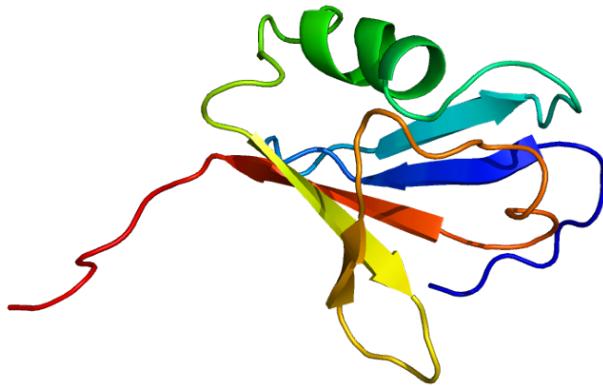


Tutorial: Protein Modelling, Part 2

All online quizzes are only activated during the tutorial. A copy of the questions and suggested answers to the quizzes will be uploaded to KEATS afterwards.

In protein modelling, we use an existing experimental structure of a protein as a template in order to build a model of a different, related protein. Based on the fact that related protein domains have similar structures, relying on homology often yields prediction of the correct fold of proteins with unknown structures. Here, we are going to be using the **protein kinase** domain of **A-Raf** as our protein of interest.



Exercise 1: (Recap) Automated Modelling and Template Search

Please follow the steps below and answer this quiz as you go along:
<https://PollEv.com/surveys/XjTXnGFFGhM899f2I91NB/respond>

In this exercise, we will quickly go over the tools and procedures we used last time in order to make sure everybody is on the same page. Answer the questions in the quiz above and make sure you have access to the following materials from last week:

- A. The ARAF_HUMAN protein kinase domain sequence in FASTA format
- B. The chosen template from your BLAST search
 - a. Sequence in FASTA format
 - b. Structure in PDB format
- C. The template information from BLAST and PDBe concerning its suitability as a template (e.g. experimental method, resolution, sequence identity)

N.B. You can download the chosen template in PDB and FASTA formats from PDBe. We will show you where to find the downloads in the tutorial.

If you were not present last week or cannot find the above materials, we have pre-computed materials below that will allow you to continue the tutorial:

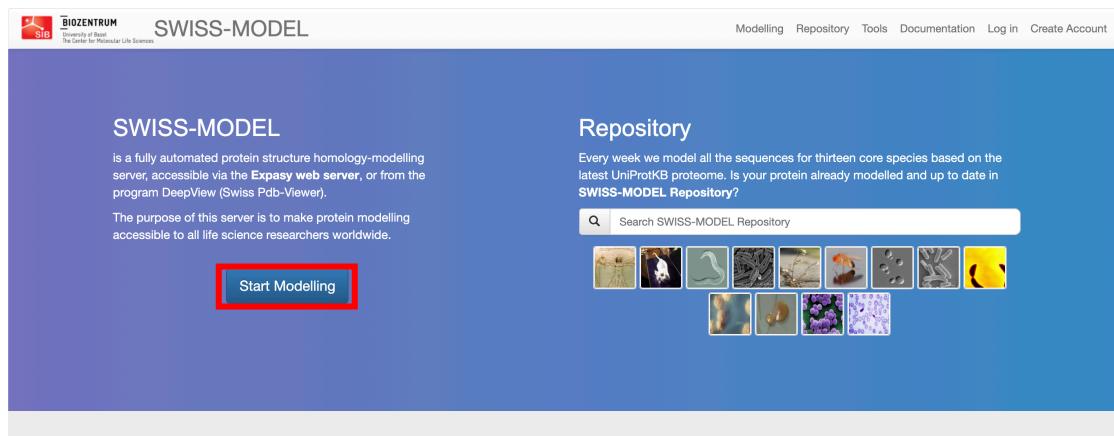
https://github.com/Fraternalilab/5BBBB0226_Protein_Modelling_Tutorial

Exercise 2: Homology Modelling: Alignment Mode

Please follow the steps below and answer this quiz as you go along:
<https://PollEv.com/surveys/MCyRk01Ota4VJiCRcHVBw/respond>

In this exercise, we will use the webserver SWISS-MODEL to perform homology modelling. As detailed in the slides, you can use SWISS-MODEL under two modes: 'automated mode' and 'alignment mode'. You should have already used 'automated mode' previously; here, we will demonstrate 'alignment mode'. You will use the structures of the previously chosen BLAST hits as templates to generate homology models of ARAF_HUMAN ('alignment mode') and compare them to the ones generated by the 'automated mode'.

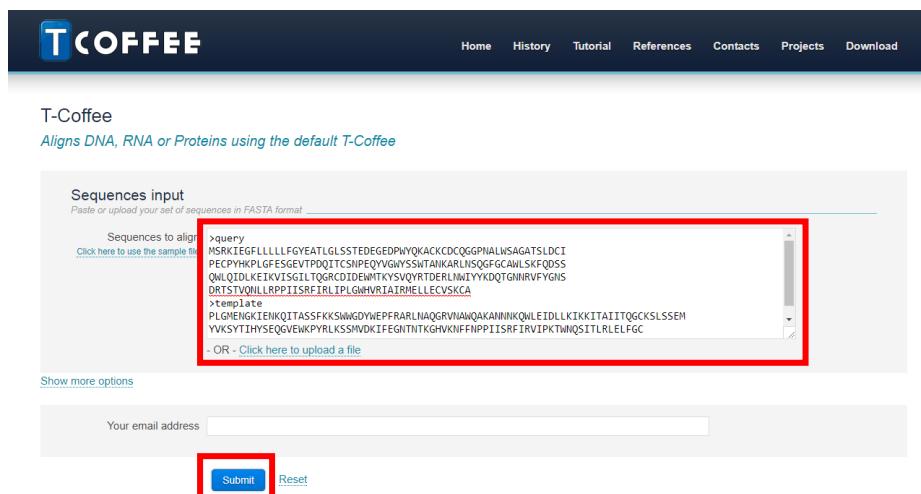
1. Navigate to the SWISS-MODEL web server (<https://swissmodel.expasy.org/>).



The screenshot shows the SWISS-MODEL homepage. The top navigation bar includes links for 'Modelling', 'Repository', 'Tools', 'Documentation', 'Log in', and 'Create Account'. The main content area has a purple header with the 'SWISS-MODEL' logo. Below it, there's a 'Start Modelling' button, which is highlighted with a red box. To the right, there's a 'Repository' section featuring a grid of small protein structure images.

2. Navigate to T-Coffee (<https://www.ebi.ac.uk/Tools/msa/tcoffee/> or <http://tcoffee.crg.cat/> [select "Simple MSA"])

Note: both links direct you to the same T-Coffee program; they are given together in case any one of them are not accessible due to server downtime/maintenance etc. Just use whichever link accessible to you.



The screenshot shows the T-Coffee homepage. The top navigation bar includes links for 'Home', 'History', 'Tutorial', 'References', 'Contacts', 'Projects', and 'Download'. The main content area has a dark header with the 'T-COFFEE' logo. Below it, there's a 'Sequences input' section where sequences can be pasted or uploaded. A 'Submit' button is highlighted with a red box at the bottom of the form. There's also a 'Reset' button.

3. Paste or upload the previous sequence for the ARAF_HUMAN protein kinase domain, and the sequences for all of your chosen hits, all in FASTA format. When you have entered all the sequences you want to align, click "Submit".

N.B. You may wish to title your ARAF_HUMAN sequence as "query" or "target", and your chosen hits as "template" to aid recognition by SWISS-MODEL. This also helps you clarify which sequence is to be modelled and which is to be used as template. HOWEVER, don't mix them up!

The screenshot shows the T-Coffee results page. At the top, there are tabs for 'Input form', 'Web services', 'Help & Documentation', 'Bioinformatics Tools FAQ', 'Feedback', and 'Share'. Below the tabs, it says 'Tools > Multiple Sequence Alignment > T-Coffee'. Under the 'Alignments' tab, there are links for 'Result Summary', 'Phylogenetic Tree', 'Results Viewers', and 'Submission Details'. A red box highlights the 'Download Alignment File' button. Below the buttons, it says 'CLUSTAL W (1.83) multiple sequence alignment'. In the 'Result files' section, there is a table with 8 output files. A red box highlights the 'fasta_aln file' link under the 'Multiple Alignment' row.

Result files	
8 output files - download them all	
Input(s)	Input sequences (400 B)
System	Command line (217 B) Log file (41KB)
Multiple Alignment	score_html file (5KB) clustalw_aln file (885 B) fasta_aln file (477 B) score_ascii file (997 B) phylip file (582 B)
Copy to your Dropbox	

4. Click "Download Alignment File" or download the "fasta_aln file", depending on the server you are using.

The screenshot shows the SWISS-MODEL homepage. At the top, there is a logo for BIOZENTRUM SIB and the text 'SWISS-MODEL'. Below the logo, there is a purple banner with the text 'SWISS-MODEL' and a brief description: 'is a fully automated protein structure homology-modelling server, accessible via the [ExPasy web server](#), or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.' A blue button labeled 'Start Modelling' is visible. In the top right corner, there is a 'Modelling' dropdown menu with options: 'myWorkspace', 'Alignment Mode' (highlighted with a red box), 'User Template Mode', and 'DeepView Project Mode'. A green box highlights the 'Modelling' dropdown menu. Below the dropdown, there is a search bar with the placeholder 'Search SWISS-MODEL Repository' and a grid of small images representing different protein structures.

5. Navigate to the SWISS-MODEL web server again in a new tab/window. From the "Modelling" drop-down menu in the upper bar, select "Alignment Mode".

Start a New Modelling Project

Target-Template Alignment:
(format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

Paste your target-template alignment here

+ Upload Target-Template Alignment File... Validate

Project Title: Untitled Project

Email: Optional

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project

6. Under "Target-Template Alignment", paste or upload the previous alignment of your target and template(s) from T-Coffee.

Start a New Modelling Project

For the uploaded target-template alignment, 2 biounits and / or chains were found to match your template. Please select which biounit you wish to use as the template.

You can avoid this step by using the SMTL ID as the template name in your input alignment, (SMTL ID is <PDB ID>.<biological assembly>.<Chain ID>)

Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project



Target	Sequence	Index
3omv.1.B	SEVQLRKIGTGSGFTVFRGRWHDGVAVKVLKVSQPTAEQAAFKNEMQVLRKTRHVNLLFMGMFTRPG SEV L RIG GSFGTV G WHGDFAVVK LKV PT EQ QAF NE VLRKTRHVNLLFMGMFTRPG	1
3omv.1.B	FAITQWCEGSSLYHHLVADTRDMVQLIDVARQTAQGMIDYLHAKNIIHRDLKSNNIFLHEGLTVKIGDFGLATVKTRWSGAQPLEQPSGSVLMMAEVIRMQD AI TQWCEGSSLY HHLV PT EQ QAF NE VLRKTRHVNLLFMGMFTRPG	2
3omv.1.B	NPYPSFQSDVYAYGVLYELMTGSLPYSHIGCRDQIIFMVGRGYLSPDLSKISSNCPKAMRRLLSCLKFQREERPLFFQILATI AT TQWCEGSSLY HHLV PT EQ QAF NE VLRKTRHVNLLFMGMFTRPG	2

Build Model



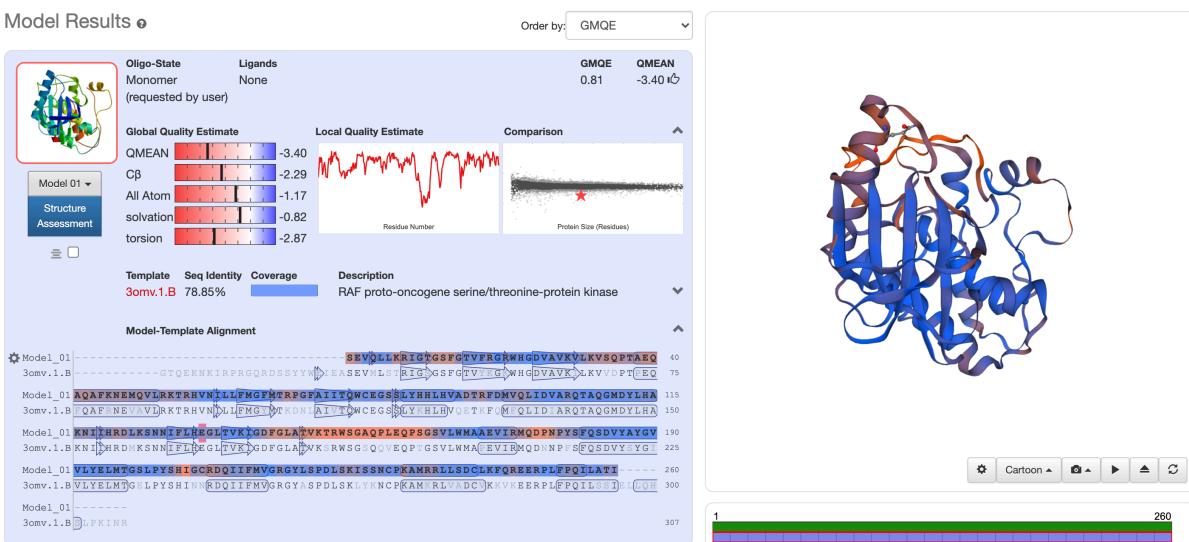
Target	Sequence	Index
3omv.1.A	SEVQLRKIGTGSGFTVFRGRWHDGVAVKVLKVSQPTAEQAAFKNEMQVLRKTRHVNLLFMGMFTRPG SEV L RIG GSFGTV G WHGDFAVVK LKV PT EQ QAF NE VLRKTRHVNLLFMGMFTRPG	1
3omv.1.A	FAITQWCEGSSLYHHLVADTRDMVQLIDVARQTAQGMIDYLHAKNIIHRDLKSNNIFLHEGLTVKIGDFGLATVKTRWSGAQPLEQPSGSVLMMAEVIRMQD AI TQWCEGSSLY HHLV PT EQ QAF NE VLRKTRHVNLLFMGMFTRPG	2
3omv.1.A	NPYPSFQSDVYAYGVLYELMTGSLPYSHIGCRDQIIFMVGRGYLSPDLSKISSNCPKAMRRLLSCLKFQREERPLFFQILATI AT TQWCEGSSLY HHLV PT EQ QAF NE VLRKTRHVNLLFMGMFTRPG	2

Build Model

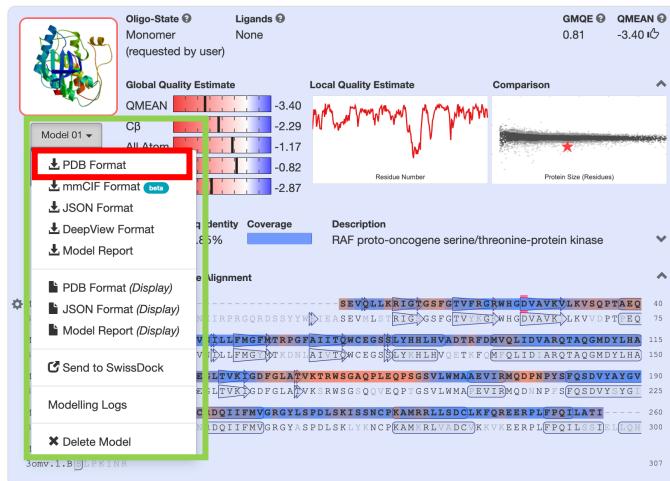
Sequence Identity: 78.85

Sequence Identity: 78.85

7. Click "Build Model". If you see multiple possible templates corresponding to different templates, PDB chains, or assemblies, select any one and click "Build Model".



8. Once SWISS-MODEL has finished running, you should get an output similar to the one above. Take a look at the model quality estimates and the local quality as mapped to the sequence and structure. You can zoom in on the individual quality plots by clicking on them, and reorientate the structure by clicking and dragging.



- From the model drop-down, save the model produced by the alignment mode of SWISS-MODEL as a PDB Format file.

Exercise 3: Protein domain databases

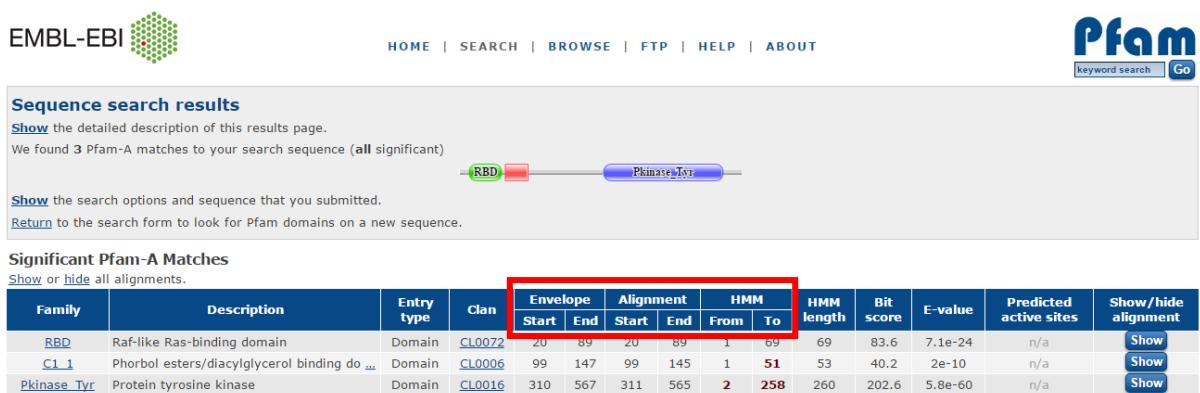
Please follow the steps below and answer this quiz as you go along:
<https://PollEv.com/surveys/u4o6K1D7mGy00ADEQ57qI/respond>

In this exercise we are going to use online databases to compare and annotate domain information on given protein sequences.

1. Obtain protein sequences in FASTA format for the following proteins from UniProt: BRAF_HUMAN, ARAF_HUMAN, RAF1_HUMAN, and KSR1_HUMAN. Check the UniProt entry pages of these proteins and note the availability of structures and their coverage of the different domains listed for the proteins.

A. Querying the Pfam database

2. Submit the sequences to Pfam (<http://pfam.xfam.org/>). On the Pfam home page, click 'SEQUENCE SEARCH' and paste one sequence at a time. Click 'Go'.
3. You should see a list of matches of segments of your submitted sequence against the Pfam domain database. Note that significant and insignificant (according to E-value) matches should be listed separately.



EMBL-EBI 

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam 

Keyword search Go

Sequence search results
Show the detailed description of this results page.
We found 3 Pfam-A matches to your search sequence (all significant)

RBD C1_1 Pkinase_Tyr

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	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
RBD	Raf-like Ras-binding domain	Domain	CL0072	20	89	20	89	1	69	69	83.6	7.1e-24	n/a	Show
C1_1	Phorbol esters/diacylglycerol binding do... ll	Domain	CL0006	99	147	99	145	1	51	53	40.2	2e-10	n/a	Show
Pkinase_Tyr	Protein tyrosine kinase	Domain	CL0016	310	567	311	565	2	258	260	202.6	5.8e-60	n/a	Show

Explanation of the different domain boundaries:

- "Envelope": Region of your submitted sequence which triggers an alignment to the respective domain sequence stored in Pfam.
- "Alignment": Region of your submitted sequence which aligns with high confidence to the Pfam domain sequence. **This should be used wherever available.**
- "HMM": Region of the Pfam domain sequence which aligns to the "Envelope" boundaries of your submitted sequence.

4. Alternatively, you can browse the Pfam entry page of a protein by going back to the Pfam home page, entering the UniProt Accession, and clicking "Go". You will see a cartoon depicting the domain architecture of your protein of interest, as well as a detailed table listing the domain boundaries.

← → ⌂ ⌂ pfam.xfam.org/protein/P10398

Apps Bookmarks Suggested Sites Quick Standalone BL download.mizuguchi European Population 1000 Genomes browser Joel on Software: Annotating TexMed - Exporting RosettaMPA

EMBL-EBI

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Protein: ARAF_HUMAN (P10398)

Summary

Sequence

Structures

TreeFam

Jump to... ↻ enter ID/acc Go

This is the summary of UniProt entry **ARAF_HUMAN** (P10398).

Description: Serine/threonine-protein kinase A-Raf EC=2.7.11.1
Source organism: *Homo sapiens (Human)* (NCBI taxonomy ID 9606) [View Pfam proteome data](#).

Length: 606 amino acids
Reference Proteome: ✓

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 here. It is important to note that, although some 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 details of each domain.

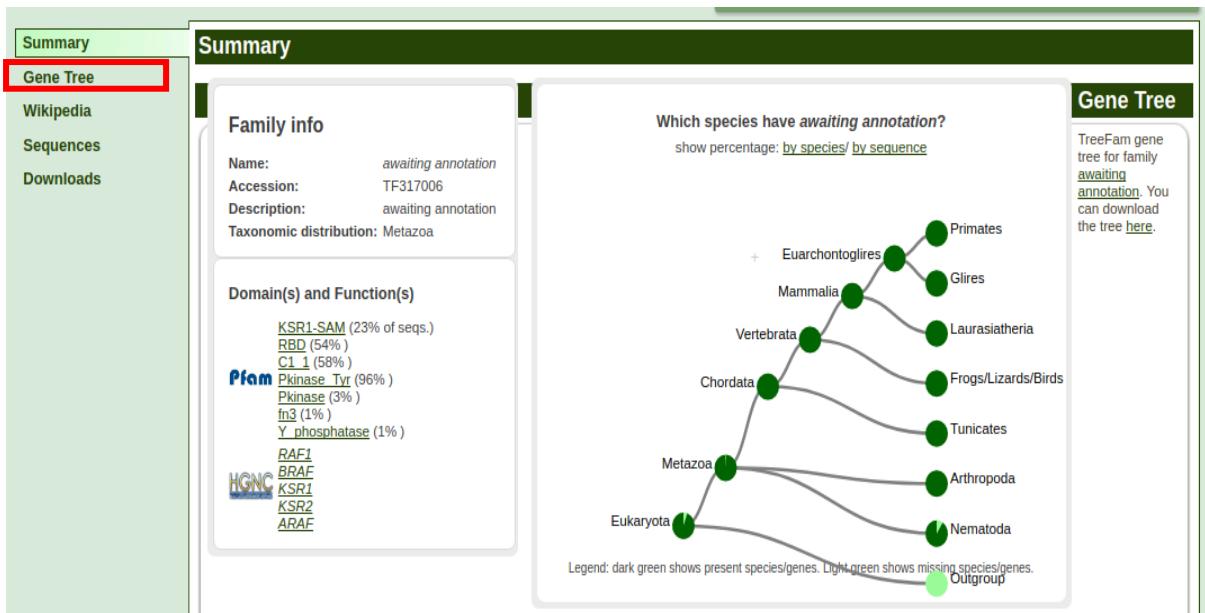
[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
disorder	n/a	1	13
Pfam	RBD	20	89
Pfam	C1_1	99	147
disorder	n/a	156	290
Pfam	Pkinase_Tyr	310	567
disorder	n/a	351	352
disorder	n/a	581	582

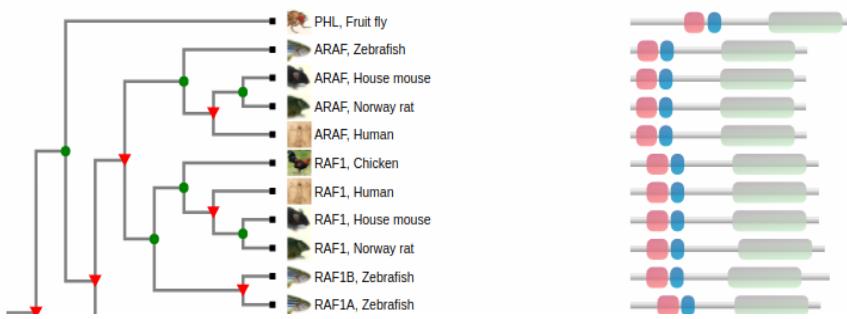
[Show or hide domain scores.](#)

- Pfam offers a gene tree detailing domain architectures of the homologues of your protein of interest. Check to see if there is a corresponding "TreeFam" link on your Pfam entry page. (**Note: Not all proteins will have one!**)

If there is one, click "TreeFam". You should see this: "TreeFam gene tree for family XXXXX", where XXXXX represents a TreeFam identifier which is also hyperlinked. Click on this hyperlink. You should then be directed to a page which looks like this:



- Click "Gene Tree" on the menu bar on the left. You would be directed to a gene tree with domain architecture on the right-hand side of each leaf of this tree:



Explanation of the annotations on the nodes:

- Green circles: speciation event
- Red triangles: duplication event

These are *inferred* by TreeFam.

B. Querying the InterPro database

- To query the InterPro database, go to InterPro (www.ebi.ac.uk/interpro/). You can either paste the FASTA sequence into the search-box to annotate or input the UniProt accession directly into the textbox under "Search by text".
- A typical InterPro entry page for a protein looks like this:

The graphical display shows the domain boundary definitions collected from several databases (including Pfam) and summarised by InterPro. Each is represented as a coloured bar; their location on the horizontal axis indicates their position within the protein of interest. InterPro generates their own "InterPro domain" definitions – these are marked by the bolded entries on the right-hand side of the graphic.