PFAM domains

Domains – modular building blocks

Domains reappear in different proteins.

Are often first identified from protein structure (common folds)

Homologous domains frequently have related functions.

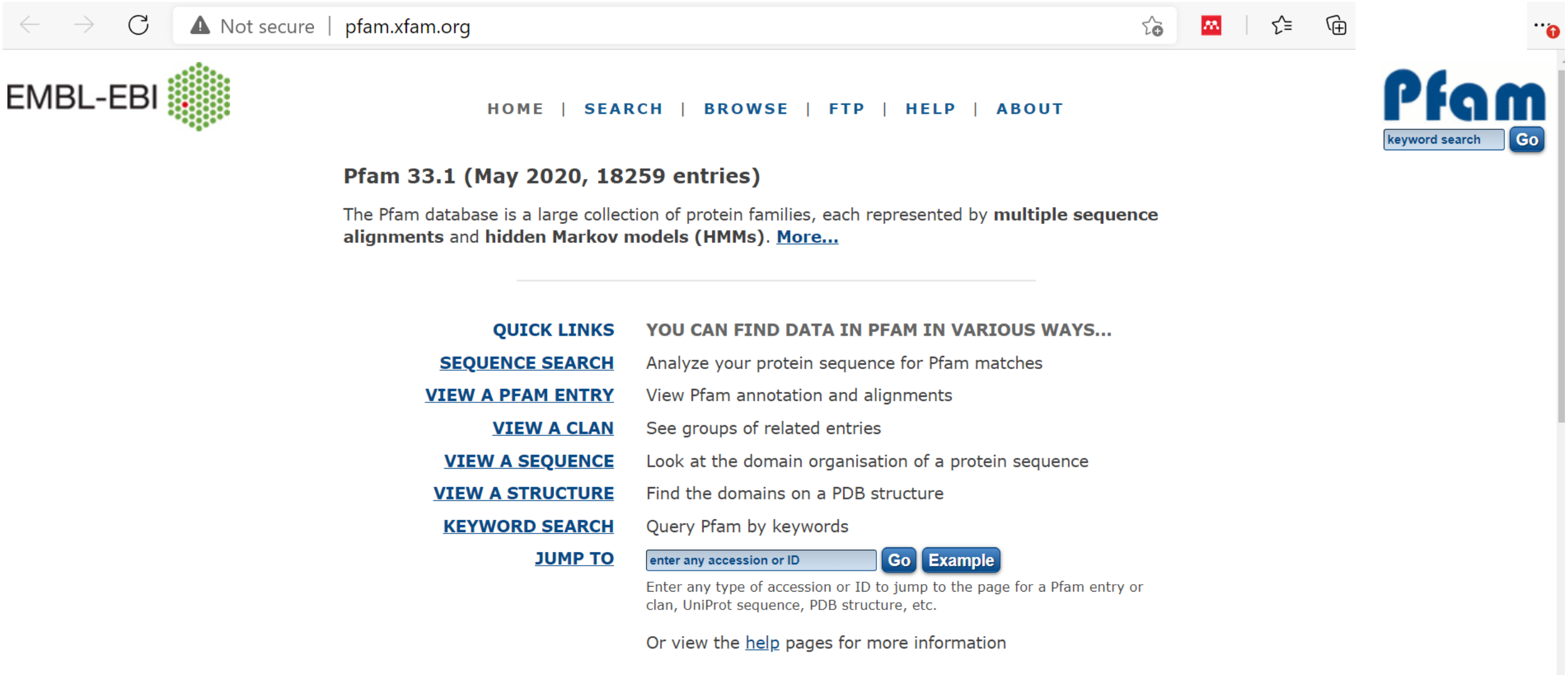
New domain combinations can yield new functions.

Domains – modular building blocks

Have conserved sequence motifs (or patterns of sequence conservation)

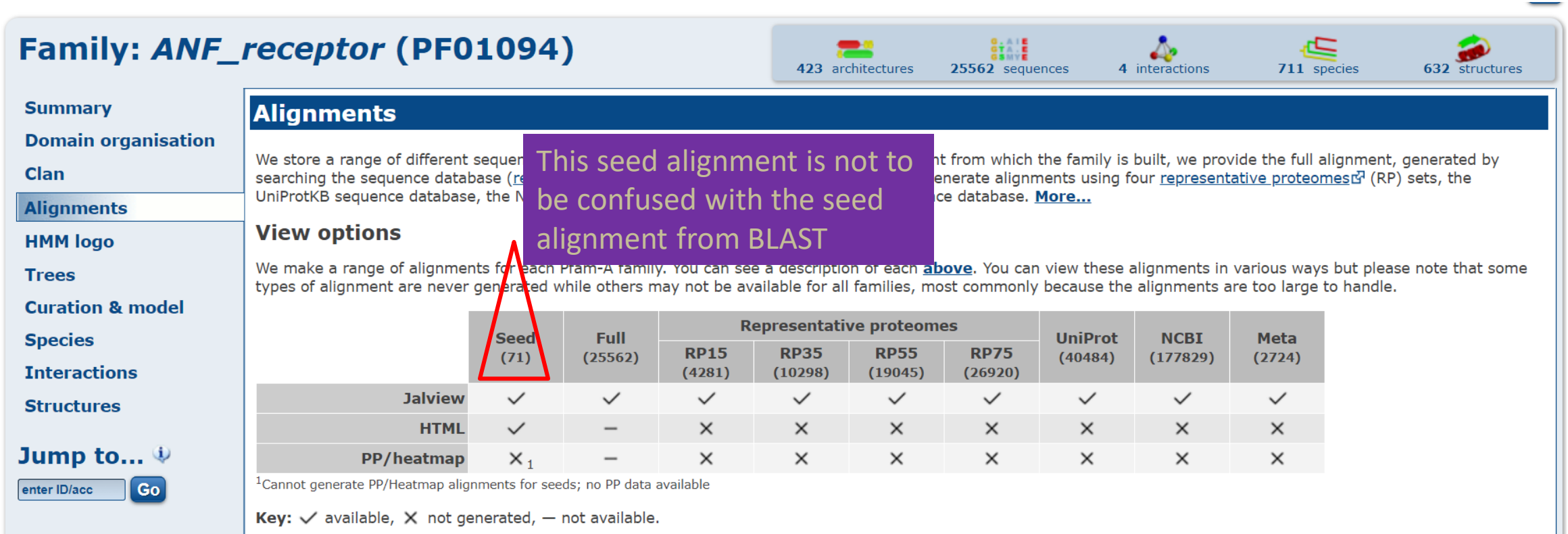
PFAM differentiates between DOMAINS as structural units and FAMILIES as collection of related protein regions. Both are modular building blocks and identifying domains and/or families help us understand a protein's function.

Pfam is the main protein domain database



Pfam uses Hidden Markov models (HMMs) to find domains in protein sequences

- A HMM is made based on a SEED alignment.
- The **SEED alignment** is a small subset of sequences (from a larger protein family alignment) that are used to make a representative position specific PROFILE HMM.



Family: *ANF_receptor* (PF01094)

423 architectures | 25562 sequences | 4 interactions | 711 species | 632 structures

Summary
Domain organisation
Clan
Alignments
HMM logo
Trees
Curation & model
Species
Interactions
Structures

Jump to... enter ID/acc Go

Alignments

This seed alignment is not to be confused with the seed alignment from BLAST

We store a range of different sequen…nt from which the family is built, we provide the full alignment, generated by searching the sequence database (re…enerate alignments using four representative proteomes (RP) sets, the UniProtKB sequence database, the N…ce 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 please note that some types of alignment are never generated while others may not be available for all families, most commonly because the alignments are too large to handle.

	Seed (71)	Full (25562)	Representative proteomes				UniProt (40484)	NCBI (177829)	Meta (2724)
			RP15 (4281)	RP35 (10298)	RP55 (19045)	RP75 (26920)			
Jalview	✓	✓	✓	✓	✓	✓	✓	✓	✓
HTML	✓	–	×	×	×	×	×	×	×
PP/heatmap	×[1]	–	×	×	×	×	×	×	×

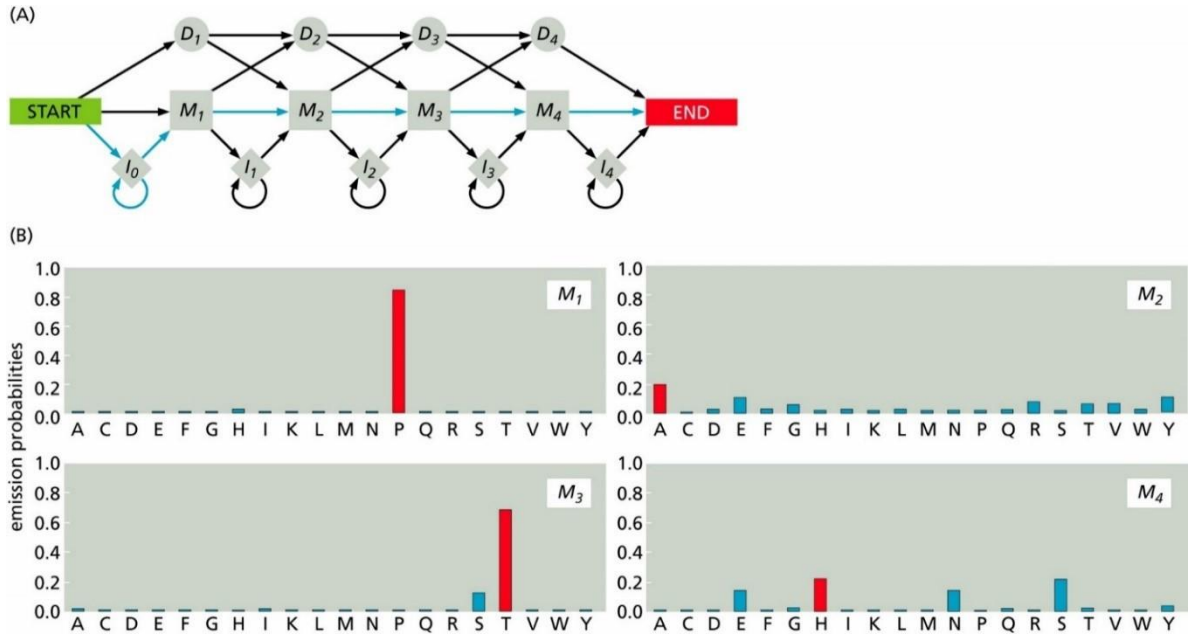
[1]Cannot generate PP/Heatmap alignments for seeds; no PP data available

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

- PFAM Seed alignment:

  The representative alignment for a family/domain. Used to make a Hidden Markov Model (HMM). The HMM is used to find other sequences that belong to this group.

Profile HMM



M – match (the amino acid residue in the sequence matches the HMM)

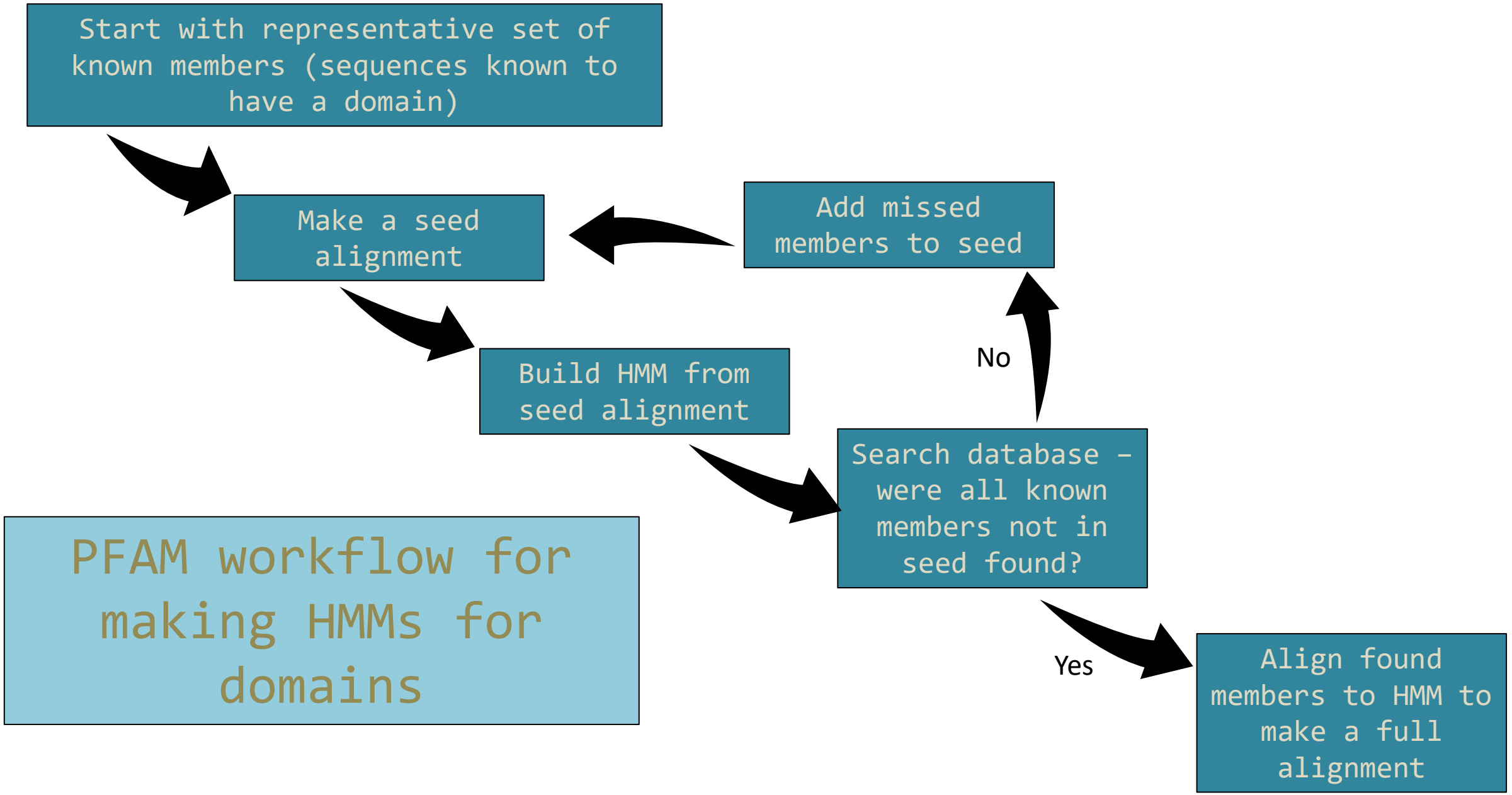
I – insert

D – delete

Each event has an output (emission) probability (even I and D).

- PFAM Full alignments:
  The alignment generated from searching a sequence database using the HMM.

From Understanding Bioinformatics. Zvelebli and Baum

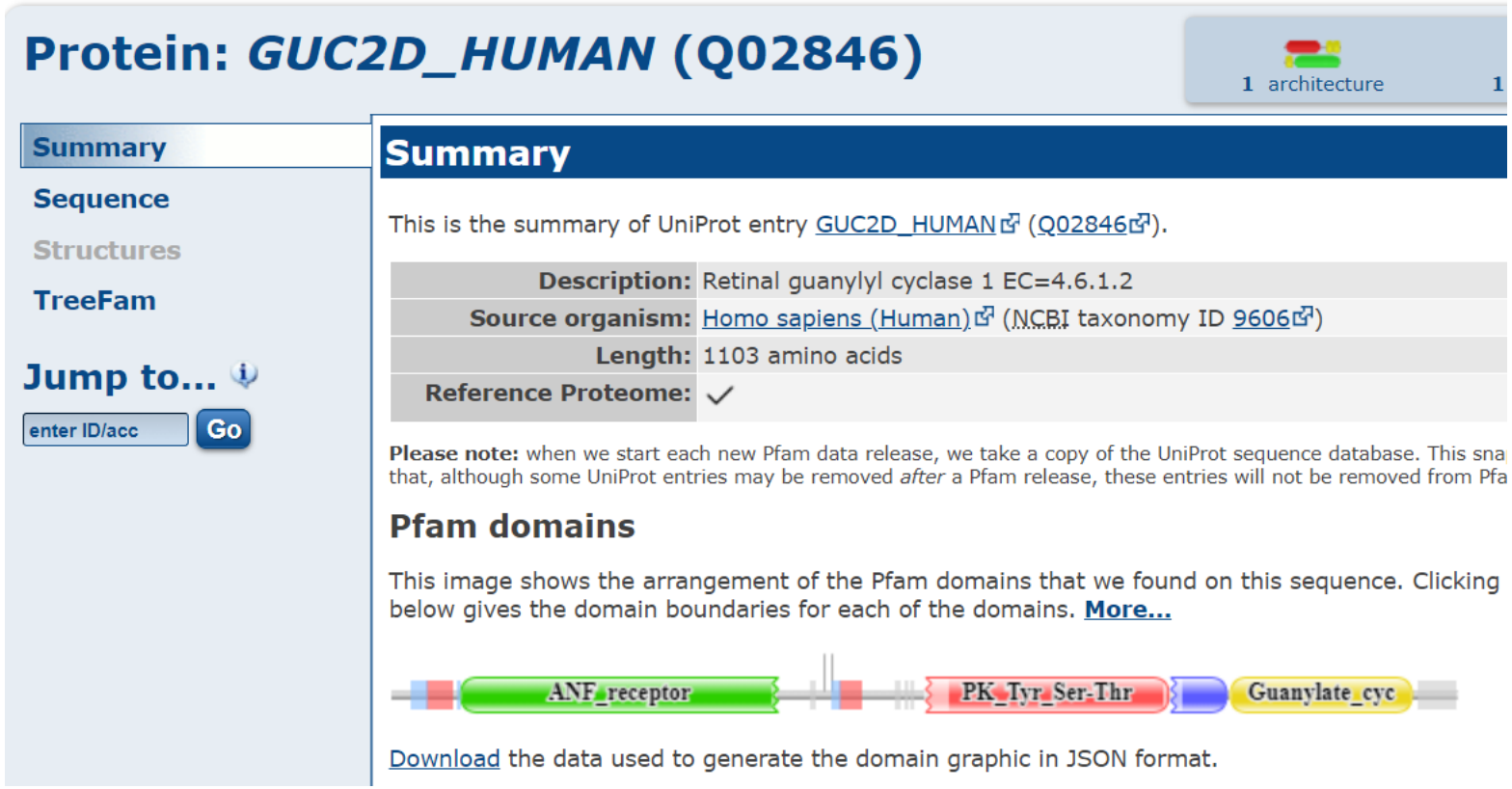


PFAM workflow for making HMMs for domains

Based on PROTEINS: Structure, Function, and Genetics, 28:405–420 (1997)

Pfam example

Based on typing in the UniProt code: Q02846



Source	Domain	Start	End
low_complexity	n/a	22	49
transmembrane	n/a	38	63
low_complexity	n/a	69	87
Pfam	ANF_receptor	72	400
low_complexity	n/a	135	146
low_complexity	n/a	156	173
low_complexity	n/a	357	370
disorder	n/a	436	439
low_complexity	n/a	457	483
transmembrane	n/a	465	487
disorder	n/a	523	527
disorder	n/a	534	535
disorder	n/a	539	540
Pfam	PK_Tyr_Ser-Thr	554	805
disorder	n/a	693	695
Pfam	HNOBA	808	865
coiled_coil	n/a	817	844
low_complexity	n/a	830	844
disorder	n/a	838	840
disorder	n/a	847	849
disorder	n/a	863	864
Pfam	Guanylate_cyc	871	1058
disorder	n/a	1065	1103

Pfam example

Based on performing a sequence search in PFAM with the FASTA sequence for Q02846.

Sequence search results

Show the detailed description of this results page.

We found **4** Pfam-A matches to your search sequence (**all** significant)

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
				Start	End	Start	End	From	To			
Guanylate_cyc	Adenylate and Guanylate cyclase catalyti ...	Domain	CL0276	871	1058	873	1057	**3**	**182**	183	210.2	1.8e-62
ANF_receptor	Receptor family ligand binding region	Domain	CL0144	72	400	73	394	**2**	**333**	354	145.7	1.8e-42
Pkinase_Tyr	Protein tyrosine kinase	Domain	CL0016	555	805	578	800	**49**	**254**	259	93.8	1.0e-26
HNOBA	Heme NO binding associated	Domain	n/a	808	865	820	865	**180**	221	221	25.4	8.3e-06

Sequence positions

Envelope shows region on the sequence where the match has been probabilistically determined to lie (whereas ***Alignment*** *coordinates* show the region over which PFAM is confident that the alignment of the sequence to the profile HMM is correct).

HMM is the part of the HMM that is matched from the sequence.

Pfam example

Based on performing a sequence search in PFAM with the FASTA sequence for Q02846.

Only the 4th domain is found to be close to complete (complete means from 1 to whatever the HMM length is).
For domains 1-3, the entire HMM is not found and these are shown whith jagged edges in the domain cartoon. The cartoon is jagged on the side that is not complete.

Sequence search results

Show the detailed description of this results page.

We found **4** Pfam-A matches to your search sequence (**all** significant)

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
				Start	End	Start	End	From	To			
Guanylate_cyc	Adenylate and Guanylate cyclase catalyti ...	Domain	CL0276	871	1058	873	1057	**3**	**182**	183	210.2	1.8e-62
ANF_receptor	Receptor family ligand binding region	Domain	CL0144	72	400	73	394	**2**	**333**	354	145.7	1.8e-42
Pkinase_Tyr	Protein tyrosine kinase	Domain	CL0016	555	805	578	800	**49**	**254**	259	93.8	1.0e-26
HNOBA	Heme NO binding associated	Domain	n/a	808	865	820	865	**180**	221	221	25.4	8.3e-06

Sequence positions

Envelope shows region on the sequence where the match has been probabilistically determined to lie (whereas ***Alignment*** *coordinates* show the region over which PFAM is confident that the alignment of the sequence to the profile HMM is correct).
HMM is the part of the HMM that is matched from the sequence.

Pfam example

>sp|Q02846|GUC2D_HUMAN Retinal guanylyl cyclase 1

Sequence search results

Show the detailed description of this results page.

We found **4** Pfam-A matches to your search sequence (**all** significant)

Show the search options and sequence that you submitted.

Return to the search form to look for Pfam domains on a new sequence.

Only the 4th domain is found to be close to complete (complete means from 1 to whatever the HMM length is).
For domains 1-3, the entire HMM is not found and these are shown whith jagged edges in the domain cartoon. The cartoon is jagged on the side that is not complete.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
ANF_receptor	Receptor family ligand binding region	Family	CL0144	72	400	73	394	**2**	**331**	352	146.0	1.4e-42	n/a	Show
Pkinase_Tyr	Protein tyrosine kinase	Domain	CL0016	554	805	577	800	**48**	**255**	260	94.1	7.5e-27	n/a	Show
HNOBA	Heme NO binding associated	Domain	n/a	808	865	820	865	**180**	221	221	25.4	7.9e-06	n/a	Show
Guanylate_cyc	Adenylate and Guanylate cyclase catalyti ...	Domain	CL0276	871	1058	873	1057	**3**	**182**	183	210.2	1.7e-62	n/a	Show

Comments or questions on the site? Send a mail to **pfam-help@ebi.ac.uk**.

European Molecular Biology Laboratory

Sequence positions

Envelope shows region on the sequence where the match has been probabilistically determined to lie (whereas ***Alignment*** *coordinates* show the region over which PFAM is confident that the alignment of the sequence to the profile HMM is correct).

HMM is the part of the HMM that is matched from the sequence.

Pfam example

>sp|Q02846|GUC2D_HUMAN Retinal guanylyl cyclase 1
http://pfam.xfam.org/family/PF00211.19

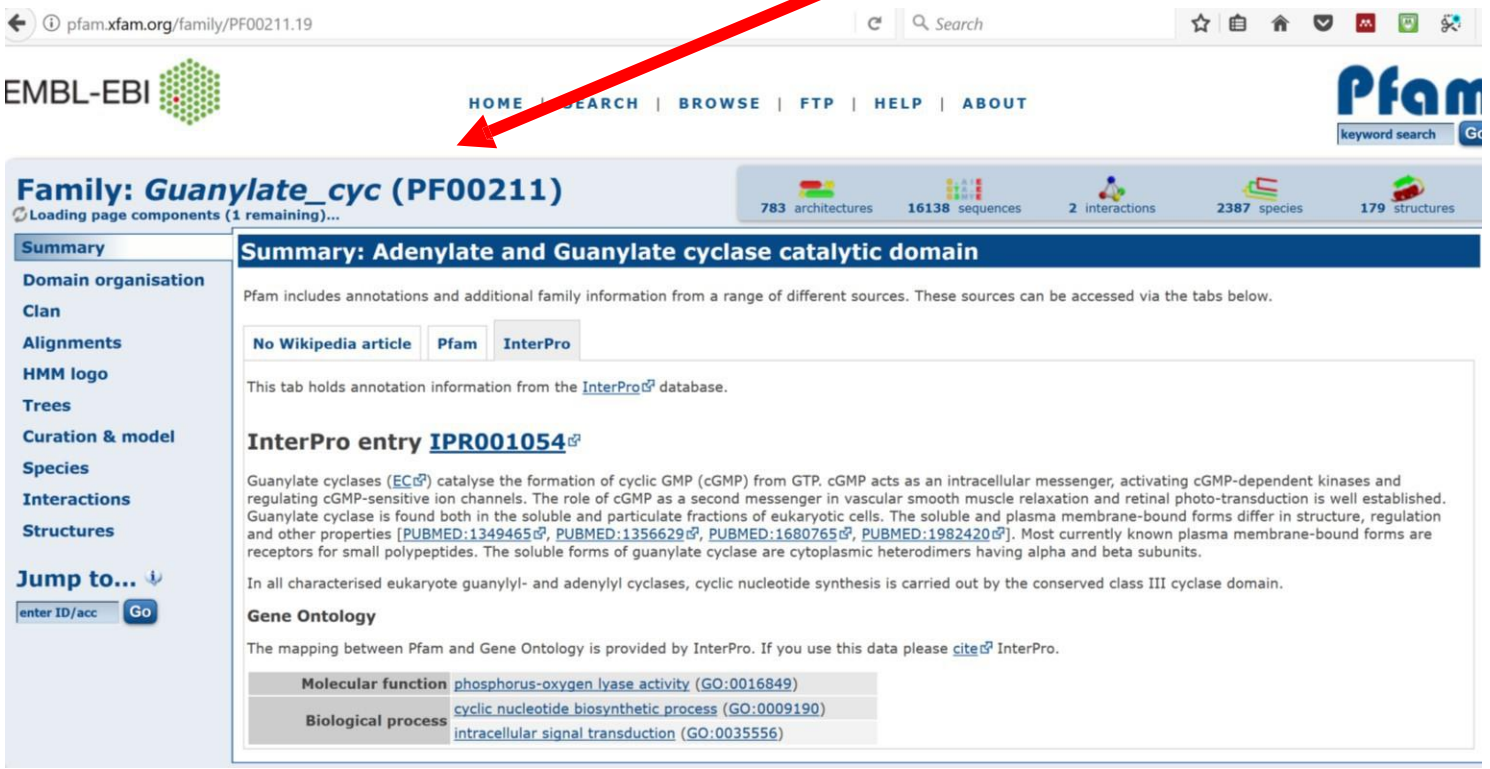
Sequence search results

Show the detailed description of this results page.

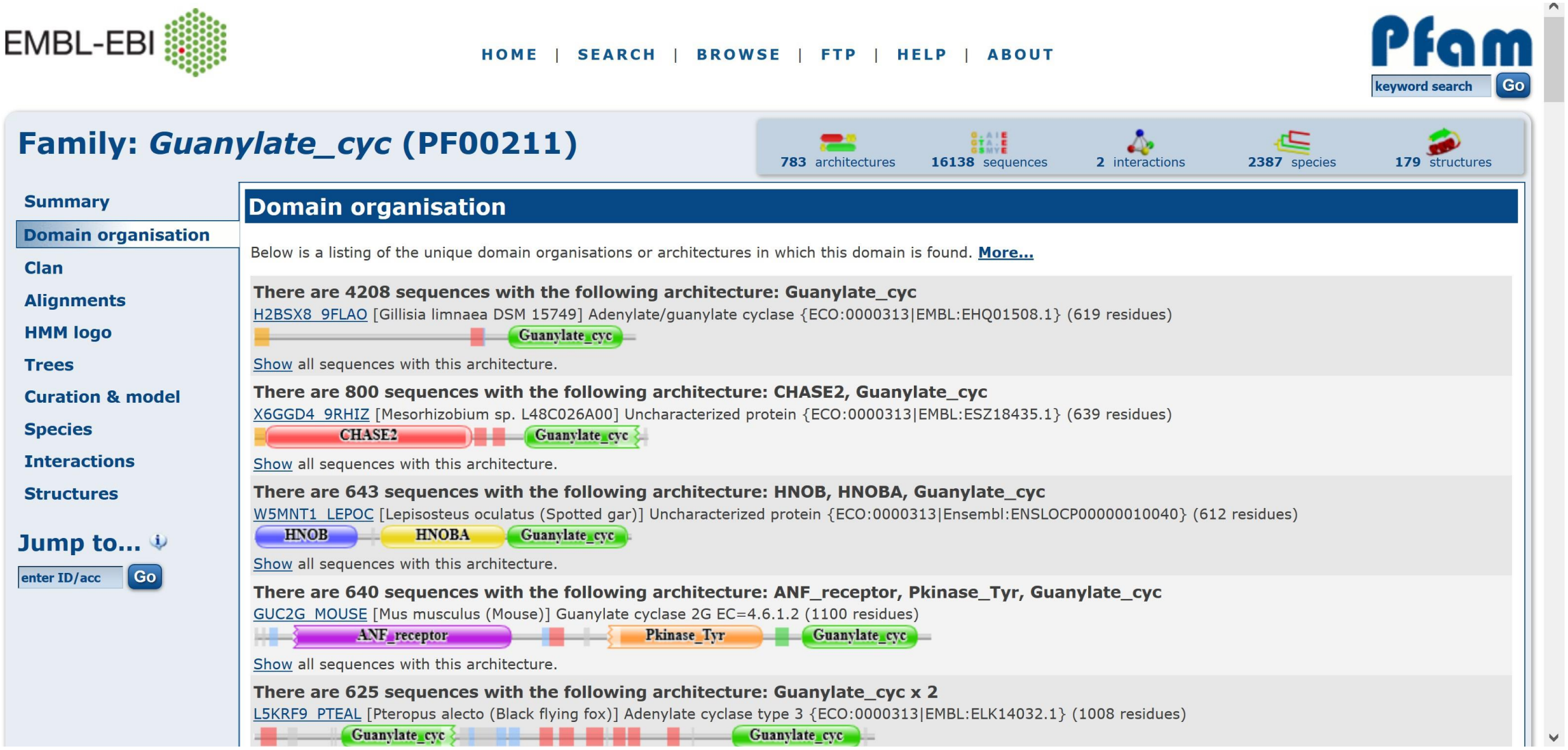
We found **4** Pfam-A matches to your search sequence (**all** significant)

Show the search options and sequence that you submitted.

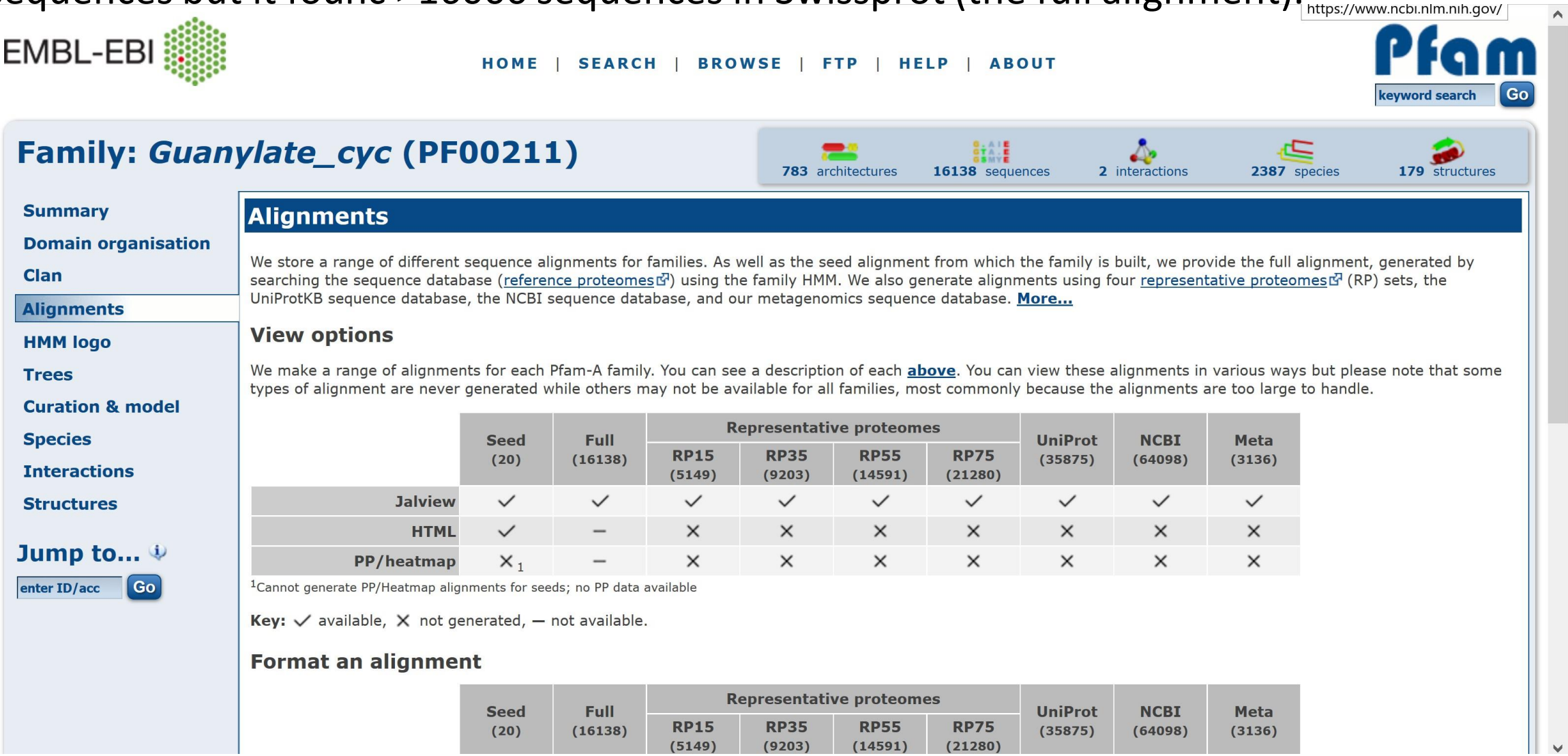
Return to the search form to look for Pfam domains on a new sequence.



Domain organization: How many different domain architecture are this domain found in? Here the first 5 of 783 domain architectures (or domain organizations) are shown.



Alignments: Shows the seed alignment and full alignment plus other alignments based on different criteria. Here the seed alignment used to make the HMM contained 20 sequences but it founc >16000 sequences in Swissprot (the full alignment).



https://www.ncbi.nlm.nih.gov/

EMBL-EBI

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam

keyword search Go

Family: *Guanylate_cyc* (PF00211)

783 architectures | 16138 sequences | 2 interactions | 2387 species | 179 structures

- 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, generated by searching the sequence database (reference proteomes) using the family HMM. We also generate alignments using four representative proteomes (RP) sets, the UniProtKB sequence database, the NCBI sequence database, and our metagenomics 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 but please note that some types of alignment are never generated while others may not be available for all families, most commonly because the alignments are too large to handle.

	Seed (20)	Full (16138)	Representative proteomes				UniProt (35875)	NCBI (64098)	Meta (3136)
			RP15 (5149)	RP35 (9203)	RP55 (14591)	RP75 (21280)			
Jalview	✓	✓	✓	✓	✓	✓	✓	✓	✓
HTML	✓	–	×	×	×	×	×	×	×
PP/heatmap	×[1]	–	×	×	×	×	×	×	×

[1]Cannot generate PP/Heatmap alignments for seeds; no PP data available

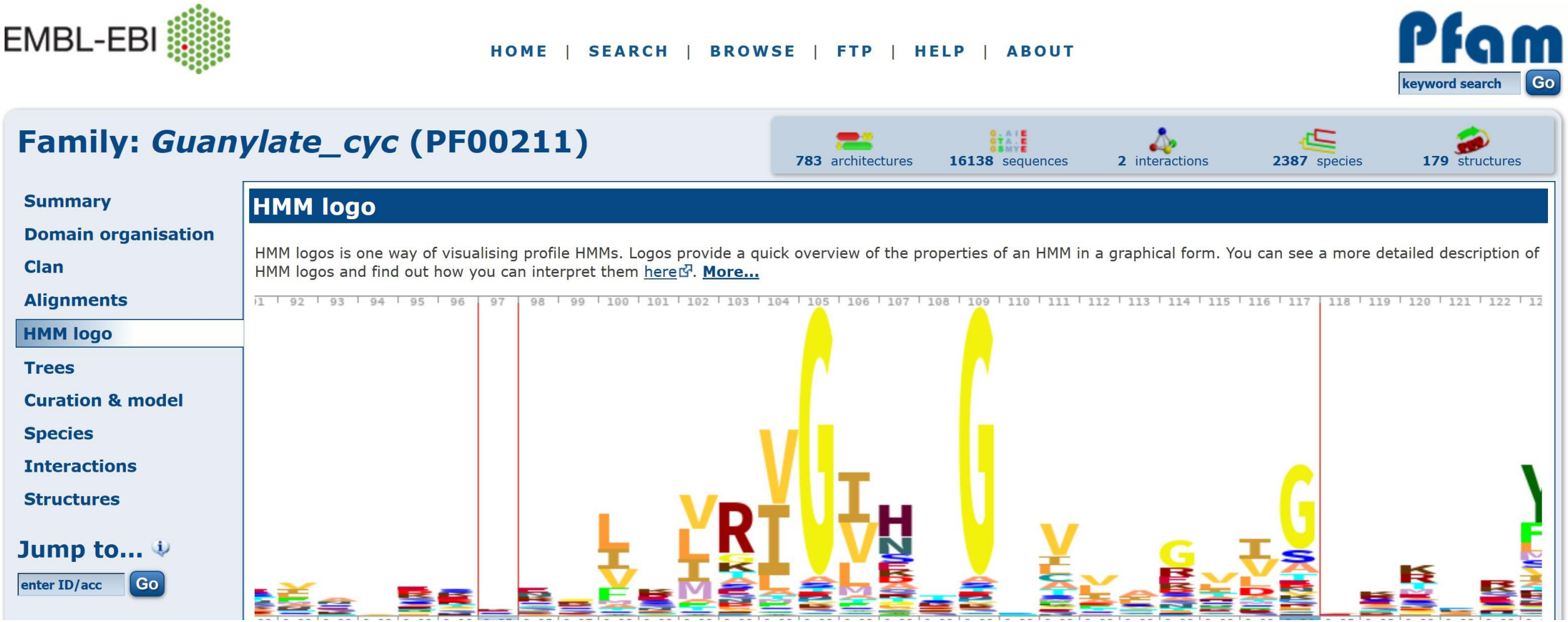
Key: ✓ available, × not generated, – not available.

Format an alignment

	Seed (20)	Full (16138)	Representative proteomes				UniProt (35875)	NCBI (64098)	Meta (3136)
			RP15 (5149)	RP35 (9203)	RP55 (14591)	RP75 (21280)			

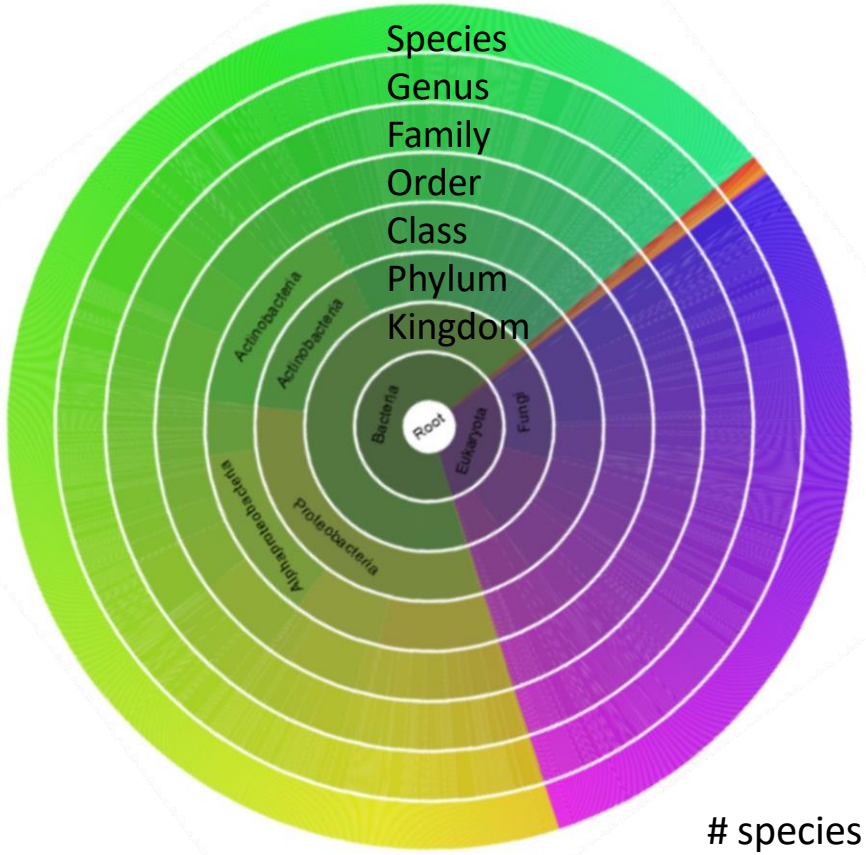
Clan: Related domains but with specific seed alignments

HMM logo: Stacks are scaled based on how much the site varies (the sequence entropy). Letters are scaled based on letter frequency.

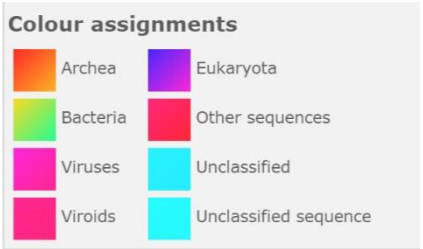


Species: The SUNBURST TREE – Overview over which species the domain has been found in and how many sequences per species

sequences



species



Interactions: Known interacting partners

Any questions?

KEGG adds CONTEXT

KEGG Home
- Release notes
- Current statistics

KEGG Database
- KEGG overview
- Searching KEGG
- KEGG mapping
- Color codes

KEGG Objects
- Pathway maps
- Brite hierarchies
- KEGG DB links

KEGG Software
- KEGG API
- KGML

KEGG FTP
- Subscription
- Background info

GenomeNet

DBGET/LinkDB

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.
See Release notes (January 1, 2021) for new and updated features.

New article KEGG: integrating viruses and cellular organisms

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents [Update notes | Release history]

Data-oriented entry points

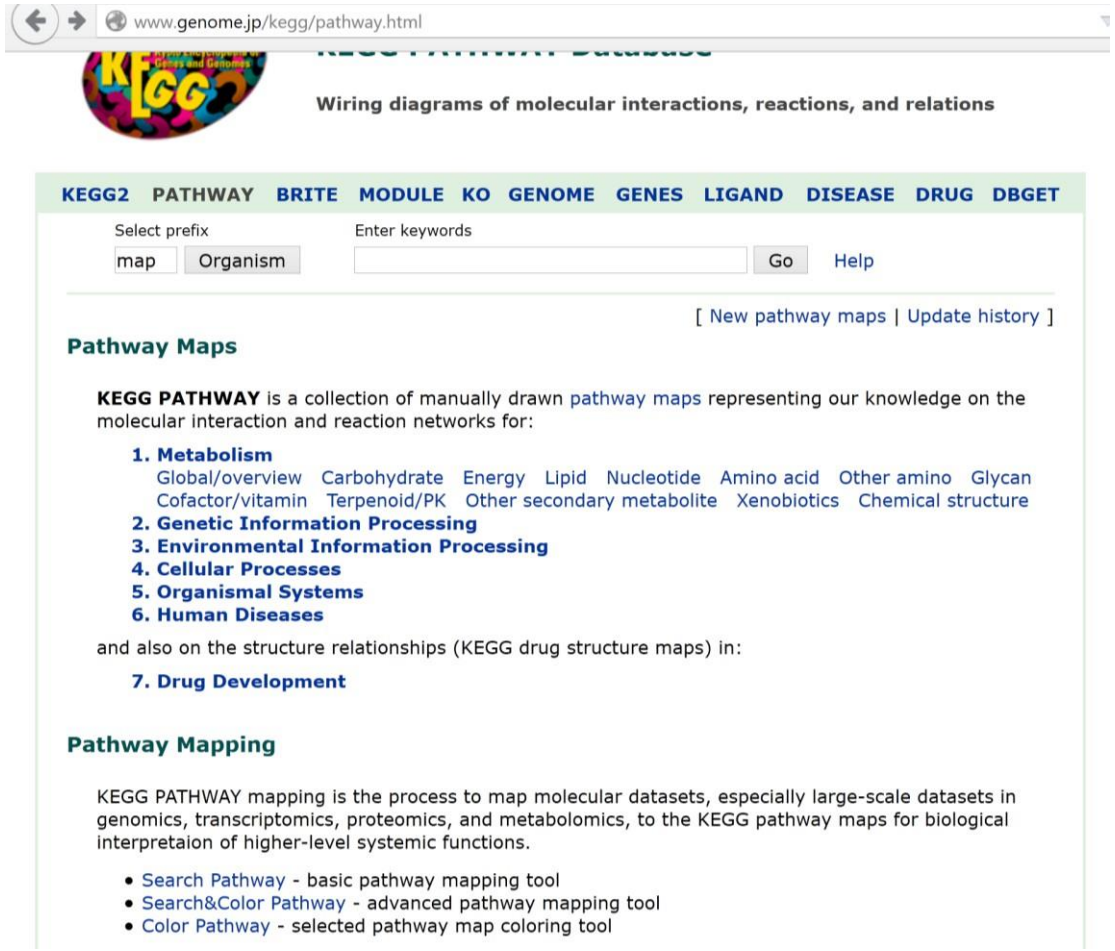
KEGG PATHWAY	KEGG pathway maps
KEGG BRITE	BRITE hierarchies and tables
KEGG MODULE	KEGG modules
KEGG ORTHOLOGY	KO functional orthologs [Annotation]
KEGG GENOME	Genomes [Pathogen | Virus | Plant]
KEGG GENES	Genes and proteins [SeqData]
KEGG COMPOUND	Small molecules
KEGG GLYCAN	Glycans
KEGG REACTION	Biochemical reactions [RModule]

Classification

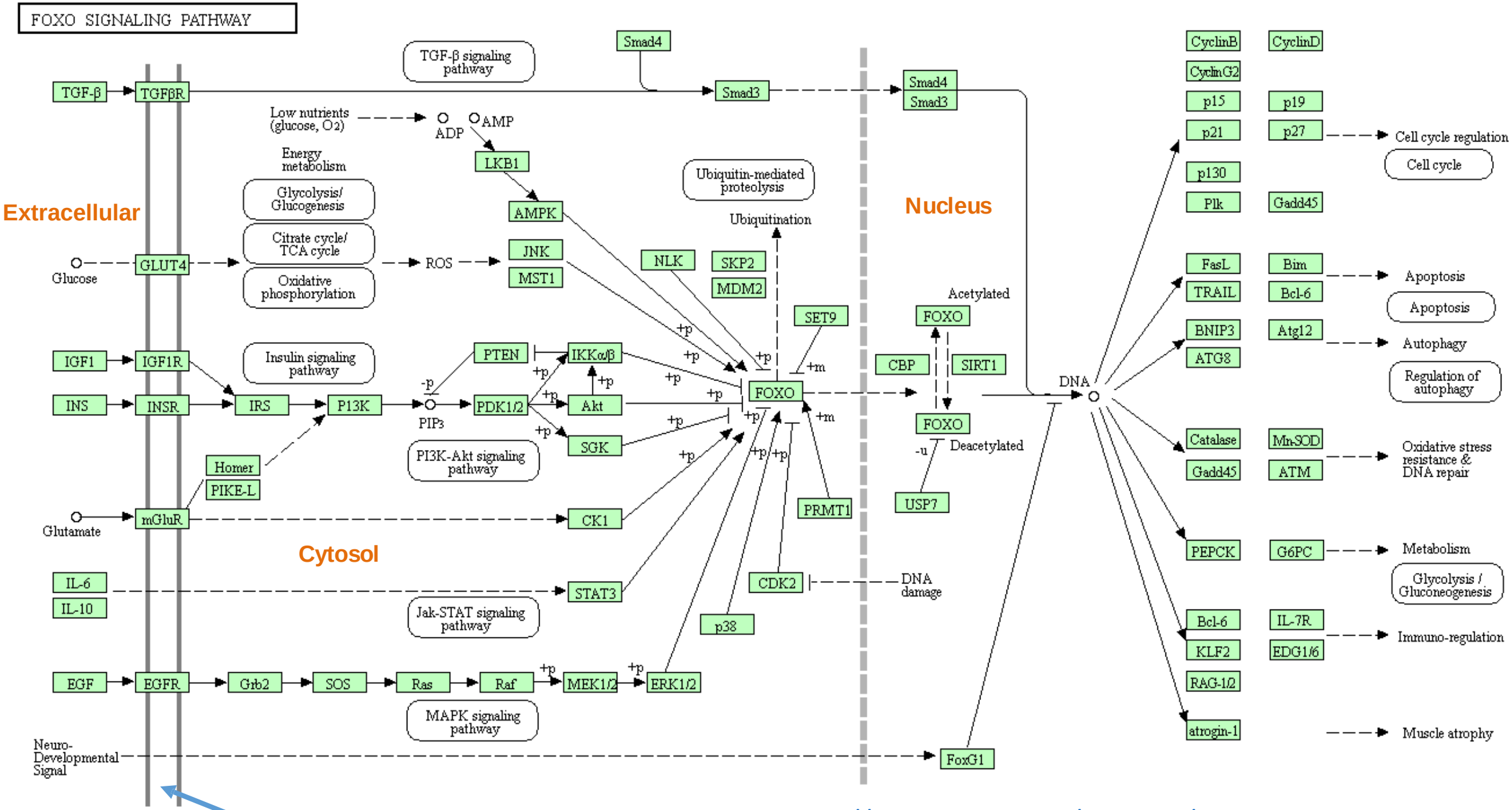
Pathway
Brite
Brite table
Module
KO (Function)
Organism
Virus
Compound
Network
Disease (ICD)

PATHWAY CONTEXT

- In which pathway does the protein and its homologs act?
- Are they enzymes? Where is the substrate coming from and where is the product going?
- Are the involved in signaling? Which proteins do they signal to/from?
- What species have a certain pathways?
- What genes/proteins of a pathway does a species have?

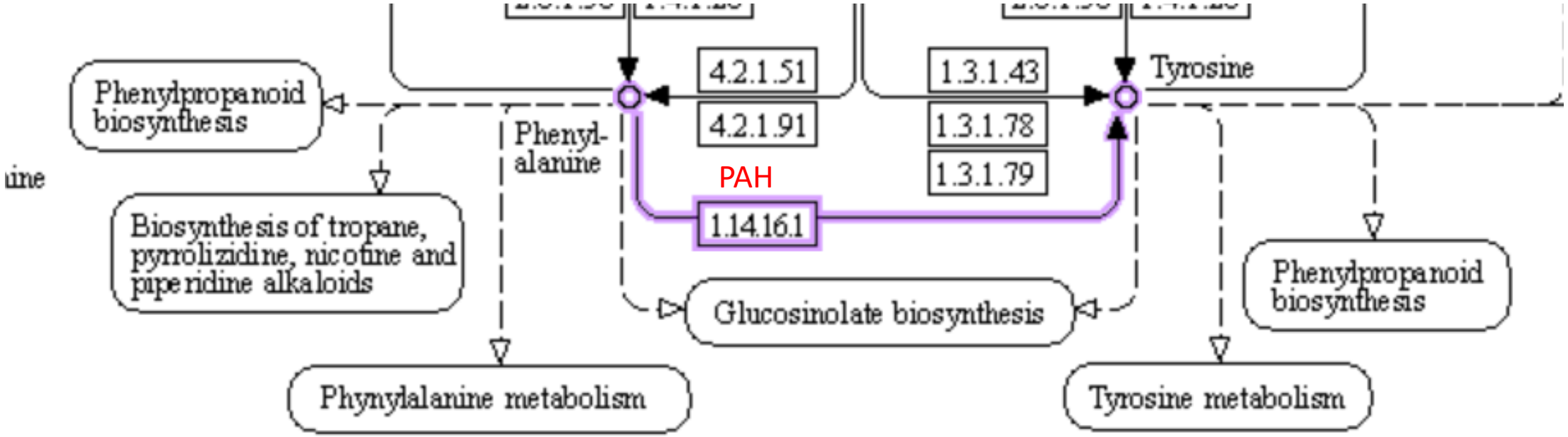


KEGG – Kyoto Encyclopedia of Genes and Genomes http://www.kegg.jp/

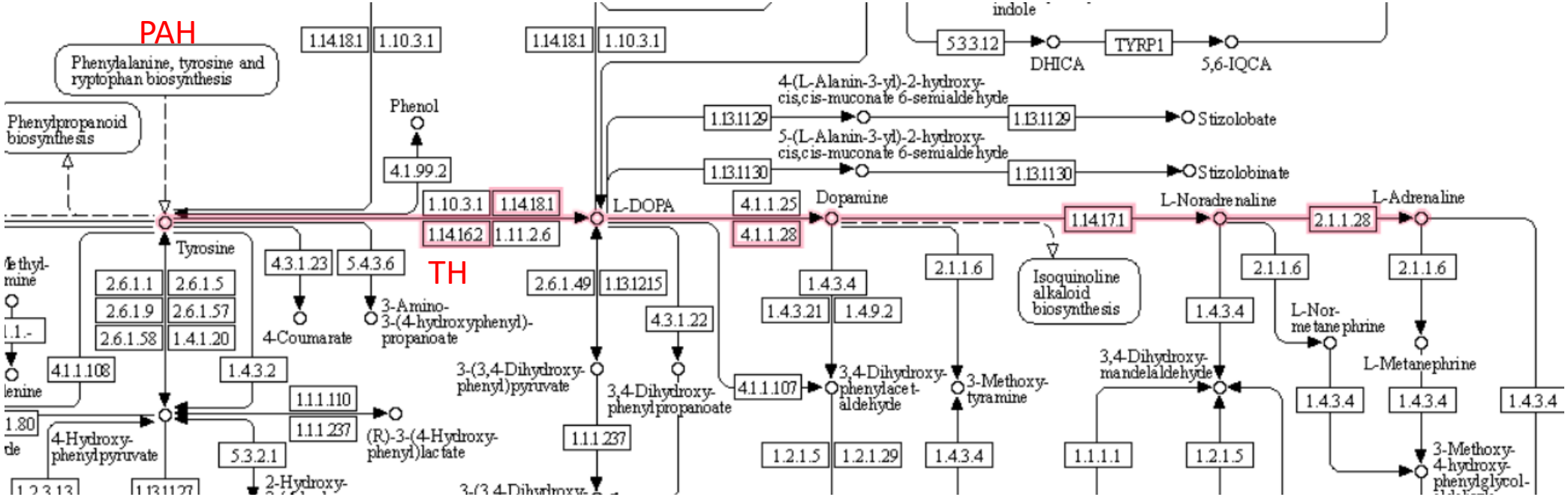


https://www.genome.jp/kegg-bin/show_pathway?hsa04068

More about the aromatic amino acid hydroxylases…



More about the aromatic amino acid hydroxylases…



https://medlineplus.gov/lab-tests/phenylketonuria-pku-screening/

U.S. National Library of Medicine

MedlinePlus
Trusted Health Information for You

Search MedlinePl

About Medlir

Health Topics | Drugs & Supplements | Genetics | Medical Tests | Videos & T

combat COVID. hhs.gov

COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: https://www.coronavirus.gov
Get the latest research information from NIH: https://covid19.nih.gov
Learn more about COVID-19 and you from HHS: https://combatcovid.hhs.gov

Home → Medical Tests → Phenylketonuria (PKU) Screening

Phenylketonuria (PKU) Screening

What is a PKU screening test?

A PKU screening test is a blood test given to newborns 24–72 hours after birth. PKU stands for phenylketonuria, a rare disorder that prevents the body from properly breaking down a substance called phenylalanine (Phe). Phe is part of proteins that are found in many foods and in an artificial sweetener called aspartame.

If you have PKU and eat these foods, Phe will build up in the blood. High levels of Phe can permanently damage the nervous system and brain, causing a variety of health problems. These include seizures, psychiatric problems, and severe intellectual disability.

PKU is caused by a genetic mutation, a change in the normal function of a gene. Genes are the basic units of heredity passed down from your mother and father. For a child to get the disorder, both the mother and father must pass down a mutated PKU gene.

Although PKU is rare, all newborns in the United States are required to get a PKU test.