order:</b>	<input checked="" type="radio"/> Tree	<input type="radio"/> Alphabetical
<b>Sequence:</b>	<input checked="" type="radio"/> Inserts lower case	<input type="radio"/> All upper case
<b>Gaps:</b>	Gaps as "." or "-" (mixed) <input type="button" value="v"/>	
<b>Download/view:</b>	<input checked="" type="radio"/> Download	<input type="radio"/> View
<input type="button" value="Generate"/>		

- EXTERNAL LINKS : collegamento a Myhits, una raccolta di tools per l'allineamento multiplo. Cliccando su Myhits, dopo aver selezionato un Pfam alignment, questo verrà direttamente inserito nel campo di ricerca di Myhits

#### External links

[MyHits](#) provides a collection of tools to handle multiple sequence alignments. For example, one can refine a seed alignment (sequence addition or removal, re-alignment or manual edition) and then search databases for remote homologs using HMMER2.

<b>Pfam alignments:</b>	<input type="radio"/> Seed (9)	<input type="radio"/> Full (1614)
<input type="button" value="Submit to MyHits"/>		





text search user: GUEST  
 [help](#)

## MSA Hub

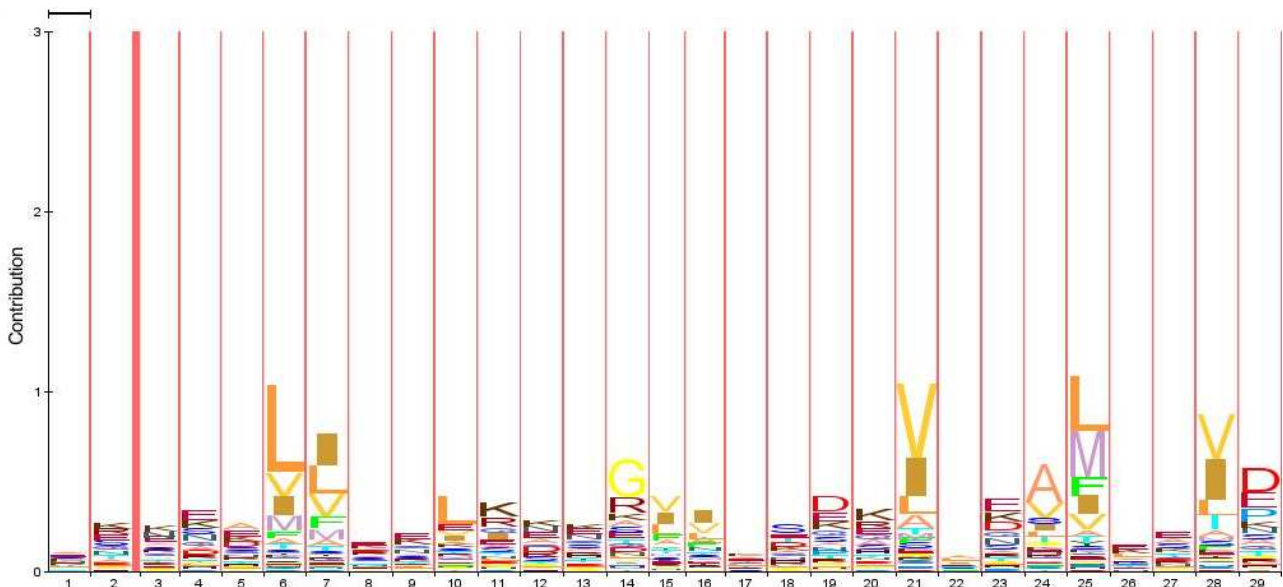
**MSA Input**  
Paste a multiple sequence alignment (MSA) in this text area (supported formats)

examples

```
>PINT_BOVIN/8-221
AS-HSELIHNLREKNGI IKTDKVFVHLATDRSHYAKC-----NPYMDSPQ
NEGAKALDVGSGSGILTACFARHVGPSG-----KVI GIDHIKELVDD SIN
APYDAIHVGAAAPVVPQALIDQLKPGGRLILPVGP-AGGNQMLEQYDKLQ
>PINT_DROME/8-226
AN-NEDLIRQLKDHGVIASDAVAQAMKETDRKHYSR-----NPYMDAPQ
KPGARILDVGS GSYLTACFYRYIKAKGVDADTRIVGIEHQAE LVRRSKA
APYNAIHVGAAAPDTPTELINQLASGGRLIVPVGP-DGGSQYNQOYDKDA
```

#### 4) HMM LOGO:

serve a visualizzare i profili HMM sotto forma grafica in base alla distribuzione degli aminoacidi nelle diverse posizioni dell'allineamento multiplo effettuato per la ricerca in banca dati. La grandezza della lettera corrispondente all'aminoacido è proporzionale alla sua frequenza in quella posizione ( in ascissa : entropia relativa; in ordinata : contributo)



#### 5) TREES:

mostra gli alberi filogenetici Neighbour Joining calcolati grazie a FAStree ( tool di EBI)sulla base del parametro di maximum likelihood ( massima verosomiglianza; con bootstrap basato su 100 ricampionamenti)a partire da allineamenti di sequenze intere ( full) o sottosequenze(seed). Si può effettuare il download dei diversi alberi

**Summary**

**Domain organisation**

**Alignments**

**HMM logo**

**Trees**

**Curation & models**

**Species**

**Interactions**

**Structures**

**Jump to...**

enter ID/acc

**Trees**

This page displays the phylogenetic tree for this family. We use [FastTree](#) to calculate neighbour join trees with a local bootstrap based on 100 resamples next to the tree nodes). FastTree calculates approximately-maximum-likelihood phylogenetic trees from our seed or full alignments.

Seed (9)  Full (1614)  NCBI (8776)  Metagenomics (3824)

Note: You can also download the data files for the [seed](#), [full](#), [NCBI](#) or [metagenomics](#) trees.

6)

**CURATION and MODELS:** informazioni accurate sul metodo utilizzato per la ricerca in banca dati, su come è stato ottenuto il profilo HMM e quali parametri sono stati usati. Si può effettuare il download del profilo HMM corrispondente alla famiglia che stiamo studiando.

### Curation

<b>Seed source:</b>	Prosites
<b>Previous IDs:</b>	none
<b>Type:</b>	Family
<b>Author:</b>	Finn RD, Bateman A
<b>Number in seed:</b>	9
<b>Number in full:</b>	1614
<b>Average length of the domain:</b>	191.70 aa
<b>Average identity of full alignment:</b>	29 %
<b>Average coverage of the sequence by the domain:</b>	73.42 %

### HMM information

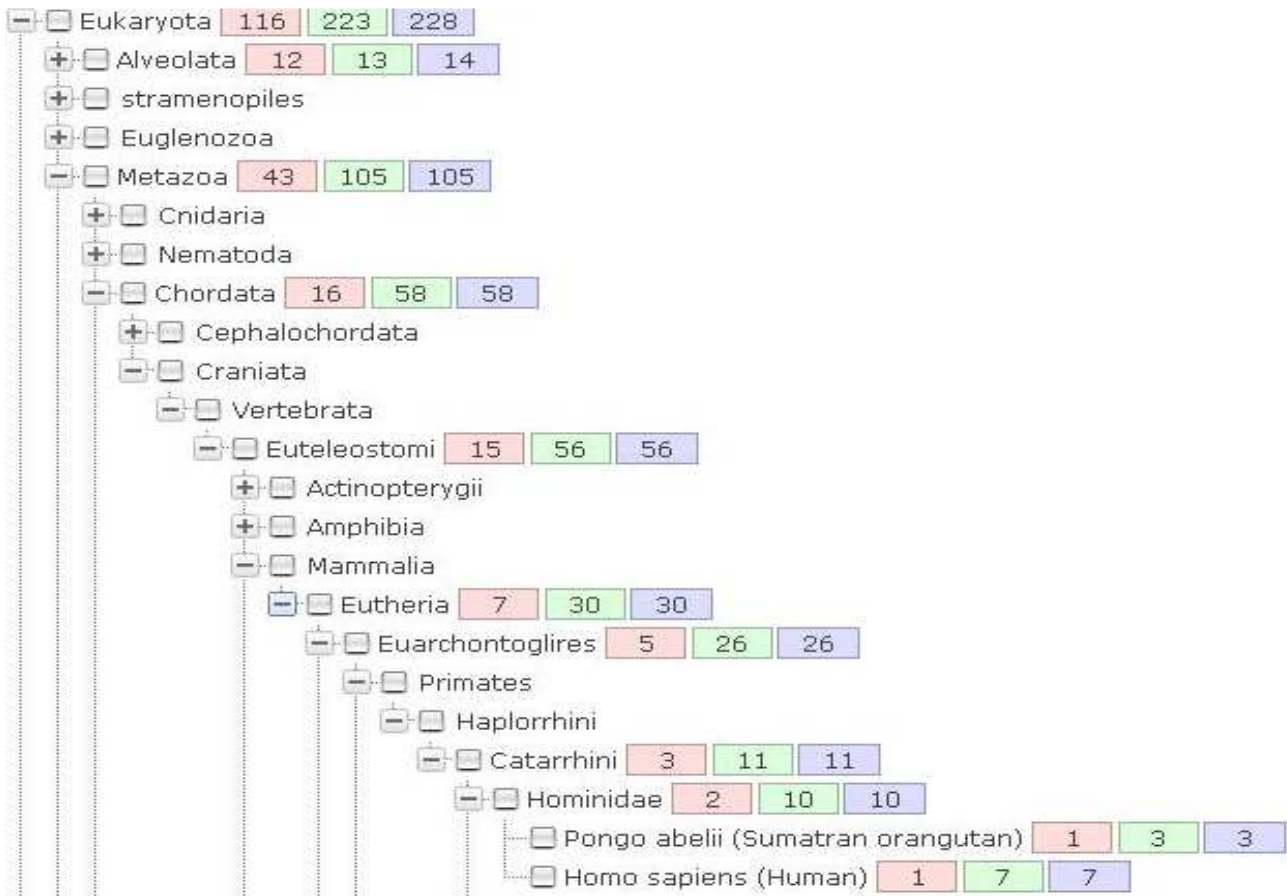
<b>HMM build commands:</b>	<i>build method:</i> hmmbuild -o /dev/null HMM SEED <i>search method:</i> hmmsearch -Z 9421015 -E 1000 HMM pfamseq		
<b>Model details:</b>	<b>Parameter</b>	<b>Sequence</b>	<b>Domain</b>
	<b>Gathering cut-off</b>	20.3	20.3
	<b>Trusted cut-off</b>	20.3	20.3
	<b>Noise cut-off</b>	20.2	20.2
<b>Model length:</b>	210		
<b>Family (HMM) version:</b>	12		
<b>Download:</b>	<a href="#">download</a> the raw HMM for this family		

### 7) SPECIES :

contiene gli alberi delle specie , secondo una classificazione tassonomica di NCBI, in cui è presente il dominio. Ad ogni unità tassonomica sono associati 3 box di colorazione diversa:

- Indaco: indica quante volte è stato trovato il dominio durante l'allineamento completo di tutte le sequenze
- Verde: numero di sequenze uniche in cui è stato trovato il dominio
- Rosa: il numero di sequenze, presenti in una stessa unità tassonomica, in cui è presente il dominio

Si può scegliere di espandere l'albero a piacere, visualizzando gli organismi di interesse cliccando sulle diverse unità tassonomiche (esempio : PCMT nell'uomo)



## 7) INTERACTIONS:

individua gli interattori della nostra proteina sulla base della struttura 3D delle proteine presenti in Pfam utilizzando iPfam.

<ul style="list-style-type: none"> <li>Summary</li> <li>Domain organisation</li> <li>Alignments</li> <li>HMM logo</li> <li>Trees</li> <li>Curation &amp; models</li> <li>Species</li> <li><b>Interactions</b></li> </ul>	<h3>Interactions</h3> <p>There is <b>1</b> interaction for this family. <a href="#">More...</a></p> <p><a href="#">PCMT</a></p>
--	---

## 7) STRUCTURES:

Viene fornita una tabella in cui vengono elencate le entries delle strutture 3D in cui è presente il nostro dominio. Queste strutture sono state ottenute utilizzando in modo coordinato PDB, Uniprot e Pfam per ricercare i domini presenti in Pfam nelle sequenze Uniprot e nelle strutture tridimensionali di PDB. La struttura del dominio può essere visualizzata con 3 programmi diversi: Jmol, AstexViewer e SPICE.

UniProt entry	UniProt residues	PDB ID	PDB chain ID	PDB residues	View
<a href="#">PIMT_DROME</a>	226 - 226	<a href="#">1r18</a>	A	0 - 0	<a href="#">Jmol</a> <a href="#">AstexViewer</a> <a href="#">SPICE</a>
<a href="#">PIMT_HUMAN</a>	8 - 221	<a href="#">1i1n</a>	A	7 - 220	<a href="#">Jmol</a> <a href="#">AstexViewer</a> <a href="#">SPICE</a>
		<a href="#">1kr5</a>	A	7 - 220	<a href="#">Jmol</a> <a href="#">AstexViewer</a> <a href="#">SPICE</a>
<a href="#">PIMT_METJA</a>	5 - 212	<a href="#">2yxe</a>	A	5 - 212	<a href="#">Jmol</a> <a href="#">AstexViewer</a> <a href="#">SPICE</a>
			B	5 - 212	<a href="#">Jmol</a> <a href="#">AstexViewer</a> <a href="#">SPICE</a>

Cliccando su ID Uniprot o PDB si può visualizzare la proteina, appartenente ad una determinata specie, contenente il dominio preso in esame. Per esempio possiamo visualizzare PIMT HUMAN

**Summary**

Features

**Sequence**

Interactions

**Structures**

TreeFam

**Jump to...**

enter ID/acc

**Summary**

**PIMT\_HUMAN**


This is the summary of UniProt entry [PIMT\\_HUMAN](#) (P22061).

<b>Description:</b>	Protein-L-isoaspartate(D-aspartate) O-methyltransferase EC=2.1.1.77
<b>Source organism:</b>	<a href="#">Homo sapiens (Human)</a> (NCBI taxonomy ID <a href="#">9606</a> ) <a href="#">View Pfam proteome data.</a>
<b>Length:</b>	227 amino acids

**Please note:** when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

**Pfam domains**

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains.



Source	Domain	Start	End
<a href="#">Pfam A</a>	PCMT	8	221

In questo caso stiamo visualizzando la pagina di Pfam relativa alla proteina contenente PCMT di Homo sapiens. Da qui possiamo ricavare una serie di altre informazioni, come la sequenza della proteina, la struttura, e l'albero filogenetico costruito se sequenze paraloghe ed ortologhe.

**Sequence:**

```

1  MDAQSGGASH SELIHRLEAQN GIINKTQVFE WMLATDRSHY AKCMPYNDSP 50
51  QSIGFQATIS APHMEAYALE LLFDQLHEGA KALDVGSGG ILTACFARMV 100
101 GCTGKVI GIDHIKELVDDSI INNVKNDIPTLLS SGRQQLAVV GDGRMGYAE 150
151 APYDAIHVGA AAPVVPQALI DQKNPGLLI LPVGPAGGNQ MLEQSDNLQD 200
201 GSIINKPLAG VIYVPLTDKEK QQSEAK 227

```

[Hide the unformatted sequence.](#)

```

MDAQSGGASHSELIHRLEAQN GIINKTQVFEWMLATDRSHYAKCMPYNDSP
QSIGFQATISAPHMEAYALELLFDQLHEGAKALDVGSGGILTACFARMV
GCTGKVI GIDHIKELVDDSI INNVKNDIPTLLS SGRQQLAVVGDGRMGYAE
APYDAIHVGA AAPVVPQALI DQKNPGLLI LPVGPAGGNQMLEQSDNLQD
GSIINKPLAGVIYVPLTDKEKQQSEAK

```

**Checksums:** CRC64: 987D63EA48422CA8  
MD5: 6485bdf0fcd1ceee4f5a914d62d261e9

# TreeFam

Below is a phylogenetic tree of animal genes, with ortholog and paralog assignments, from [TreeFam](#).

