

BGGN213 – Foundations of Bioinformatics

# Find-A-Gene Project

---

Amy Prichard

[aprichar@ucsd.edu](mailto:aprichar@ucsd.edu)

PID: A53263684

[Q1] Tell me the name of a protein you are interested in. Include the species and the accession number. This can be a human protein or a protein from any other species as long as its function is known.

If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online.

Name: Argonaute-1 (AGO1)

Accession: NP\_001246314.1

Species: *Drosophila melanogaster*

[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism).

Method: TBLASTN (NCBI)

Database: Expressed Sequence Tags (est)

Organism: Any

Also include the output of that BLAST search in your document. If appropriate, change the font to Courier New size 10 so that the results are displayed neatly. You can also screen capture a BLAST output. It is **not** necessary to print out all of the blast results if there are many pages.

Translated BLAST: tblastn

blastn blastp blastx **tblastn** tblastx

TBLASTN search translated nucleotide databases using a protein query [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) [Query subrange](#)

NP\_001246314.1 From To

Or, upload file [Browse...](#) No file selected.

Job Title NP\_001246314:Argonaute-1, isoform D [Drosophila... Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database [+](#) Expressed sequence tags (est)

Organism Optional Enter organism name or id--completions will be suggested ☐ exclude Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional ☐ Sequences from type material

Entrez Query Optional Enter an Entrez query to limit search [YouTube](#) [Create custom database](#)

**BLAST** Search database Expressed sequence tags (est) using Tblastn (search translated nucleotide databases using a protein query) ☐ Show results in a new window

[Algorithm parameters](#) [sign](#) Note: Parameter values that differ from the default are highlighted in yellow and marked with [+](#)

On the BLAST results, clearly indicate a match that represents a protein sequence, encoded from some DNA sequence, that is homologous to your query protein. I need to be able to inspect the pairwise alignment you have selected, including the E value and score. It should be labeled a "genomic clone" or "mRNA sequence", etc. - but include no functional annotation.

**Chosen match:** Accession HO777791.1, a 3644 base pair cDNA clone from *Phaseolus acutifolius* (Tepary bean). See below for alignment details.

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

**Job title:** NP\_001246314:Argonaute-1, isoform D [Drosophila...]

**RID:** 6WEA1CHA015 (Expires on 02-22 12:38 pm)

**Query ID:** NP\_001246314.1  
**Description:** Argonaute-1, isoform D [Drosophila melanogaster]  
**Molecule type:** amino acid  
**Query Length:** 984

**Database Name:** est  
**Description:** Database of GenBank+EMBL+DDBJ sequences from EST Divisions  
**Program:** TBLASTN 2.8.1+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#)

**Graphic Summary**

Distribution of the top 106 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

■ <40 ■ 40-50 ■ 50-80 ■ 80-200 ■ >=200

Query

1 150 300 450 600 750 900

**Descriptions**

Sequences producing significant alignments:

Select: **All** None Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	vsu-ars_002824_L10Contig_4970 VSU-ARS-L10 Phaseolus acutifolius cDNA, mRNA sequence	665	665	86%	0.0	42.52%	HO777791.1
<input type="checkbox"/>	EST1221428 Glossina morsitans morsitans Fat body cDNA Glossina morsitans morsitans cDNA clone GLAEP94 3' end, mRNA sequence	621	621	32%	0.0	95.25%	DV618432.1
<input type="checkbox"/>	AM092154 Lutzomyia longipalpis Jacobina whole fly Lutzomyia longipalpis cDNA clone NSF14-48q06, mRNA sequence	615	615	30%	0.0	97.66%	AM092154.1
<input type="checkbox"/>	F03_F03sq1c6_pDNRF_515165 Myzus persicae, tobacco lineage, aphid salivary gland library Myzus persicae cDNA clone F03_F03sq1c6_pDNRF_515165, mRNA	576	576	28%	0.0	97.84%	EE264736.1
<input type="checkbox"/>	EST1208580 Glossina morsitans morsitans Fat body cDNA Glossina morsitans morsitans cDNA clone GLAAX77 3' end similar to CG8671-PA (Drosophila melan	559	559	29%	0.0	93.77%	DV605584.1
<input type="checkbox"/>	FY948757 planarian head cDNA library Dugesia japonica cDNA clone DJ_aH_326_G10, mRNA sequence	550	550	32%	0.0	82.19%	FY948757.1
<input type="checkbox"/>	EST1214812 Glossina morsitans morsitans Fat body cDNA Glossina morsitans morsitans cDNA clone GLACG65 3' end, mRNA sequence	548	548	26%	0.0	99.62%	DV611816.1
<input type="checkbox"/>	ID04DD4CF12RM2 ID04DD Acyrthosiphon pisum cDNA clone ID04DD4CF12 5', mRNA sequence	533	533	28%	0.0	90.49%	CV842652.1
<input type="checkbox"/>	LC03020B1G12.11 Schistocerca gregaria central nervous system cDNA library Schistocerca gregaria cDNA, mRNA sequence	521	521	26%	1e-176	95.31%	JG687356.1
<input type="checkbox"/>	FY953176 planarian head cDNA library Dugesia japonica cDNA clone DJ_aH_506_I04, mRNA sequence	515	515	33%	2e-173	73.25%	FY953176.1
<input type="checkbox"/>	BX743547 XGC-ladpole Xenopus tropicalis cDNA clone TTPA030m04 5', mRNA sequence	512	512	29%	1e-172	84.43%	BX743547.1

Download [GenBank](#) [Graphics](#)

vsu-ars\_002824\_Li10Contig\_4970 VSU-ARS-L10 Phaseolus acutifolius cDNA, mRNA sequence

Sequence ID: [HO777791.1](#) Length: 3644 Number of Matches: 1

Range 1: 925 to 3534 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps	Frame
665 bits(1717)	0.0	Compositional matrix adjust.	382/896(43%)	529/896(59%)	74/896(8%)	+1
Query 136	RRRPNLGREGRPIVLRANHFQVIMPRGYVHHYDINIQPKCPRKVNREIIETMVHAYSKI				195	
Sbjct 925	PLRPGKGSYGKICIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRVMEQLVRLYRES				1104	
Query 196	-FGVLKPFVDGRNNLYTRDPLPIGNERLELEVLPGEG-----KDRIFRVTIKWQAQVS				248	
Sbjct 1105	HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDGAAGGQRRDREFKVVIKLAARAD				1284	
Query 249	LFNLEEALEGRTRQIPYDAILALDVVMRHLPSMTYTFVGRSFFSSPEGYYHPLGGGREVW				308	
Sbjct 1285	LHHLGLFLQGKQTDAPQALQVLDIVLRELPTTRYCPVGRSFSYSPDLGRRQPLGEGLESW				1464	
Query 309	FGFHQSVRPSQWKMMNLIDVSATAFYKAQFVIDFMCEVLDIRDINEQRKPLTDSQRVKFT				368	
Sbjct 1465	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLN-RDVSA--RPLSDADRVKIK				1635	
Query 369	KEIKGLKIEITHCGQMRKRYRVCNVTTRPAQMOSFPLQLENGQVTECTVAKYFLDKYRMK				428	
Sbjct 1636	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV				1806	
Query 429	LRYPHLPCQLQVGQEHKHTYLPLEVCNIVAGQRCIKKLTDMQISTMIKATARSAPDREREI				488	
Sbjct 1807	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTICQRFVERERDI				1986	
Query 489	NNLVKRAFNDNSYVQEFGLTISNSMMEVGRVLPPLQYGGRVSTGLTGQQLFPPQNK				548	
Sbjct 1987	MQTVYHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYH---DTGREKDCL-----				2139	
Query 549	VSLASPNQGVWDMRGKQFFTGVVEIRIWAIACFAPQRTVREDALRNFTQQLQKISNDAGMP				608	
Sbjct 2140	-----PQVGQWNNMNMKNVNGGTNNWF--CINFSRSVQDSVARGFCYELAQMCYISGMA				2298	
Query 609	IIGQPCFCFKYATGPDQVEPMF--RY--LKITFPGLqlvvvvv1---PGKTPVYAEVKRVGD				661	
Sbjct 2299	FNPEFVVPVSARPDQVEKVLKTRYHDAKNKLQGRELDLLIVILPDNNGSLYGLDKRICE				2478	
Query 662	TVLGMATQCQVQAKNVNKTSPQTLNCLKINVKLGGINSLVPSIR---PKVFNEPVI FL				718	
Sbjct 2479	TDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRRNTVLVDALSRRIPLVGDRPTIIF				2658	
Query 719	GADVTHPPAGDNKKPSIAAVVGSMDAHP--SRYAATVRVQQHRQEIIQEL-----				766	
Sbjct 2659	GADVTHPPHGEDSSPSIAAVVASQD-YPEITKYAGLVCAQHRQELIQDLFKQWQDPVRG				2835	
Query 767	---SSMRELLIMFYKSTGGYKPHRIILYRDGVSEGQFPHVLQHELTAREACIKLEPEY				823	
Sbjct 2836	TVTGGMIKELLISFRATG-QKPQRIIFRYRDGVSEGQFYQVLLFELDAIRKACASLEPNY				3012	
Query 824	RPGITFIVVQKRHHTRLFCAEKKEQSG--KSGNIPAGTIVDVGITHPTFDFYLCSHQGI				881	
Sbjct 3013	QPPVTFVTVQKRHHTRLFASNNHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGI				3192	
Query 882	QGTSRPSHYHVLWDDNHFDSDQLCTLYQLCHTYVRCRSVSIPAPAYYAHLVAFRARYH				941	
Sbjct 3193	QGTSRPAHYHVLWDDNNFTADALQTLINNLCYIARCTRSVSIVPPAYYAHLAARFRARY				3372	
Query 942	LVEKEHDSGEGSHQSGC-----SEDRTPGAMARAITVHA---DTKKVMYF				983	
Sbjct 3373	MEPETSDS--GSMTSGAVAGRGMXGGMGRSTRAPGANAAVRPLPALKENVKRVVMFY				3534	

[Q3] Gather information about this “novel” **protein**. At a minimum, show me the protein sequence of the “novel” protein as displayed in your BLAST results from [Q2] as FASTA format (you can copy and paste the aligned sequence subject lines from your BLAST result page if necessary) or translate your novel DNA sequence using a tool called EMBOSS Transeq at the EBI. Don’t forget to translate all six reading frames; the ORF (open reading frame) is likely to be the longest sequence without a stop codon. It may not start with a methionine if you don’t have the complete coding region. Make sure the sequence you provide includes a header/subject line and is in traditional FASTA format.

**Chosen sequence:**

```
>Phaseolus acutifolius protein (sequence taken from BLAST result)
PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYDGRKSLYT
AGPLPFISKEFRITLIDDDGAAGGQRRDREFKVVIKLAARADLHHLGLFLQGKQTDAPQEALQVLDIVLRELPTTR
YCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLNRDVSARPLSDA
DRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMKSVVEYFYETYGFVIQHTQWPCLQVGNT
QRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDIMQTVYHNAYHEDPYAKEFGIKISEKLAQVEA
RILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNGGTVNNWFCINFSSRSVQDSVARGFCYELAQMCIYISGMAFNPE
PVVPPVSARPDQVEKVLKTRYHDAKNKLQGRELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQ
YLANVALKINVKVGGRNTVLVDALSRRIPLVGDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCA
QHRQELIQDLFKQWQDPVRGTVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLE
PNYQPPVTFVQVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAGIQGTSRPAHYHVLWD
ENNFTADALQTLTNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVAGRGMXGGMGRSTRA
PGANA AVRPLPALKENVKRMFY
```

Here, tell me the name of the novel protein, and the species from which it derives. It is very unlikely (but still definitely possible) that you will find a novel gene from an organism such as *S. cerevisiae*, human or mouse, because those genomes have already been thoroughly annotated. It is more likely that you will discover a new gene in a genome that is currently being sequenced, such as bacteria or plants or protozoa.

**Name:** *Phaseolus* protein

**Species:** *Phaseolus acutifolius*  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetales;  
rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50 kb inversion clade; NPAAA  
clade; indigoferoid/millettioid clade; Phaseoleae; Phaseolus.

[Q4] Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, “novel” is defined as follows. Take the protein sequence (your answer to [Q3]), and use it as a query in a blastp search of the nr database at NCBI.

- If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as “unknown”). Someone has already found and annotated this sequence, and assigned it an accession number.
- If the top match reported has less than 100% identity, then it is likely that your protein is novel, and you have succeeded.
- If there is a match with 100% identity, but to a different species than the one you started with, then you have likely succeeded in finding a novel gene.
- If there are no database matches to the original query from [Q1], this indicates that you have partially succeeded: yes, you may have found a new gene, but no, it is not actually homologous to the original query. You should probably start over.

**Details:** Searching this *Phaseolus* protein using BLASTP against non-redundant (NR) database reveals no hits with 100% identity. Although some results are close to 100% identity (the top hit has 99.89% identity), none of them are found in *Phaseolus acutifolius*. The top hit is found in *Phaseolus vulgaris*, and the other hits all appear to be found in different genera. (See BLASTP setup, BLASTP hits, and alignment of the top hit with the query below.)

## BLASTP setup

blastnblastpblastxtblastntblastx

Standard Protein BLAST

BLASTP programs search protein databases using a protein query. [more...](#)

[Reset page](#)[Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) [Query subrange](#)

From   
To

Or, upload file

Browse...

No file selected.

Job Title

Protein Sequence

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database

Non-redundant protein sequences (nr)

[?](#)

Organism

Optional

Enter organism name or id—completions will be suggested [?](#) ☐ exclude [+](#)  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude

Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Entrez Query

Optional

[YouTube](#) [Create custom database](#)  
Enter an Entrez query to limit search [?](#)

Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST) **New**  
☒ blastp (protein-protein BLAST)  
☐ PSI-BLAST (Position-Specific Iterated BLAST)  
☐ PHI-BLAST (Pattern Hit Initiated BLAST)  
☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)  
Choose a BLAST algorithm [?](#)

BLAST

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)  
☐ Show results in a new window

[+ Algorithm parameters](#)

## BLASTP hits

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments	Download	GenPept	Graphics	Distance tree of results	Multiple alignment		
	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">hypothetical protein PHAVU_004G142900g [Phaseolus vulgaris]</a>	1813	1813	100%	0.0	99.89%	<a href="#">XP_007152592.1</a>
<input type="checkbox"/>	<a href="#">protein argonaute 1 [Vigna radiata var. radiata]</a>	1808	1808	100%	0.0	99.43%	<a href="#">XP_014512211.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein LR48_Vigan10q045800 [Vigna angularis]</a>	1808	1808	100%	0.0	99.43%	<a href="#">KOM54566.1</a>
<input type="checkbox"/>	<a href="#">PREDICTED: protein argonaute 1 [Vigna angularis]</a>	1807	1807	100%	0.0	99.43%	<a href="#">XP_017439908.1</a>
<input type="checkbox"/>	<a href="#">Protein argonaute [Cajanus cajan]</a>	1783	1783	100%	0.0	97.94%	<a href="#">KYP61687.1</a>
<input type="checkbox"/>	<a href="#">protein argonaute 1 [Cajanus cajan]</a>	1783	1783	100%	0.0	97.94%	<a href="#">XP_020220591.1</a>
<input type="checkbox"/>	<a href="#">protein argonaute 1 [Abrus precatorius]</a>	1780	1780	100%	0.0	97.82%	<a href="#">XP_027359534.1</a>
<input type="checkbox"/>	<a href="#">protein argonaute 1 [Glycine max]</a>	1779	1779	100%	0.0	97.71%	<a href="#">XP_003534084.1</a>
<input type="checkbox"/>	<a href="#">Protein argonaute 1 [Glycine soja]</a>	1778	1778	100%	0.0	97.71%	<a href="#">KHN23595.1</a>
<input type="checkbox"/>	<a href="#">argonaute 1 [Glycine max]</a>	1772	1772	100%	0.0	97.59%	<a href="#">NP_001340257.1</a>
<input type="checkbox"/>	<a href="#">Protein argonaute 1 [Glycine soja]</a>	1772	1772	100%	0.0	97.59%	<a href="#">KHN24531.1</a>
<input type="checkbox"/>	<a href="#">Protein argonaute 1 [Mucuna pruriens]</a>	1760	1760	100%	0.0	97.24%	<a href="#">RDY09546.1</a>
<input type="checkbox"/>	<a href="#">protein argonaute 1 [Cicer arietinum]</a>	1753	1753	100%	0.0	96.44%	<a href="#">XP_012567214.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein TSUD_295900 [Trifolium subterraneum]</a>	1739	1739	100%	0.0	95.06%	<a href="#">GAU30109.1</a>



## Alignment of top hit

[Download](#) [GenPept](#) [Graphics](#)

hypothetical protein PHAVU\_004G142900g [Phaseolus vulgaris]

Sequence ID: [XP\\_007152592.1](#) Length: 1063 Number of Matches: 1

[▶ See 1 more title\(s\)](#)

Range 1: 193 to 1062 [GenPept](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
1813 bits(4695)	0.0	Compositional matrix adjust.	869/870(99%)	869/870(99%)	0/870(0%)
Query 1	PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES				60
Sbjct 193	PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES				252
Query 61	HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDGAAGGQRRDREFKVVIKLAARAD				120
Sbjct 253	HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDGAAGGQRRDREFKVVIKLAARAD				312
Query 121	LHHLGLFLQGKQTDAPQEAQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW				180
Sbjct 313	LHHLGLFLQGKQTDAPQEAQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW				372
Query 181	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLNRDVSARPLSDADRVKIKKAL				240
Sbjct 373	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLNRDVSARPLSDADRVKIKKAL				432
Query 241	RGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMKSVVEYFYETYGFVIOHTQW				300
Sbjct 433	RGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMKSVVEYFYETYGFVIOHTQW				492
Query 301	PCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTQCRPVERERDIMQTVYH				360
Sbjct 493	PCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTQCRPVERERDIMQTVYH				552
Query 361	NAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVN				420
Sbjct 553	NAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVN				612
Query 421	GGTVNNWFCINFSRVSQDSVARGFCYELAQMCYISGMAFNPEPVVPPVSARPDQVEKVLK				480
Sbjct 613	GGTVNNWFCINFSRVSQDSVARGFCYELAQMCYISGMAFNPEPVVPPVSARPDQVEKVLK				672
Query 481	TRYHDAKNKLQGRELDLLIVILPDNNGSLYGDLCRICEIDLGLVSQCCLTKHVFKMSKQY				540
Sbjct 673	TRYHDAKNKLQGRELDLLIVILPDNNGSLYGDLCRICEIDLGLVSQCCLTKHVFKMSKQY				732
Query 541	LANVALKINVKVGGRNITVLVDALSRRIPLVGDRPTIIFGADVTHPHPGEDSSPSIAAVVA				600
Sbjct 733	LANVALKINVKVGGRNITVLVDALSRRIPLVGDRPTIIFGADVTHPHPGEDSSPSIAAVVA				792
Query 601	SQDYPEITKYAGLVCAQAHQELIQDLFKQWQDPVRGTVTGGMKELLISFRATGQKPQ				660
Sbjct 793	SQDYPEITKYAGLVCAQAHQELIQDLFKQWQDPVRGTVTGGMKELLISFRATGQKPQ				852
Query 661	RIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVTFVWVQKRHHTRLFASNHH				720
Sbjct 853	RIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVTFVWVQKRHHTRLFASNHH				912
Query 721	KSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADALQ				780
Sbjct 913	KSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADALQ				972
Query 781	TLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVAGRMXGG				840
Sbjct 973	TLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVAGRMXGG				1032
Query 841	MGRSTRAPGANAAVRPLPALKENVKRMVY		870		
Sbjct 1033	MGRSTRAPGANAAVRPLPALKENVKRMVY		1062		



[Q5] Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family from different species. A typical number of proteins to use in a multiple sequence alignment for this assignment purpose is a minimum of 5 and a maximum of 20 - although the exact number is up to you. Include the multiple sequence alignment in your report. Use Courier font with a size appropriate to fit page width.

Side-note: Indicate your sequence in the alignment by choosing an appropriate name for each sequence in the input unaligned sequence file (i.e. edit the sequence file so that the species, or short common, names (rather than accession numbers) display in the output alignment and in the subsequent answers below). The goal in this step is to create an interesting alignment for building a phylogenetic tree that illustrates species divergence.

### Re-labeled sequences for alignment:

(15 total sequences; all of the plants are from the legume family except cork oak and peach.)

```
>Fruit_fly (original query; Drosophila melanogaster Argonaute-1)
PRRPNLGREGRPVLRANHFQVTMPRGYVHHYDINIQPDCKPRKVNREIIETMVHAYSKI FGVLPKPVFDG
RNNLYTRDPLPIGNERLELEVTLPGEGKDRIFRVTIKWQAQVSLFNLEEALEGRTRQIPYDAILALDVVM
RHLPSMTYTPVGRSFFSSPEGYYHPLGGGREVWFGFHQSVRPSQWKMLNIDVSATAFYKAQPVIDFMCE
VLDIRDINEQRKPLTDSQVRKFTKEIKGLKIEITHCGQMRRKYRVCNVTRRPAQMQSFPLQLENGQTVEC
TVAKYFLDKYRMKLRYPHLPCLQVGQEHKHTYLPLEVCNIVAGQRCIKKLTDMQTSTMIKATARSAPDRE
REINNVLKRAFDNDSYVQEFGLTISNSMMEVRGRVLPFPKLQYGGRVSTGLTGQQLFPPQNKVSLASPN
QGVWDMRGKQFFTGVEIRIWAIACFAPQRTVREDALRNFTQQLQKISNDAGMPIIGQPCFCKYATGPDQV
EPMFRYLKITFPGLqlvvvvlPGKTPVYAEVKRVGDTVLMGATQCVQAKNVNKTSPQTLNCLCLINVKL
GGINSILVPSIRPKVFNEPVI FLGADVTHPPAGDNKKPSIAAVVGSMDAHPSPRYAATVRVQQRQEIIQE
LSSMVRELLIMFYKSTGGYKPHRIILYRDGVSEGQFPVHLQHELTAREACIKLEPEYRPGITFIVVQKR
HHTRLFCAEKKEQSGKSGNIPAGTTVDVGITHPTFEFDLYLCSHQGIQGTSRPSHYHVLWDDNHFDSDDELQ
CLTYQLCHTYVRCTRSVSIPAPAYY AHLVAFRARYHLVEKEHDSGEGSHQSGCEDRTPGAMARAITVHA
DTKKVMYF
```

```
>Tepary_bean (novel protein; from Phaseolus acutifolius cDNA)
PLRPGKGSYGIKICIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD
GRKSLYTAGPLPFISKEFRITLIDDEGAAGGQRRDREFKVVIKLAARADLHHLGLFLQGKQTDAPQEAL
QVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLP
VIDFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTM
KSVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVER
ERDIMQTVYHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVN
GGTVNNWFCINF SRSVQDSVARGFCYELAQMCIISGMAFNPEPVPPVSARPDQVEKVLKTRYHDAKNKL
QGREL DLLIVILPDNNGSLYGLDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLV
DALSRRIPLVGDRPTIIFGADVTHPPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQELIQDLFKQ
WQDPVRGTVTGGMKELLISFRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPV
TFV VVQKRHHTRLFASNHDKSSVDRSGNILPGTVVDSKICHPTFEFDLYLCSHAGIQGTSRPAHYHVLWD
ENNFTADALQTLTNLCYTYARCTRSVSIVPPAYY AHLAAFRARFYMEPETS DSGSMTSGAVAGRMXGG
MGRSTRAPGANAAVRPLPALKENVKRVMFY
```

```
>Mung_bean XP_014512211.1:190-1059 protein argonaute 1 [Vigna radiata var.
radiata]
PLRPGKGSYGIKICIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD
GRKSLYTAGPLPFISKEFRITLIDDEGAAGGQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEAL
```

QVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLP  
VIDFVTQLLNRDV SARPLSDADRVKIKKALRGKIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTM  
KSVVEYFYETYGFVVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVER  
ERDIMQTVYHNAYHEDPYAQEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVN  
GGTVNHWFCINF SRSVQDSVARGFCYELAQMCIYISGMAFNPEPVPPVSARPDQVEKVLKTRYHDAKNKL  
QGREL DLLIVILPDNNGSLYGDLKRICETDLGLV SQCLTKHVFKMSKQYLANVALKINVKVGGRNTVLV  
DALSRRIPLVGDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQELIQDLFKQ  
WQDPVRGTVTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPV  
TFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLC SHAGIQGTSRPAHYHVLWD  
ENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVAGRGMGGG  
IGRSTRAPGANAAVRPLPALKENVKRMVFY

>Pigeon\_pea KYP61687.1:183-1053 Protein argonaute [Cajanus cajan]  
PLRPGKGSYGIKICIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRVMEQLVKLYRESHLGKRLPAYD  
GRKSLYTAGPLPFLSKEFRITLVDDDEGAGGQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGDGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IDFVTQLLNRDVSTRPLSDADRVKIKKALRGKIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFVVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE  
RDIMQTVHHNAYHDDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNNWFVCNFSRNVDTVARGFCYELAQMCIYISGMAFNPEPVPPVSARPDQVEKVLKTRYHDAKNKLQ  
GREL DLLIVILPDNNGSLYGDLKRICETDLGLV SQCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD  
ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQELIQDLFKQW  
QDPVRGTVTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLC SHAGIQGTSRPAHYHVLWDE  
NNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVAGRGMGGGG  
GLGRSTRAPGASAAVRPLPALKENVKRMVFY

>Jequirity\_bean XP\_027359534.1:181-1050 protein argonaute 1 [Abrus  
preparatorius]  
PLRPGKGSYGTRCIVKANHFFAELPNKDLHQYDVTITPEVPSRGVNRVMEQLVRLYRESHLGKRLPAYD  
GRKSLYTAGPLPFISKEFRITLIDDEGAGGQRRDREFKVVIKFAARADLHHLGLFLQGRQTDAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IDFVTQLLNRDVSSRPLSDADRVKIKKALRGKIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYDITYGFVVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE  
RDIMQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNNWFVCINF SRSNVQDSVARGFCYELAQMCIYISGMAFNPEPVPPLSARPDQVEKVLKTRYHDAKNKLQ  
GREL DLLIVILPDNNGSLYGDLKRICETDLGLV SQCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD  
AISRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQELIQDLFKQW  
QDPVRGTLTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLC SHAGIQGTSRPAHYHVLWDE  
NNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVAGRGMGGGG  
AGRSTRAPGASAAVRPLPALKENVKRMVFY

>Soybean XP\_003534084.1:185-1057 protein argonaute 1 [Glycine max]  
PLRPGKGSYGTCKCVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRVMEQLVRLYRESHLGKRLPAYD  
GRKSLYTAGPLPFMSKEFRIVLADDEGAGGQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IDFVNQLLNRDVSARPLSDADRVKIKKALRGKIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFVVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE  
RDIMQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNNWFVCINF SRSNVQDSVARGFCYELAQMCIYISGMAFTPEPVPPVSARPDQVEKVLKTRYHDAKNKLQ  
GKEL DLLIVILPDNNGSLYGDLKRICETDLGLV SQCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD  
ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQELIQDLFKQW  
QDPVRGTVTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASNHHDKSSFDRSGNILPGTVVDSKICHPTEFDYLC SHAGIQGTSRPAHYHVLWDE  
NNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVAGRGMGGGG  
GGGVGRSTRAPGANAAVRPLPALKENVKRMVFY

>Velvet\_bean RDY09546.1:232-1100 Protein argonaute 1, partial [Mucuna pruriens]  
PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD  
GRKSLYTAGALPFISKEFRITLIDDDGAGQQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IDFVTQLLNRDVSSRPLSDSDRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE  
RDIMQTVHHNAYHDDPYAKEFGIKISEKLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNNWFCINFNRNVQDSVARGFCYELAQMCIYISGMAFNPEPVPPVSSRPDQVEKVLKTRYHDAKSKLQ  
GRDLDLLIVILPDNNGSLYGLDLKRICETDLGLVSCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD  
ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQW  
QDPVRGTVTGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTFEFDYLCSHAGIQGTSRPAHYHVLWDE  
NNFTADGLQTLTNNLCYTYARCTRSVSIVPPAYYAHLAARFRFYMEPETS DSGSMTSGAIGRGMGGGGA  
GRSTRAPGANAAVRPLPALKENVKRTSSY

>Chickpea XP\_012567214.1:225-1093 protein argonaute 1 [Cicer arietinum]  
PLRPGKGCGRKCVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD  
GRKSLYTAGPLPFISKDFRITLVDDDDGTGGQRRDREFKVVIKLAARADLHHLGLFLEGRQTDAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IDFVTQLLNRDVSTRPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPDRE  
RDIMQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNNWFCVNFNRNVQDSVARGFCCELAQMCIYISGMAFNPEPVPPLSARPDQVDKVLKTRYQDAKNKLQ  
GRELDLLIVILPDNNGSLYGLDLKRICETELGLVSCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD  
ALSRRIPLVSDRPTIIFGADVTHPHPGEDMSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQW  
QDPVRGTLTGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASDHRDKSSVDRSGNILPGTVVDSKICHPTFEFDYLCSHAGIQGTSRPAHYHVLWDE  
NNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAARFRFYMEPETS DSGSMTSGAVSGRGMGGGV  
GRSTRAPGANAAVRPLPALKENVKRVMFY

>Red\_clover PNY07943.1:221-1089 protein argonaute 1-like [Trifolium pratense]  
PLRPGKGSTGKRCIVKANHFIAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDSHLGKRLPAYD  
GRKSLYTAGPLPFISKDFRITLVDEDDGSGSQRRDREFKVVIKLAARADLHHLGLFLEGRQTDAPQEALQ  
VLDIVLRELPTSRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IEFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFISIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPDRE  
RDIMQTVHHNAYHDDPYAKEFGIKISDKLAQVEARILPPPWLKYHDTGKEKDCLPQVGQWNMMNKKMVNG  
GTVNNWFCVNFNRNVQDSVARGFCSELAHMCYVSGMAFNPEPVPPLSARPDQVEKVLKTRYHDAKNKLQ  
GRELDLLIVILPDNNGSLYGLDLKRICETDLGVVSCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLID  
ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLFKQW  
QDPARGTLTGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTFEFDYLCSHAGIQGTSRPAHYHVLWDE  
NNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAARFRFYMEPETS DSGSMTSGAVSRGMGAAA  
GRSTRAPGANAAVRPLPALKENVKRVMFY

>Barrelclover KEH26855.1:245-1114 argonaute protein 1A [Medicago truncatula]  
PLRPGKGSYGRKTLVKANHFFAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDSHLGKRLPAYD  
GRKSLYTAGPLPFISKDFRITLVDEDDGSGGQRRDREFKVVIKLAARADLHHLGLFLEGRQTDAPQEALQ  
VLDIVLRELPTSRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IEFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRIAGLTSQATRELTFPVDERGTMK  
SVVEYFFETYGFVIQHTQWPCLQVGNPQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPDRE  
RDIMQTVHHNAYHEDPYAKEFGIKISDRLAQVEARILPAPWLKYNDTGREKDCLPQVGQWNMMNKKMFNG  
GSVKYWLVCVNFRTVQDSVARGFCYELAQMCIYVSGMEFNAEPVVPALTARPDQVEKVLKNRYHDAKSKMP  
KDKELDLLIVILPDNNGSLYGLDLKRICETDLGVVSCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLV  
DALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQ

WQDPVRGTLTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPV  
TFVVVQKRHHTRLFASNHQDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWD  
ENNFSADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAVSRGGAGAA  
VGRSTRAPGANA A VRPLPALKDNVKKVMFY

>Wild\_peanut XP\_016203325.1:189-1055 protein argonaute 1 isoform X2 [Arachis  
ipensis]

PLRPGKGSYGT KCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRESHLGKRLPAYD  
GRKSLYTAGPLPFVSKEFRIMLIDEDDGTGAQRREREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IEFVTQLLNRDVSSRPLSDADRVKIKKALRGKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFSETYGFHIIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE  
RDIMQTVHHNAYDQDPYAKEFGIKISEKLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNHWFCINFSRVQDSVARTFCYELAQMCQVSGMAFNLDPVVPPVSARPDQVEKVLKTRYDADKNKLP  
GKDLDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVYKMSKQYLANVALKINVKVGGRNTVLVD  
ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQELIQDLFKQW  
QDPNRGQVTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASNHHDNRNSIDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE  
NNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAAAGRGMGAGR  
STRAPGASAAVRPLPSLKDNVCRMVY

>Peanut XP\_025701446.1:190-1056 protein argonaute 1-like [Arachis hypogaea]

PLRPGKGSYGT KCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRESHLGKRLPAYD  
GRKSLYTAGPLPFVSKEFRIMLIDEDDGTGAQRREREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ  
VLDIVLRELPTTRYSPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IEFVTQLLNRDVSTRPLSDADRVKIKKALRGKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFSETYGFHIIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE  
RDIMQTVHHNAYDQDPYAKEFGIKISEKLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNHWFCINFSRVQDSVARTFCYELAQMCQVSGMAFNLDPVVPPVSARPDQVEKVLKTRYDADKNKLP  
GKDLDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVYKMSKQYLANVALKINVKVGGRNTVLVD  
ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQELIQDLFKQW  
QDPNRGQVTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT  
FVVVQKRHHTRLFASNHHDNRNSIDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE  
NNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGAAAGRGMGAGR  
STRAPGASAAVRPLPSLKDNVCRMVY

>Castor Bean XP\_002526275.1:191-1062 protein argonaute 1 [Ricinus communis]

PLRPGKGSTGIR CIVKANHFFAELPKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRESHLGKRLPAYD  
GRKSLYTAGPLPFISKEFKITLIDEDDGSGGQRREREFKVVIKLAARADLHHLGLFLQGRQADAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IDFVNQLLNRDVSSRPLSDADRVKIKKALRGVKEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFVIQHTQWPCLQVGNQQRPNYLPMEVCKVVEGQRYSKRLNERQITALLKVTCQRPQERE  
RDIMQTVHHNAYGNDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNNWICINFSRVQDSVARGFCYELAQMCYISGMAFNPEPVLPPVSARPEQVEKVLKTRYHDAMTKLQ  
QGKELDLLIVILPDNNGSLYGELKRICETDLGLVSQCCLTKHVFRMNKQYLANVALKINVKVGGRNTVLV  
DALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHQELIQDLFKE  
WQDPVRGRVTGGMKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPV  
TFVVVQKRHHTRLFANNHNDNRNAVDSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWD  
ENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHAAFRARFYMEPETS DSGSMTSGPVGGRGGMGG  
GAGARSTRGPAASAAVRPLPALKENVCRMVY

>Cork\_oak XP\_023880904.1:197-1067 protein argonaute 1 [Quercus suber]

PLRPGKGSTGTCMVKANHFFAELPKDLHQYDVSITPEVTSRGVNRAVMEQLVKLYKETHLGKRLPAYD  
GRKSLYTAGPLPFLSKEFKIPLIDEDDGSGGQRREREFKVVIKLAARADLHHLGLFLQGRQADAPQEALQ  
VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IEFVTQLLNRDVTSRPLSDADRVKIKKALRGVKEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFAIQHTQWPCLQVGNQQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPQERE

RDIMQTV EHNAYHNDPYAKEFGIKISERLASVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GTVNNWICINF SRNVQDSVARGFCYELAQMCYISGMAFNPEPVLPLSARPDQVERVLKTRYHDAMTKIQ  
PQGKELDLLIVILPDNNGSLYGLDKRICETDLGLVSQCCLTKHVFRMSKQYLANVALKINVKVGG RNTVL  
VDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQ AHRQELIQDLFK  
TWQDPVRGTVSGMIKELLISFRRATGQKPKRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPP  
VTFVVVQKRHHTRLFANNHHD RNAVDKSGNILPGTVVDSKICHPTEFDFYLC SHAGIQGTSRPAHYHVLW  
DENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETS DSGSMTSGAAAGRGMGG  
AGPRSTRPPGANA AVRPLPALKENVKRVMFY

>Peach XP\_007210410.1:201-1068 protein argonaute 1 [Prunus persica]  
PLRPGKGSTGIRCTVKANHFFAELPKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRESHLGKRLPAYD  
GRKSLYTAGPLPFLSKEFKIILIDEDDGGQRREREF RVVIKFAARADLHHLGLFLQGRQADAPQEALQ  
VLDIVLRELPTSRYCPVGRSFYAPDLGRRQSLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV  
IEFVTQLLNRDVT HRPLSDSDRVKIKKALRGVKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK  
SVVEYFYETYGFVIQHTQWPCLQVGNQQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVT CQRPHDRE  
QDIMRTVRHNAYHEDPYAKEFGIKISENLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG  
GKVNNWICINF SRNVQDSVARGFCSELAQMCYISGMAFNPEPVLPPISARPDQVEKVLKTRYHDAMTKLR  
VQGKELDLLVILPDNNGSLYGLDKRICETDLGLVSQCCLTKHVFRMSKQYLANVALKINVKVGG RNTVL  
VDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQ AHRQELIQDLFK  
TWQDPARGTVSGMIKELLISFRRATGQKPKRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPP  
VTFVVVQKRHHTRLFANNHHD RNTVDRSGNILPGTVVDSKICHPTEFDFYLC SHAGIQGTSRPAHYHVLW  
DENKFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETS DSGSMTSGAPGRGGMGA  
RSTRAPGANAAVRPLPALKENVKRVMFY

**Alignment:**  
(Obtained using MUSCLE version 3.8 at EBI)

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Fruit_fly	PRRPNLGREGRPIVLRANHFQVTMPRGYVHHYDINIQPKDCPRKVNREIIETMVHAYSKI
Barrelclover	PLRPGKGSYGRKTLVKANHFFAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDS
Wild_peanut	PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRES
Peanut	PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRES
Red_clover	PLRPGKGSTGKRCIVKANHFIAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDS
Chickpea	PLRPGKGSYGRKCVVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES
Velvet_bean	PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES
Jequirity_bean	PLRPGKGSYGTRCIVKANHFFAELPNKDLHQYDVTITPEVPSRGVNRAVMEQLVRLYRES
Pigeon_pea	PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRES
Soybean	PLRPGKGSYGTKCVVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES
Tepary_bean	PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES
Mung_bean	PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES
Peach	PLRPGKGSTGIRCTVKANHFFAELPKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRES
Castor_bean	PLRPGKGSTGIRCIVKANHFFAELPKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRES
Cork_oak	PLRPGKGSTGTCMVKANHFFAELPKDLHQYDVSITPEVTSRGVNRAVMEQLVKLYKET
	* *. * * :.***** . : * :*:***:.* *: .* *** :*: *. * .

Fruit_fly	-FGVLKP VFDGRNNLYTRDPLPIGNERLELEVTLPGEG-----KDRIFRVTIKWQAQVS
Barrelclover	HLGKRLPAYDGRKSLYTAGPLPFISKDFRITLVDEDDG-SGGQRRDREFKVVIKLAARAD
Wild_peanut	HLGKRLPAYDGRKSLYTAGPLPFVFSKEFRIMLIDEDDG-TGAQRREREFKVVIKLAARAD
Peanut	HLGKRLPAYDGRKSLYTAGPLPFVFSKEFRIMLIDEDDG-TGAQRREREFKVVIKLAARAD
Red_clover	HLGKRLPAYDGRKSLYTAGPLPFISKDFRITLVDEDDG-SGSQRRDREFKVVIKLAARAD
Chickpea	HLGKRLPAYDGRKSLYTAGPLPFISKDFRITLVDDDDG-TGGQRRDREFKVVIKLAARAD
Velvet_bean	HLGKRLPAYDGRKSLYTAGALPFISKEFRITLIDDDG-AGQQRDREFKVVIKLAARAD
Jequirity_bean	HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDG-AGGQRRDREFKVVIKFAARAD
Pigeon_pea	HLGKRLPAYDGRKSLYTAGPLPFLSKEFRITLVDDDEG-AGGQRRDREFKVVIKLAARAD
Soybean	HLGKRLPAYDGRKSLYTAGPLPFMSKEFRIVLADDDEG-AGGQRRDREFKVVIKLAARAD

Tepary_bean	HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDDEGAAGGQRRDREFKVVIKLAARAD
Mung_bean	HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDDEGAAGGQRRDREFKVVIKLAARAD
Peach	HLGKRLPAYDGRKSLYTAGPLPFLSKEFKIILIDEDDG-PGGQRREREFVVIKFAARAD
Castor_bean	HLGKRLPAYDGRKSLYTAGPLPFISKEFKITLIDEDDG-SGGQRREREFVVIKLAARAD
Cork_oak	HLGKRLPAYDGRKSLYTAGPLPFLSKEFKIPLIDEDDG-SGGQRREREFKVVIKLAARAD
	:* *.:***:.* **.: : : : : :.* :.* *.*.* ** ..

Fruit_fly	LFNLEEALGRTRQIPYDAILALDVMRHLPSMTYTPVGRSFFSSPEGYHYPLGGGREVW
Barrelclover	LHHLGLFLEGRQTDAPQEALQVLDIVLRELPTSRYCPVGRSFYSPDLGRRQPLGEGLESW
Wild_peanut	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Peanut	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYSPVGRSFYSPDLGRRQPLGEGLESW
Red_clover	LHHLGLFLEGRQTDAPQEALQVLDIVLRELPTSRYCPVGRSFYSPDLGRRQPLGEGLESW
Chickpea	LHHLGLFLEGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Velvet_bean	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Jequirity_bean	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Pigeon_pea	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGDGLESW
Soybean	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Tepary_bean	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Mung_bean	LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Peach	LHHLGLFLQGRQADAPQEALQVLDIVLRELPTSRYCPVGRSFYAPDLGRRQSLGEGLESW
Castor_bean	LHHLGLFLQGRQADAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
Cork_oak	LHHLGLFLQGRQADAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW
	*.:* *.:* : * :*: .*:*: * *: * *****:.. * :.* * * *

Fruit_fly	FGFHQSVRPSQWKMLNIDVSATAFYKAQPVIDFMCEVLDIRDINEQRKPLTDSQRVKFT
Barrelclover	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIEFVTQLLN-RDVSA--RPLSDADRVKIK
Wild_peanut	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIEFVTQLLN-RDVSS--RPLSDADRVKIK
Peanut	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIEFVTQLLN-RDVST--RPLSDADRVKIK
Red_clover	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIEFVTQLLN-RDVSA--RPLSDADRVKIK
Chickpea	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLN-RDVST--RPLSDADRVKIK
Velvet_bean	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLN-RDVSS--RPLSDSDRVKIK
Jequirity_bean	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLN-RDVSS--RPLSDADRVKIK
Pigeon_pea	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLN-RDVST--RPLSDADRVKIK
Soybean	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVNQLLN-RDVSA--RPLSDADRVKIK
Tepary_bean	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLN-RDVSA--RPLSDADRVKIK
Mung_bean	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLN-RDVSA--RPLSDADRVKIK
Peach	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIEFVTQLLN-RDVTH--RPLSDSDRVKIK
Castor_bean	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVNQLLN-RDVSS--RPLSDADRVKIK
Cork_oak	RGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIEFVTQLLN-RDVTS--RPLSDADRVKIK
	**:*:*:*:* : *****:*** :. ***: :*: :*: :.* :.*:***:..

Fruit_fly	KEIKGLKIEITHCGQMRRKYRVCNVTRRPAQMOSFPLQLENGQTVECTVAKYFLDKYRMK
Barrelclover	KALRGIKVEVTHRGNMRRKYRIAGLTSQATRELTFPVD-ERGTMK--SVVEYFFETYGFV
Wild_peanut	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFSETYGFH
Peanut	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFSETYGFH
Red_clover	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFS
Chickpea	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Velvet_bean	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Jequirity_bean	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYDITYGFV
Pigeon_pea	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Soybean	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Tepary_bean	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Mung_bean	KALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Peach	KALRGVKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Castor_bean	KALRGVKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFV
Cork_oak	KALRGVKVEVTHRGNMRRKYRISGLTSQATRELTFPVD-ERGTMK--SVVEYFYETYGFA
	* :.*:*:*:* *:*:*:*:..* :. :. :*: :.* :.*:*** :.* :

Fruit_fly	LRYPHLPCLQVGQEHKHTYLPLEVCNIVAGQRCIKKLTDMQTSTMIKATARSAPDREREI
Barrelclover	IQHTQWPCQLQGNPQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPLDRERDI

Wild_peanut	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Peanut	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Red_clover	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPPLDRERDI
Chickpea	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPPLDRERDI
Velvet_bean	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Jequirity_bean	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Pigeon_pea	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Soybean	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Tepary_bean	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Mung_bean	IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI
Peach	IQHTQWPCLQVGNGQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPDREQDI
Castor_bean	IQHTQWPCLQVGNGQRPNYLPMEVCKVVEGQRYSKRLNERQITALLKVTCQRPQERERDI
Cork_oak	IQHTQWPCLQVGNGQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPQERERDI
	::: *****: .: .*****: * ** * *: *: * *: * *: * *: *

Fruit_fly	NNLVKRADFNND SYVQEFGLTISNSMMEVGRVLPPPKLQYGGRVSTGLTGQQLFPPQNK
Barrelclover	MQTVHHNAYHEDPYAKEFGIKISDRLAQVEARILPAPWLKYN---DTGREKDCL-----
Wild_peanut	MQTVHHNAYDQDPYAKEFGIKISEKLAQVEARILPPPWLKYH---DTGREKDCL-----
Peanut	MQTVHHNAYDQDPYAKEFGIKISEKLAQVEARILPPPWLKYH---DTGREKDCL-----
Red_clover	MQTVHHNAYHDDPYAKEFGIKISDKLAQVEARILPPPWLKYH---DTGKEKDCL-----
Chickpea	MQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYH---DTGREKDCL-----
Velvet_bean	MQTVHHNAYHDDPYAKEFGIKISEKLAQVEARILPPPWLKYH---DTGREKDCL-----
Jequirity_bean	MQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYH---DTGREKDCL-----
Pigeon_pea	MQTVHHNAYHDDPYAKEFGIKISEKLAQVEARILPAPWLKYH---DTGREKDCL-----
Soybean	MQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYH---DTGREKDCL-----
Tepary_bean	MQTVYHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYH---DTGREKDCL-----
Mung_bean	MQTVYHNAYHEDPYAQEFGIKISEKLAQVEARILPAPWLKYH---DTGREKDCL-----
Peach	MRTVRHNAYHEDPYAKEFGIKISENLAQVEARILPPPWLKYH---DTGREKDCL-----
Castor_bean	MQTVHHNAYGNDPYAKEFGIKISEKLASVEARILPAPWLKYH---DTGREKDCL-----
Cork_oak	MQTVHEHNAYHNDPYAKEFGIKISERLASVEARILPAPWLKYH---DTGREKDCL-----
	* * * * *

Fruit_fly	VSLASPNQGVWDMRGKQFFTGVEIRIWAIACFAPQRTVREDALRNFTQQQLQKISNDAGMP
Barrelclover	-----PQVGQWNMMNKKMFNGGSVKYW--LCVNFSRTVQDSVARGFCYELAQMCIYVSGME
Wild_peanut	-----PQVGQWNMMNKKMVNGGTVNHWW--FCINFSRVNQDSVARTFCYELAQMCQVSGMA
Peanut	-----PQVGQWNMMNKKMVNGGTVNHWW--FCINFSRVNQDSVARTFCYELAQMCQVSGMA
Red_clover	-----PQVGQWNMMNKKMVNGGTVNNW--FCVNFSRVNQDSVARGFCSELAHMCYVSGMA
Chickpea	-----PQVGQWNMMNKKMVNGGTVNNW--FCVNFSRVNQDSVARGFCCELAQMCIYISGMA
Velvet_bean	-----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRVNQDSVARGFCYELAQMCIYISGMA
Jequirity_bean	-----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRVNQDSVARGFCYELAQMCIYISGMA
Pigeon_pea	-----PQVGQWNMMNKKMVNGGTVNNW--FCVNFSRVNQDVTARGFCYELAQMCIYISGMA
Soybean	-----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRVNQDSVARGFCYELAQMCIYISGMA
Tepary_bean	-----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRSVQDSVARGFCYELAQMCIYISGMA
Mung_bean	-----PQVGQWNMMNKKMVNGGTVNHWW--FCINFSRSVQDSVARGFCYELAQMCIYISGMA
Peach	-----PQVGQWNMMNKKMVNGGKVNNW--ICINFSRVNQDSVARGFCSELAQMCIYISGMA
Castor_bean	-----PQVGQWNMMNKKMVNGGTVNNW--ICINFSRVNQDSVARGFCYELAQMCIYISGMA
Cork_oak	-----PQVGQWNMMNKKMVNGGTVNNW--ICINFSRVNQDSVARGFCYELAQMCIYISGMA
	*:

Fruit_fly	IIQQPCFCKYATGPDQVEPMF--RY----LKIT--FPGQLQLVVVLPGKT-PVYAEVKRV
Barrelclover	FNAEPVVPALTARPDQVEKVLKNRYHDAKSKMP-KDKELDLLIVILPDNNGSLYGDLKRI
Wild_peanut	FNLDPVVPPVSARPDQVEKVLKTRYDDAKNKL--GKDLDLLIVILPDNNGSLYGDLKRI
Peanut	FNLDPVVPPVSARPDQVEKVLKTRYDDAKNKL--GKDLDLLIVILPDNNGSLYGDLKRI
Red_clover	FNPEPVVPLSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI
Chickpea	FNPEPVVPLSARPDQVDKVLKTRYQDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI
Velvet_bean	FNPEPVVPPVSSRPDQVEKVLKTRYHDAKSKLQ--GRDLDLLIVILPDNNGSLYGDLKRI
Jequirity_bean	FNPEPVVPLSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI
Pigeon_pea	FNPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI
Soybean	FTPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ--GKELDLLIVILPDNNGSLYGDLKRI
Tepary_bean	FNPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI



Mung_bean	FNPEPVPPVPSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI
Peach	FNPEPVLPPISARPDQVEKVLKTRYHDAMTKLRFVQGKELDLLVILPDNNGSLYGDLKRI
Castor_bean	FNPEPVLPPVSARPEQVEKVLKTRYHDAMTKLQ--QGKELDLLIVILPDNNGSLYGELKRI
Cork_oak	FNPEPVLPLSARPDQVEKVLKTRYHDAMTKIQFGKELDLLIVILPDNNGSLYGDLKRI
	: * : : * : : : * : *
	: * : : * : : : * : *

Fruit_fly	GDTVLGMATQCVQAKNVNKTSPQTLNLCLKINVKVLGGINSILVPSIR---PKVFNEPVI
Barrelclover	CETDLGVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Wild_peanut	CETDLGLVVSQCCLTKHVYKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Peanut	CETDLGLVVSQCCLTKHVYKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Red_clover	CETDLGVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLIDALSRRIPLVSDRPTI
Chickpea	CETELGLVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Velvet_bean	CETDLGLVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Jequirity_bean	CETDLGLVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Pigeon_pea	CETDLGLVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Soybean	CETDLGLVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Tepary_bean	CETDLGLVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVGDRPTI
Mung_bean	CETDLGLVVSQCCLTKHVFKMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVGDRPTI
Peach	CETDLGLVVSQCCLTKHVFRMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Castor_bean	CETDLGLVVSQCCLTKHVFRMNKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
Cork_oak	CETDLGLVVSQCCLTKHVFRMSKQYLANVALKINVKVGGGRNTVLVDALSRRIPLVSDRPTI
	: * * *: : * * : * : * : * * : : * * : : * * : * * : : * * : * * : : * * : * *

Fruit_fly	FLGADVTHPPAGDNKKPSIAAVVGSMD-AHPSRYAATVRVQQHRQEI IQEL-----
Barrelclover	IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Wild_peanut	IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQRQELIQDLFKQWQDPNR
Peanut	IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQRQELIQDLFKQWQDPNR
Red_clover	IFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHQRQELIQDLFKQWQDPAR
Chickpea	IFGADVTHPHPGEDMSPSIAAVVASQDWPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Velvet_bean	IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Jequirity_bean	IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Pigeon_pea	IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Soybean	IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Tepary_bean	IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Mung_bean	IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Peach	IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHQRQELIQDLFKTWQDPAR
Castor_bean	IFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHQRQELIQDLFKQWQDPVR
Cork_oak	IFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHQRQELIQDLFKTWQDPVR
	::***** *:: ***** * *:: ***** * ***** *::

[illegible]

Fruit_fly	YRPGITFIVVQKRHHTRLFCAEKKEQSG--KSGNIPAGTTVDVGITHPTFDFYLCSHQG
Barrelclover	YQPPVTFVVVQKRHHTRLFASNHQDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG
Wild peanut	YQPPVTFVVVQKRHHTRLFASNHHDRNSIDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Peanut	YQPPVTFVVVQKRHHTRLFASNHHDRNSIDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Red_clover	YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Chickpea	YQPPVTFVVVQKRHHTRLFASDHRDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Velvet_bean	YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Jequirity_bean	YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Pigeon_pea	YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Soybean	YQPPVTFVVVQKRHHTRLFASNHHDKSSFDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Tepary_bean	YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Mung_bean	YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Peach	YQPPVTFVVVQKRHHTRLFANNHHDNRNTVDRSGNILPGTVVDSKICHPTEFDYLCSHAG
Castor_bean	YQPPVTFVVVQKRHHTRLFANNHHDNRNAVDSGNILPGTVVDSKICHPTEFDYLCSHAG
Cork_oak	YQPPVTFVVVQKRHHTRLFANNHHDNRNAVDSGNILPGTVVDSKICHPTEFDYLCSHAG
	*.* :*:*****. :.:... .**** .**.* * ***** *
Fruit_fly	IQGTSRPSHYHVLWDDNHFDSDDELQCLTYQLCHTYVRCTRSVSIPAPAYYAHLLVAFRARY
Barrelclover	IQGTSRPAHYHVLWDENNFSADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Wild_peanut	IQGTSRPAHYHVLWDENNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Peanut	IQGTSRPAHYHVLWDENNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Red_clover	IQGTSRPAHYHVLWDENNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Chickpea	IQGTSRPAHYHVLWDENNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Velvet_bean	IQGTSRPAHYHVLWDENNFTADGLQTLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Jequirity_bean	IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Pigeon_pea	IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Soybean	IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Tepary_bean	IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Mung_bean	IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Peach	IQGTSRPAHYHVLWDENKFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Castor_bean	IQGTSRPAHYHVLWDENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
Cork_oak	IQGTSRPAHYHVLWDENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLLAFAFRARF
	*****:*****:.* :* ** * :*:**.****** .*****.*****:
Fruit_fly	HLVEKEHDSGE-----GSHQSGCEDRTPGAMA--RAI-TVHADTKKVMFY
Barrelclover	YMEPETSDSGSMTSGAVSRGG---AGAAVG-RSTRAPGANAAVRPLPALKDNVKKVMFY
Wild_peanut	YMEPETSDSGSMTSGAAAGRGM-----GAG-RSTRAPGASAAVRPLPSLKDENVKRMFY
Peanut	YMEPETSDSGSMTSGAAAGRGM-----GAG-RSTRAPGASAAVRPLPSLKDENVKRMFY
Red_clover	YMEPETSDSGSMTSGAVSRGGM----GAAAG-RSTRAPGANAAVRPLPALKENVKRMFY
Chickpea	YMEPETSDSGSMTSGAVSRGGM----GGGVG-RSTRAPGANAAVRPLPALKENVKRMFY
Velvet_bean	YMEPETSDSGSMTSGAI-GRGM---GGGGAG-RSTRAPGANAAVRPLPALKENVKRTSSY
Jequirity_bean	YMEPETSDSGSMTSGAVAGRGM---GGGGAG-RSTRAPGASAAVRPLPALKENVKRMFY
Pigeon_pea	YMEPETSDSGSMTSGAVAGRGM--GGGGGLG-RSTRAPGASAAVRPLPALKENVKRMFY
Soybean	YMEPETSDSGSMTSGAVAGRGMGGGGGGVG-RSTRAPGANAAVRPLPALKENVKRMFY
Tepary_bean	YMEPETSDSGSMTSGAVAGRGM----XGGMG-RSTRAPGANAAVRPLPALKENVKRMFY
Mung_bean	YMEPETSDSGSMTSGAVAGRGM----GGGIG-RSTRAPGANAAVRPLPALKENVKRMFY
Peach	YMEPETSDSGSMTSGAP-----GRGGMGARSTRAPGANAAVRPLPALKENVKRMFY
Castor_bean	YMEPETSDSGSMTSGPVGGRRG---MGGGAGARSTRGPAASAAVRPLPALKENVKRMFY
Cork_oak	YMEPETSDSGSMTSGAAAGRGM-----GGAGPRSTRPPGANAAVRPLPALKENVKRMFY
	: : * : * . * . * * * * . : : : : . * . :

[Q6] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use “simple phylogeny” online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.

I imported the output from MUSCLE into SeaView to create a distance-based phylogenetic tree (using bootstrapping). See next page for distance analysis parameters and phylogenetic tree.

Distance analysis

☐ NJ ☒ BioNJ ☐ Save to File

Distance Observed

Ignore all gap sites ☐

# of replicates

☒ Bootstrap 100

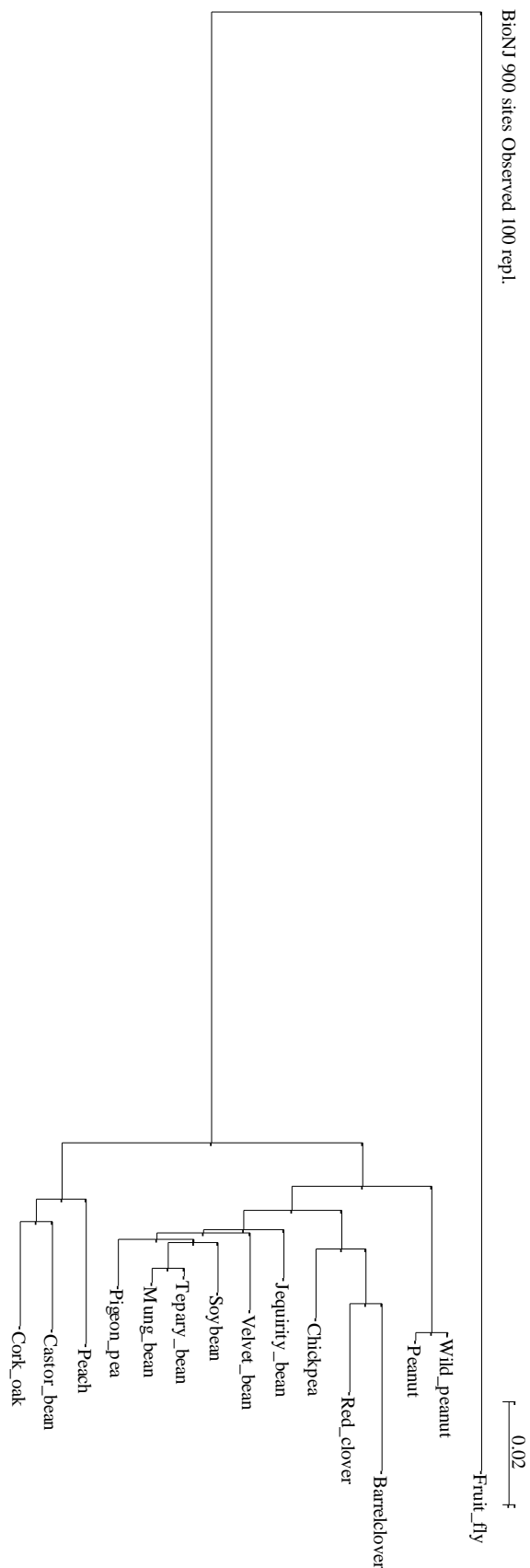
☐ Transfer Bootstrap Expectation method

☐ Show bootstrap trees

Optimize branch lengths

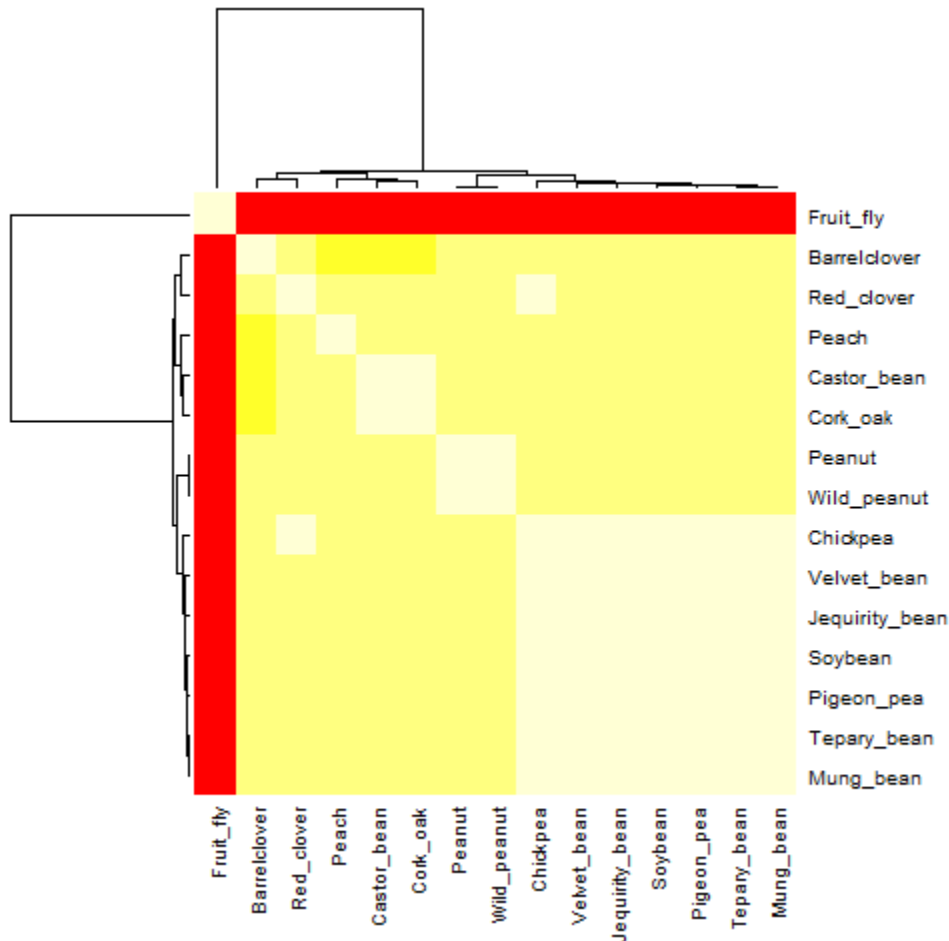
☐ User tree:

Cancel Go



[Q7] Generate a sequence identity based **heatmap** of your aligned sequences using R.

If necessary, convert your sequence alignment to the ubiquitous FASTA format (Seaview can read in clustal format and “Save as” FASTA format for example). Read this FASTA format alignment into R with the help of functions in the Bio3D package. Calculate a sequence identity matrix (again using a function within the **Bio3D package**). Then generate a heatmap plot and add to your report. Do make sure your labels are visible and not cut at the figure margins.



[Q8] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.

List the top 3 *unique* hits (i.e. not hits representing different chains from the same structure) along with their Evalue and sequence identity to your query. Please also add annotation details of these structures. For example, include the annotation terms PDB identifier (structureId),

Method used to solve the structure (experimentalTechnique), resolution (resolution), and source organism (source).

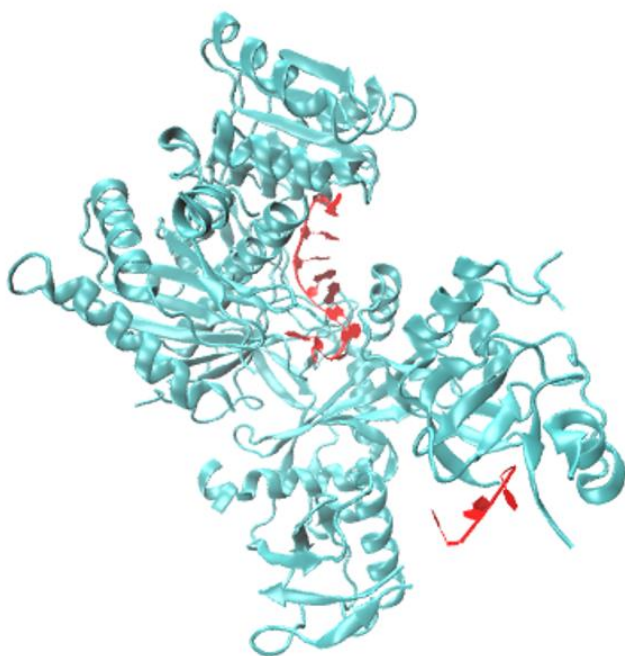
HINT: You can use a single sequence from your alignment or generate a consensus sequence from your alignment using the Bio3D function `consensus()`. The Bio3D functions `blast.pdb()`, `plot.blast()` and `pdb.annotate()` are likely to be of most relevance for completing this task. Note that the results of `blast.pdb()` contain the hits PDB identifier (or `pdb.id`) as well as Evalue and identity. The results of `pdb.annotate()` contain the other annotation terms noted above.

Note that if your consensus sequence has lots of gap positions then it will be better to use an original sequence from the alignment for your search of the PDB. In this case you could chose the sequence with the highest identity to all others in your alignment by calculating the row-wise maximum from your sequence identity matrix.

ID	Technique	Resolution	Source	Evalue	Identity
5VM9	X-RAY DIFFRACTION	3.28	<i>Homo sapiens</i>	0	44.549
3VNA	X-RAY DIFFRACTION	2	<i>Arabidopsis thaliana</i>	7.67E-84	84.459
5THE	X-RAY DIFFRACTION	2.1	<i>Vanderwaltozyma polyspora</i>	1.44E-52	33.506

[Q9] Generate a molecular figure of one of your identified PDB structures using **VMD**. You can optionally highlight conserved residues that are likely to be functional. Please use a white or transparent background for your figure (i.e. not the default black).

Based on sequence similarity. How likely is this structure to be similar to your “novel” protein?



5VM9

This structure might be somewhat similar to the tepary bean argonaute, but since its identity is only 44.549%, it would probably have some differences.

In the figure to the left, the cyan cartoon represents the argonaute protein, and the red ribbon portion represents the guide RNA that is bound to argonaute.

[Q10] Perform a “Target” search of ChEMBL (<https://www.ebi.ac.uk/chembl/>) with your novel sequence. Are there any **Target Associated Assays** and **ligand efficiency data** reported that may be useful starting points for exploring potential inhibition of your novel protein?

ChEMBL has information about 6 binding assays (CHEMBL2311232), but no data on ligand efficiency.

<https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL2311232>

The first listed binding assay discovered inhibitors of pre-miRNA degradation as well as enhancers of miRNA processing in a high-throughput screen of a few thousand small molecules. The overarching goal is to develop drugs that could be used in miRNA therapy to rescue global miRNA defects.

K. Asada, E. Canestrari, Z. Paroo, *Bioorganic & Medicinal Chemistry Letters* A druggable target for rescuing microRNA defects. *Bioorg. Med. Chem. Lett.* **26**, 4942–4946 (2016).

<http://europepmc.org/abstract/MED/27641467>

The remaining five binding assays detail xanthone derivatives (specifically, an aminoalkoxy-substituted thioxanthone derivative) inhibiting miRNA processing. They believe this molecule could be developed as a drug to treat miRNA related disease.

A. Murata, T. Fukuzumi, S. Umemoto, K. Nakatani, *Bioorganic & Medicinal Chemistry Letters* Xanthone derivatives as potential inhibitors of miRNA processing by human Dicer : Targeting secondary structures of pre-miRNA by small molecules. *Bioorg. Med. Chem. Lett.* **23**, 252–255 (2013).

<http://europepmc.org/abstract/MED/23164709>