BGGN213 – Foundations of Bioinformatics

Find-A-Gene Project

Amy Prichard

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*

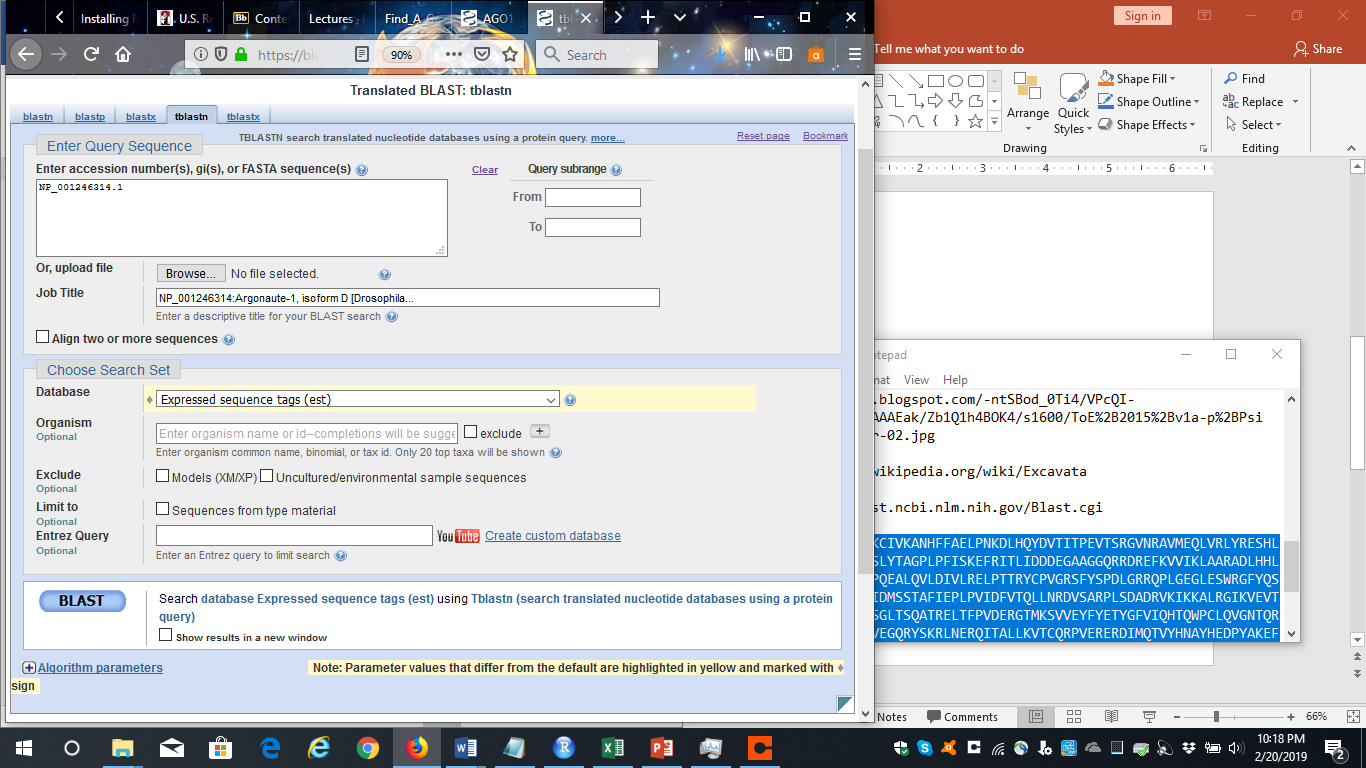
[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched and any limits applied (e.g. Organism).

**Method**: TBLASTN (NCBI)

**Database**: Expressed Sequence Tags (est)

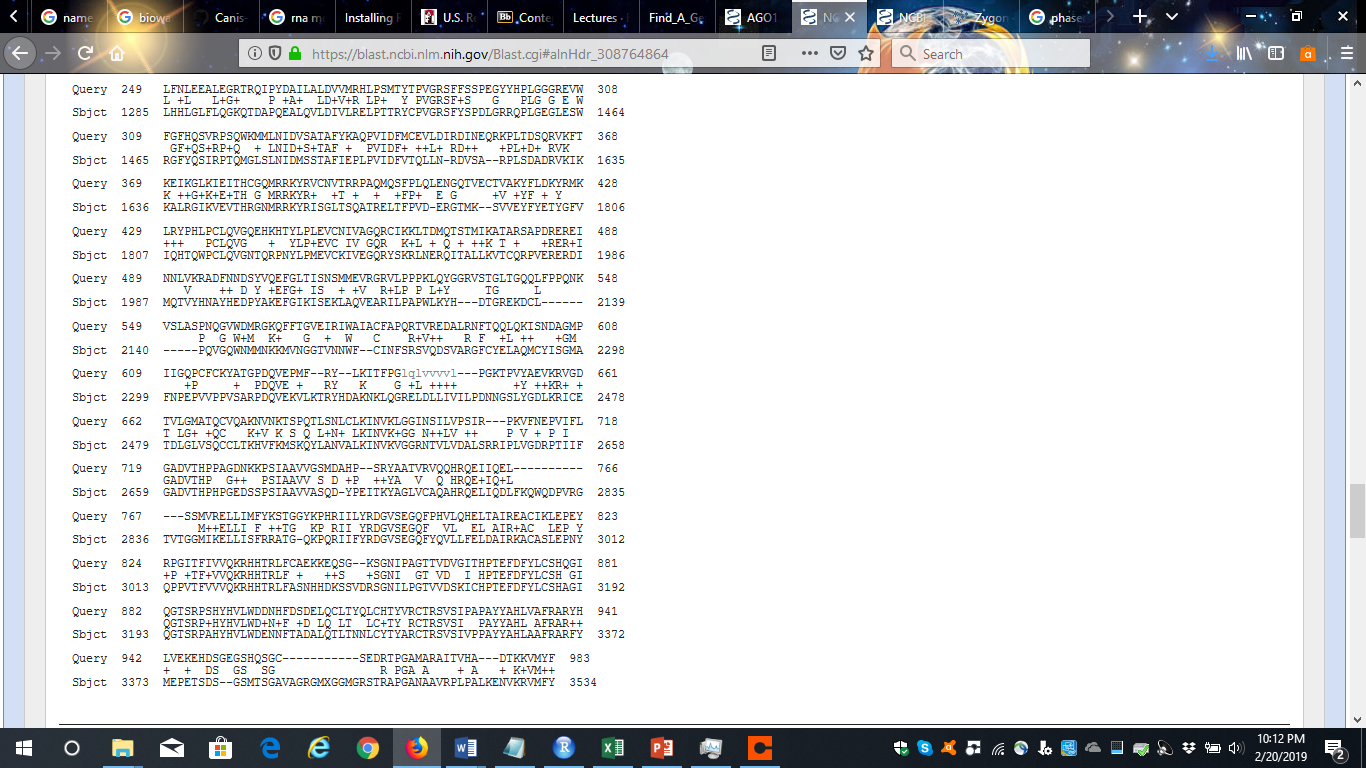
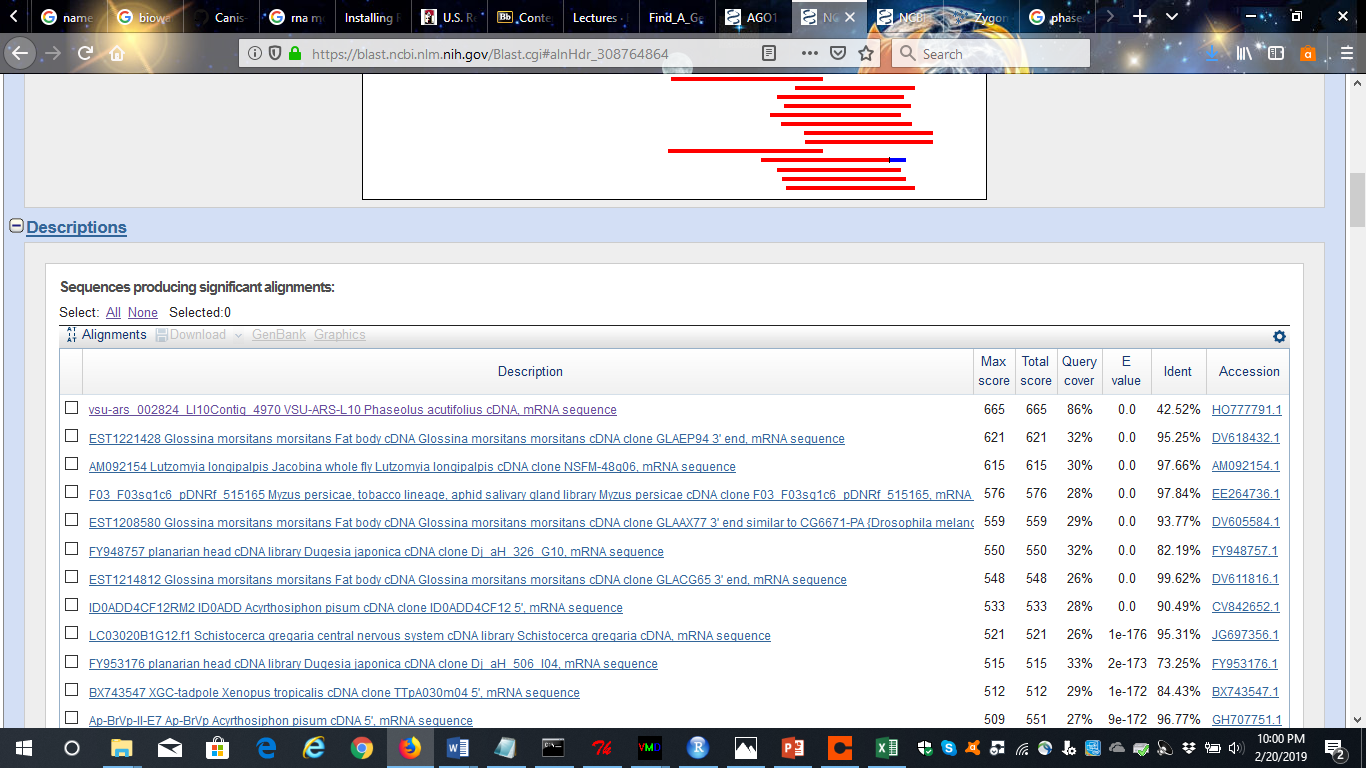
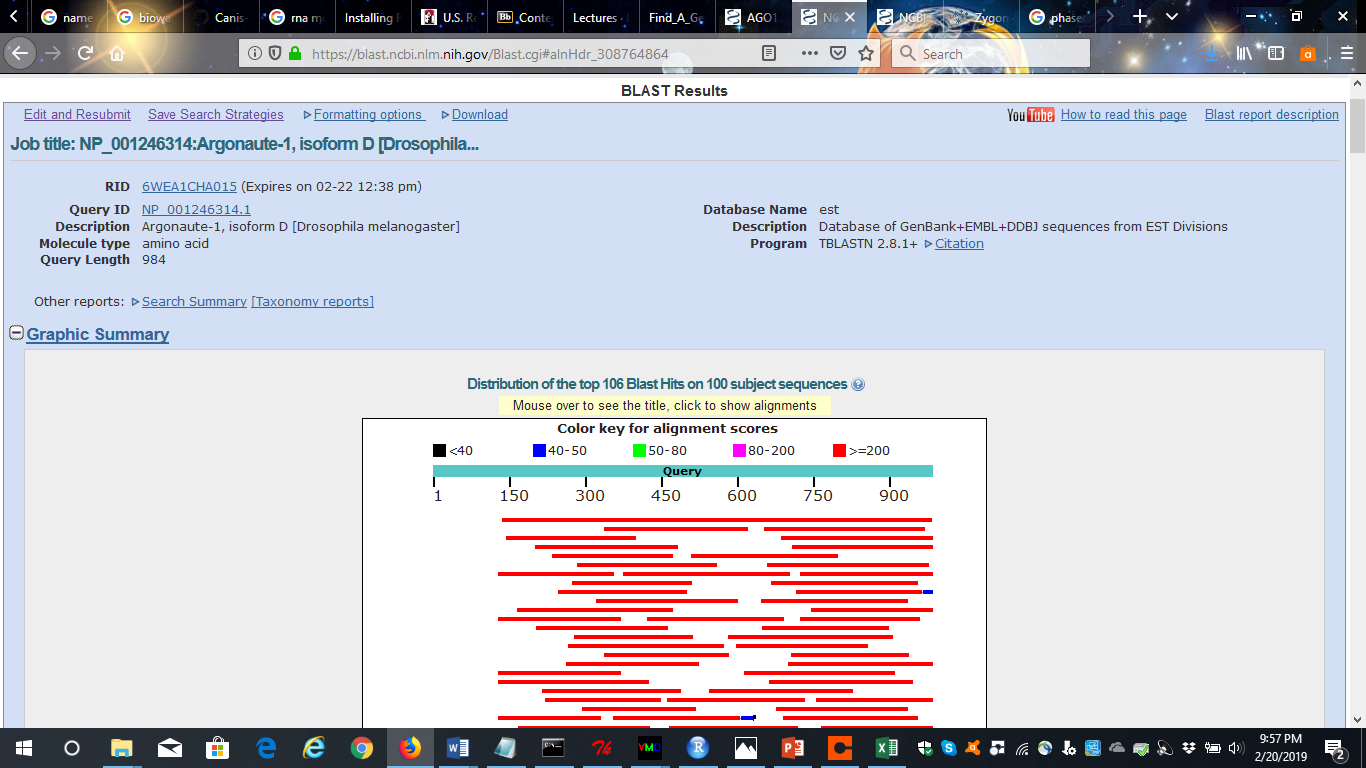
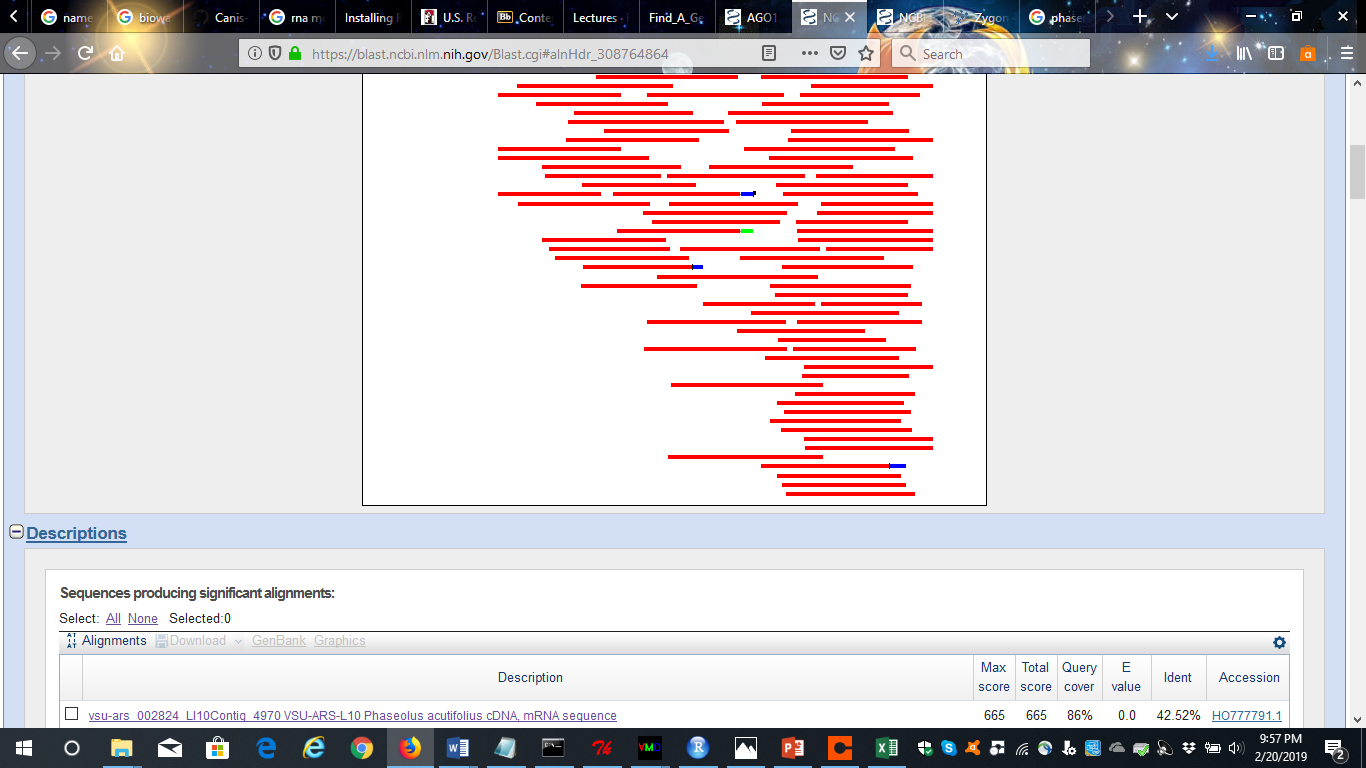
**Organism**: Any

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.



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.



[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)  
PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDEGAAGGQRRDREFKVVIKLAARADLHHLGLFLQGKQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPVIDFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMKSVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDIMQTVYHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNGGTVNNWFCINFSRSVQDSVARGFCYELAQMCYISGMAFNPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQGRELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVGDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQWQDPVRGTVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVAGRGMXGGMGRSTRAPGANAAVRPLPALKENVKRVMFY

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

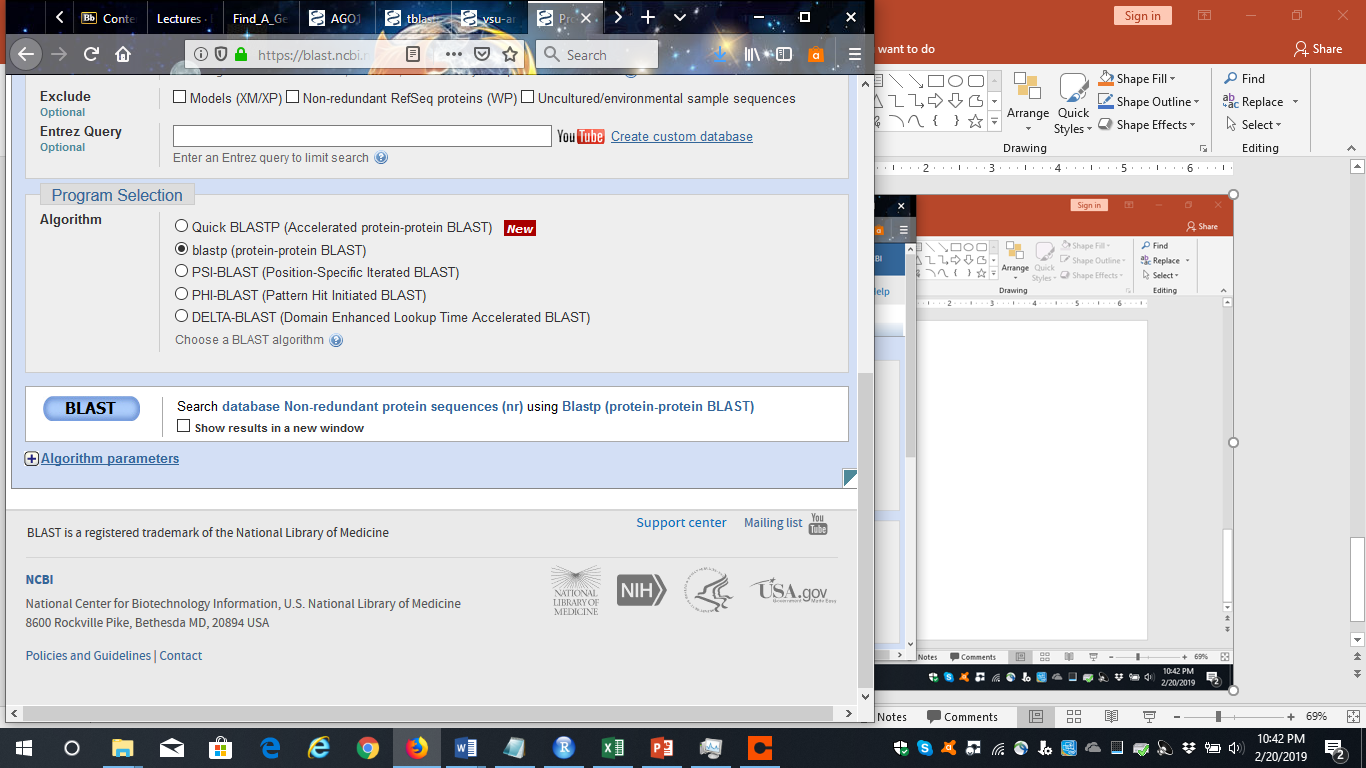
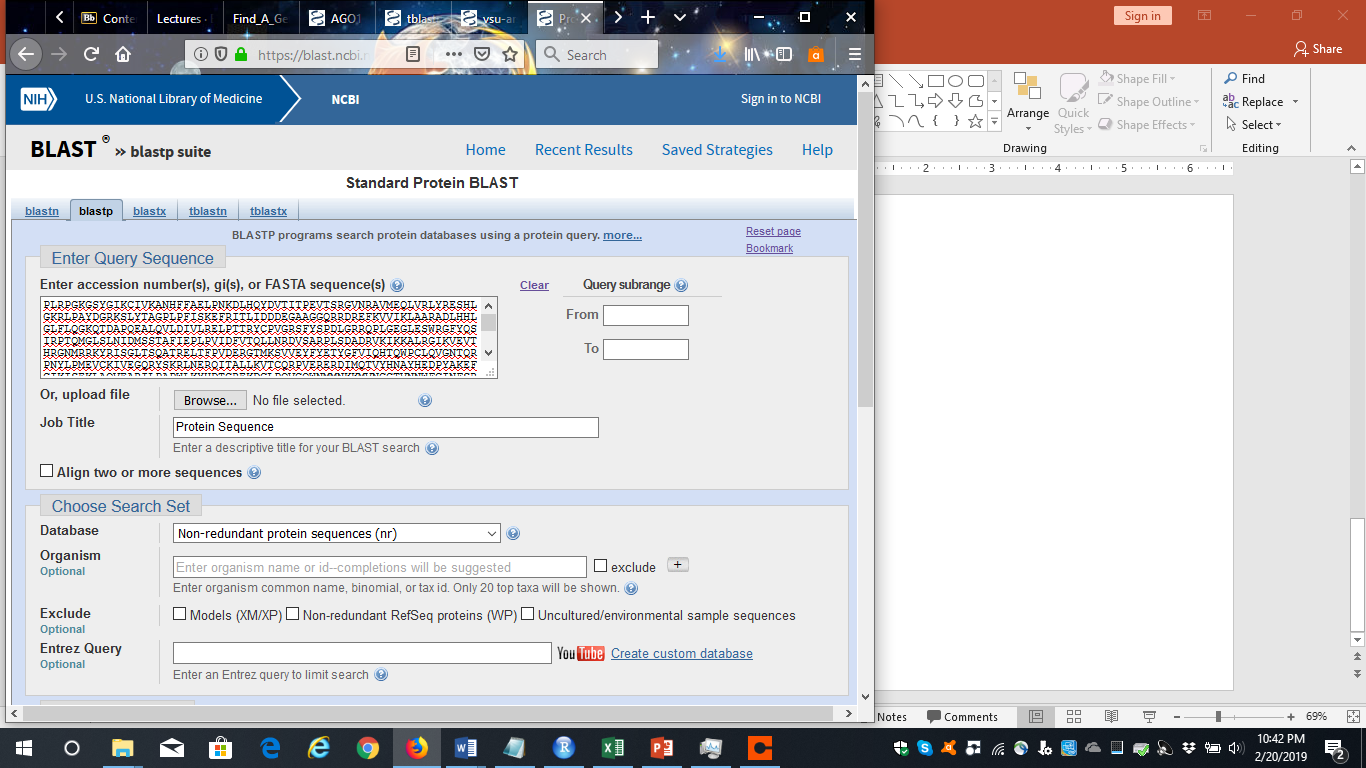
**Species**: *Phaseolus acutifolius*  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Fabales; Fabaceae; Papilionoideae; 50 kb inversion clade; NPAAA clade; indigoferoid/millettioid clade; Phaseoleae;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

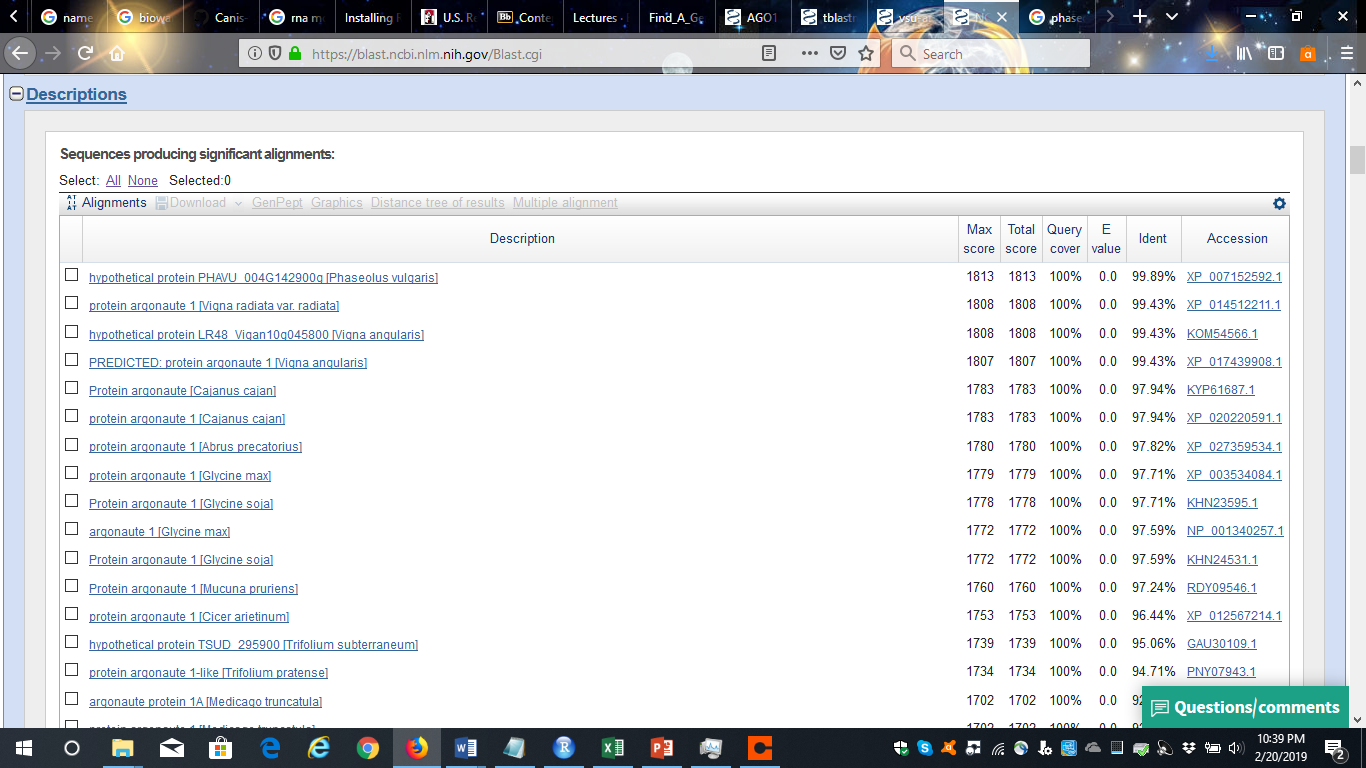
* If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as “unknown”). Someone has already found and annotated this sequence, and assigned it an accession number.
* If the 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.)

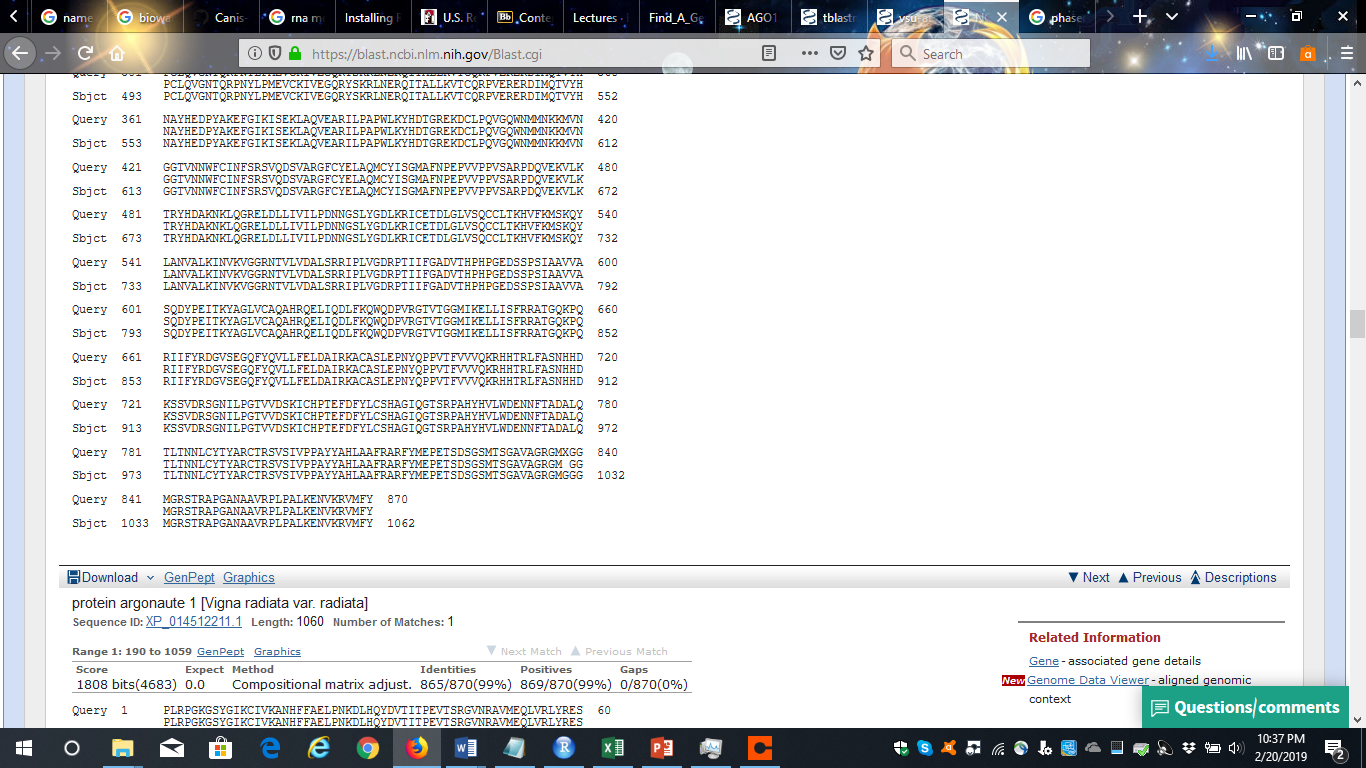
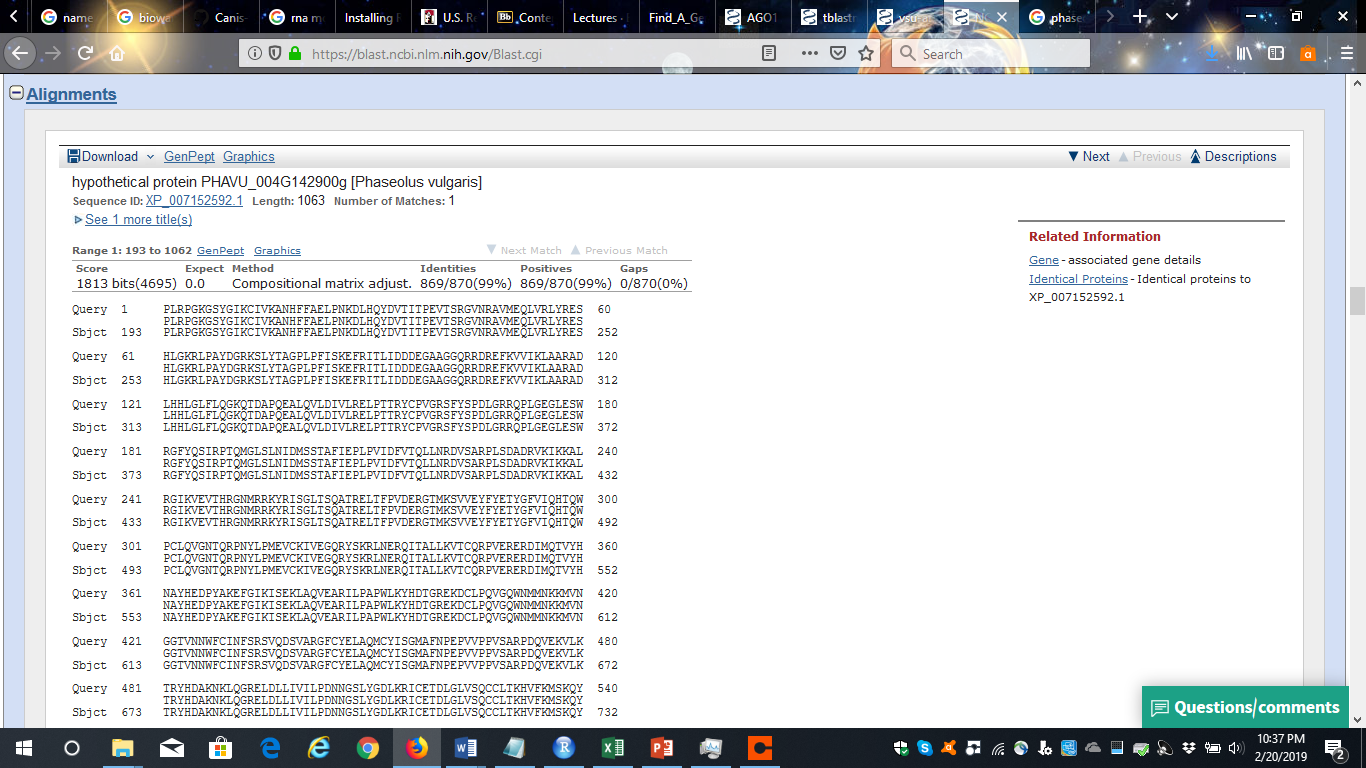
**BLASTP setup**



**BLASTP hits**



**Alignment of top hit**



[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)

PRRPNLGREGRPIVLRANHFQVTMPRGYVHHYDINIQPDKCPRKVNREIIETMVHAYSKIFGVLKPVFDG

RNNLYTRDPLPIGNERLELEVTLPGEGKDRIFRVTIKWQAQVSLFNLEEALEGRTRQIPYDAILALDVVM

RHLPSMTYTPVGRSFFSSPEGYYHPLGGGREVWFGFHQSVRPSQWKMMLNIDVSATAFYKAQPVIDFMCE

VLDIRDINEQRKPLTDSQRVKFTKEIKGLKIEITHCGQMRRKYRVCNVTRRPAQMQSFPLQLENGQTVEC

TVAKYFLDKYRMKLRYPHLPCLQVGQEHKHTYLPLEVCNIVAGQRCIKKLTDMQTSTMIKATARSAPDRE

REINNLVKRADFNNDSYVQEFGLTISNSMMEVRGRVLPPPKLQYGGRVSTGLTGQQLFPPQNKVSLASPN

QGVWDMRGKQFFTGVEIRIWAIACFAPQRTVREDALRNFTQQLQKISNDAGMPIIGQPCFCKYATGPDQV

EPMFRYLKITFPGlqlvvvvlPGKTPVYAEVKRVGDTVLGMATQCVQAKNVNKTSPQTLSNLCLKINVKL

GGINSILVPSIRPKVFNEPVIFLGADVTHPPAGDNKKPSIAAVVGSMDAHPSRYAATVRVQQHRQEIIQE

LSSMVRELLIMFYKSTGGYKPHRIILYRDGVSEGQFPHVLQHELTAIREACIKLEPEYRPGITFIVVQKR

HHTRLFCAEKKEQSGKSGNIPAGTTVDVGITHPTEFDFYLCSHQGIQGTSRPSHYHVLWDDNHFDSDELQ

CLTYQLCHTYVRCTRSVSIPAPAYYAHLVAFRARYHLVEKEHDSGEGSHQSGCSEDRTPGAMARAITVHA

DTKKVMYF

>Tepary\_bean (novel protein; from Phaseolus acutifolius cDNA)  
PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD

GRKSLYTAGPLPFISKEFRITLIDDDEGAAGGQRRDREFKVVIKLAARADLHHLGLFLQGKQTDAPQEAL

QVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLP

VIDFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTM

KSVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVER

ERDIMQTVYHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVN

GGTVNNWFCINFSRSVQDSVARGFCYELAQMCYISGMAFNPEPVVPPVSARPDQVEKVLKTRYHDAKNKL

QGRELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLV

DALSRRIPLVGDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQ

WQDPVRGTVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPV

TFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWD

ENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVAGRGMXGG

MGRSTRAPGANAAVRPLPALKENVKRVMFY

>Mung\_bean XP\_014512211.1:190-1059 protein argonaute 1 [Vigna radiata var. radiata]

PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD

GRKSLYTAGPLPFISKEFRITLIDDDEGAAGGQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEAL

QVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLP

VIDFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTM

KSVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVER

ERDIMQTVYHNAYHEDPYAQEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVN

GGTVNHWFCINFSRSVQDSVARGFCYELAQMCYISGMAFNPEPVVPPVSARPDQVEKVLKTRYHDAKNKL

QGRELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLV

DALSRRIPLVGDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQ

WQDPVRGTVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPV

TFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWD

ENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVAGRGMGGG

IGRSTRAPGANAAVRPLPALKENVKRVMFY

>Pigeon\_pea KYP61687.1:183-1053 Protein argonaute [Cajanus cajan]

PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRESHLGKRLPAYD

GRKSLYTAGPLPFLSKEFRITLVDDDEGAGGQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGDGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IDFVTQLLNRDVSTRPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE

RDIMQTVHHNAYHDDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNNWFCVNFSRNVQDTVARGFCYELAQMCYISGMAFNPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ

GRELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD

ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQW

QDPVRGTVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVAGRGMGGGG

GLGRSTRAPGASAAVRPLPALKENVKRVMFY

>Jequirity\_bean XP\_027359534.1:181-1050 protein argonaute 1 [Abrus precatorius]

PLRPGKGSYGTRCIVKANHFFAELPNKDLHQYDVTITPEVPSRGVNRAVMEQLVRLYRESHLGKRLPAYD

GRKSLYTAGPLPFISKEFRITLIDDDEGAGGQRRDREFKVVIKFAARADLHHLGLFLQGRQTDAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IDFVTQLLNRDVSSRPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYDTYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE

RDIMQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNNWFCINFSRNVQDSVARGFCYELAQMCYISGMAFNPEPVVPPLSARPDQVEKVLKTRYHDAKNKLQ

GRELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD

AISRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQW

QDPVRGTLTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVAGRGMGGGG

AGRSTRAPGASAAVRPLPALKENVKRVMFY

>Soybean XP\_003534084.1:185-1057 protein argonaute 1 [Glycine max]

PLRPGKGSYGTKCVVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD

GRKSLYTAGPLPFMSKEFRIVLADDDEGAGGQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IDFVNQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE

RDIMQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNNWFCINFSRNVQDSVARGFCYELAQMCYISGMAFTPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ

GKELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD

ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQW

QDPVRGTVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASNHHDKSSFDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVAGRGMGGGG

GGGVGRSTRAPGANAAVRPLPALKENVKRVMFY

>Velvet\_bean RDY09546.1:232-1100 Protein argonaute 1, partial [Mucuna pruriens]

PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD

GRKSLYTAGALPFISKEFRITLIDDDEGAGQQRRDREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IDFVTQLLNRDVSSRPLSDSDRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE

RDIMQTVHHNAYHDDPYAKEFGIKISEKLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNNWFCINFSRNVQDSVARGFCYELAQMCYISGMAFNPEPVVPPVSSRPDQVEKVLKTRYHDAKSKLQ

GRDLDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD

ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQW

QDPVRGTVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADGLQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAIGRGMGGGGA

GRSTRAPGANAAVRPLPALKENVKRTSSY

>Chickpea XP\_012567214.1:225-1093 protein argonaute 1 [Cicer arietinum]

PLRPGKGSCGRKCVVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRESHLGKRLPAYD

GRKSLYTAGPLPFISKDFRITLVDDDDGTGGQRRDREFKVVIKLAARADLHHLGLFLEGRQTDAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IDFVTQLLNRDVSTRPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFVIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPLDRE

RDIMQTVHHNAYHEDPYAKEFGIKISEKLAQVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNNWFCVNFSRNVQDSVARGFCCELAQMCYISGMAFNPEPVVPPLSARPDQVDKVLKTRYQDAKNKLQ

GRELDLLIVILPDNNGSLYGDLKRICETELGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVD

ALSRRIPLVSDRPTIIFGADVTHPHPGEDMSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQW

QDPVRGTLTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASDHRDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVSGRGMGGGV

GRSTRAPGANAAVRPLPALKENVKRVMFY

>Red\_clover PNY07943.1:221-1089 protein argonaute 1-like [Trifolium pratense]

PLRPGKGSTGKRCIVKANHFIAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDSHLGKRLPAYD

GRKSLYTAGPLPFISKDFRITLVDEDDGSGSQRRDREFKVVIKLAARADLHHLGLFLEGRQTDAPQEALQ

VLDIVLRELPTSRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IEFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFSIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPLDRE

RDIMQTVHHNAYHDDPYAKEFGIKISDKLAQVEARILPPPWLKYHDTGKEKDCLPQVGQWNMMNKKMVNG

GTVNNWFCVNFSRNVQDSVARGFCSELAHMCYVSGMAFNPEPVVPPLSARPDQVEKVLKTRYHDAKNKLQ

GRELDLLIVILPDNNGSLYGDLKRICETDLGVVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLID

ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLFKQW

QDPARGTLTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVSRGGMGAAA

GRSTRAPGANAAVRPLPALKENVKRVMFY

>Barrelclover KEH26855.1:245-1114 argonaute protein 1A [Medicago truncatula]

PLRPGKGSYGRKTLVKANHFFAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDSHLGKRLPAYD

GRKSLYTAGPLPFISKDFRITLVDEDDGSGGQRRDREFKVVIKLAARADLHHLGLFLEGRQTDAPQEALQ

VLDIVLRELPTSRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IEFVTQLLNRDVSARPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRIAGLTSQATRELTFPVDERGTMK

SVVEYFFETYGFVIQHTQWPCLQVGNPQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPLDRE

RDIMQTVHHNAYHEDPYAKEFGIKISDRLAQVEARILPAPWLKYNDTGREKDCLPQVGQWNMMNKKMFNG

GSVKYWLCVNFSRTVQDSVARGFCYELAQMCYVSGMEFNAEPVVPALTARPDQVEKVLKNRYHDAKSKMP

KDKELDLLIVILPDNNGSLYGDLKRICETDLGVVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLV

DALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQ

WQDPVRGTLTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPV

TFVVVQKRHHTRLFASNHQDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWD

ENNFSADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAVSRGGAGAA

VGRSTRAPGANAAVRPLPALKDNVKKVMFY

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

PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRESHLGKRLPAYD

GRKSLYTAGPLPFVSKEFRIMLIDEDDGTGAQRREREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IEFVTQLLNRDVSSRPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFSETYGFHIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE

RDIMQTVHHNAYDQDPYAKEFGIKISEKLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNHWFCINFSRNVQDSVARTFCYELAQMCQVSGMAFNLDPVVPPVSARPDQVEKVLKTRYYDAKNKLP

GKDLDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVYKMSKQYLANVALKINVKVGGRNTVLVD

ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQW

QDPNRGQVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASNHHDRNSIDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAAAGRGMGAGR

STRAPGASAAVRPLPSLKDNVKRVMFY

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

PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRESHLGKRLPAYD

GRKSLYTAGPLPFVSKEFRIMLIDEDDGTGAQRREREFKVVIKLAARADLHHLGLFLQGRQTDAPQEALQ

VLDIVLRELPTTRYSPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IEFVTQLLNRDVSTRPLSDADRVKIKKALRGIKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFSETYGFHIQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERE

RDIMQTVHHNAYDQDPYAKEFGIKISEKLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNHWFCINFSRNVQDSVARTFCYELAQMCQVSGMAFNLDPVVPPVSARPDQVEKVLKTRYYDAKNKLP

GKDLDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVYKMSKQYLANVALKINVKVGGRNTVLVD

ALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQW

QDPNRGQVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPNYQPPVT

FVVVQKRHHTRLFASNHHDRNSIDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWDE

NNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAAAGRGMGAGR

STRAPGASAAVRPLPSLKDNVKRVMFY

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

PLRPGKGSTGIRCIVKANHFFAELPDKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRESHLGKRLPAYD

GRKSLYTAGPLPFISKEFKITLIDEDDGSGGQRREREFRVVIKLAARADLHHLGLFLQGRQADAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IDFVNQLLNRDVSSRPLSDADRVKIKKALRGVKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFVIQHTQWPCLQVGNQQRPNYLPMEVCKVVEGQRYSKRLNERQITALLKVTCQRPQERE

RDIMQTVHHNAYGNDPYAKEFGIKISEKLASVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNNWICINFSRNVQDSVARGFCYELAQMCYISGMAFNPEPVLPPVSARPEQVEKVLKTRYHDAMTKLQ

QGKELDLLIVILPDNNGSLYGELKRICETDLGLVSQCCLTKHVFRMNKQYLANVALKINVKVGGRNTVLV

DALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLFKE

WQDPVRGRVTGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPPV

TFVVVQKRHHTRLFANNHNDRNAVDKSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLWD

ENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGPVGGRGGMGG

GAGARSTRGPAASAAVRPLPALKENVKRVMFY

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

PLRPGKGSTGTKCMVKANHFFAELPDKDLHQYDVSITPEVTSRGVNRAVMEQLVKLYKETHLGKRLPAYD

GRKSLYTAGPLPFLSKEFKIPLIDEDDGSGGQRREREFKVVIKLAARADLHHLGLFLQGRQADAPQEALQ

VLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IEFVTQLLNRDVTSRPLSDADRVKIKKALRGVKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFAIQHTQWPCLQVGNQQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPQERE

RDIMQTVEHNAYHNDPYAKEFGIKISERLASVEARILPAPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GTVNNWICINFSRNVQDSVARGFCYELAQMCYISGMAFNPEPVLPPLSARPDQVERVLKTRYHDAMTKIQ

PQGKELDLLIVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFRMSKQYLANVALKINVKVGGRNTVL

VDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLFK

TWQDPVRGTVSGGMIKELLISFRRATGQKPKRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPP

VTFVVVQKRHHTRLFANNHHDRNAVDKSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLW

DENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAAAGRGMGG

AGPRSTRPPGANAAVRPLPALKENVKRVMFY

>Peach XP\_007210410.1:201-1068 protein argonaute 1 [Prunus persica]

PLRPGKGSTGIRCTVKANHFFAELPDKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRESHLGKRLPAYD

GRKSLYTAGPLPFLSKEFKIILIDEDDGPGGQRREREFRVVIKFAARADLHHLGLFLQGRQADAPQEALQ

VLDIVLRELPTSRYCPVGRSFYAPDLGRRQSLGEGLESWRGFYQSIRPTQMGLSLNIDMSSTAFIEPLPV

IEFVTQLLNRDVTHRPLSDSDRVKIKKALRGVKVEVTHRGNMRRKYRISGLTSQATRELTFPVDERGTMK

SVVEYFYETYGFVIQHTQWPCLQVGNQQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPHDRE

QDIMRTVRHNAYHEDPYAKEFGIKISENLAQVEARILPPPWLKYHDTGREKDCLPQVGQWNMMNKKMVNG

GKVNNWICINFSRNVQDSVARGFCSELAQMCYISGMAFNPEPVLPPISARPDQVEKVLKTRYHDAMTKLR

VQGKELDLLVVILPDNNGSLYGDLKRICETDLGLVSQCCLTKHVFRMSKQYLANVALKINVKVGGRNTVL

VDALSRRIPLVSDRPTIIFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFK

TWQDPARGTVSGGMIKELLISFRRATGQKPQRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPNYQPP

VTFVVVQKRHHTRLFANNHHDRNTVDRSGNILPGTVVDSKICHPTEFDFYLCSHAGIQGTSRPAHYHVLW

DENKFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARFYMEPETSDSGSMTSGAPGRGGMGA

RSTRAPGANAAVRPLPALKENVKRVMFY

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

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Fruit\_fly PRRPNLGREGRPIVLRANHFQVTMPRGYVHHYDINIQPDKCPRKVNREIIETMVHAYSKI

Barrelclover PLRPGKGSYGRKTLVKANHFFAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDS

Wild\_peanut PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRES

Peanut PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEELVRLYRES

Red\_clover PLRPGKGSTGKRCIVKANHFIAELPKKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRDS

Chickpea PLRPGKGSCGRKCVVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES

Velvet\_bean PLRPGKGSYGTKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES

Jequirity\_bean PLRPGKGSYGTRCIVKANHFFAELPNKDLHQYDVTITPEVPSRGVNRAVMEQLVRLYRES

Pigeon\_pea PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRES

Soybean PLRPGKGSYGTKCVVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES

Tepary\_bean PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES

Mung\_bean PLRPGKGSYGIKCIVKANHFFAELPNKDLHQYDVTITPEVTSRGVNRAVMEQLVRLYRES

Peach PLRPGKGSTGIRCTVKANHFFAELPDKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRES

Castor\_bean PLRPGKGSTGIRCIVKANHFFAELPDKDLHQYDVTITPEVTSRGVNRAVMEQLVKLYRES

Cork\_oak PLRPGKGSTGTKCMVKANHFFAELPDKDLHQYDVSITPEVTSRGVNRAVMEQLVKLYKET

\* \*\*. \* \* :.\*\*\*\* . :\* :\*:\*\*:.\* \*: .\* \*\*\* ::\* :\*. \* .

Fruit\_fly -FGVLKPVFDGRNNLYTRDPLPIGNERLELEVTLPGEG------KDRIFRVTIKWQAQVS

Barrelclover HLGKRLPAYDGRKSLYTAGPLPFISKDFRITLVDEDDG-SGGQRRDREFKVVIKLAARAD

Wild\_peanut HLGKRLPAYDGRKSLYTAGPLPFVSKEFRIMLIDEDDG-TGAQRREREFKVVIKLAARAD

Peanut HLGKRLPAYDGRKSLYTAGPLPFVSKEFRIMLIDEDDG-TGAQRREREFKVVIKLAARAD

Red\_clover HLGKRLPAYDGRKSLYTAGPLPFISKDFRITLVDEDDG-SGSQRRDREFKVVIKLAARAD

Chickpea HLGKRLPAYDGRKSLYTAGPLPFISKDFRITLVDDDDG-TGGQRRDREFKVVIKLAARAD

Velvet\_bean HLGKRLPAYDGRKSLYTAGALPFISKEFRITLIDDDEG-AGQQRRDREFKVVIKLAARAD

Jequirity\_bean HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDEG-AGGQRRDREFKVVIKFAARAD

Pigeon\_pea HLGKRLPAYDGRKSLYTAGPLPFLSKEFRITLVDDDEG-AGGQRRDREFKVVIKLAARAD

Soybean HLGKRLPAYDGRKSLYTAGPLPFMSKEFRIVLADDDEG-AGGQRRDREFKVVIKLAARAD

Tepary\_bean HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDEGAAGGQRRDREFKVVIKLAARAD

Mung\_bean HLGKRLPAYDGRKSLYTAGPLPFISKEFRITLIDDDEGAAGGQRRDREFKVVIKLAARAD

Peach HLGKRLPAYDGRKSLYTAGPLPFLSKEFKIILIDEDDG-PGGQRREREFRVVIKFAARAD

Castor\_bean HLGKRLPAYDGRKSLYTAGPLPFISKEFKITLIDEDDG-SGGQRREREFRVVIKLAARAD

Cork\_oak HLGKRLPAYDGRKSLYTAGPLPFLSKEFKIPLIDEDDG-SGGQRREREFKVVIKLAARAD

:\* \*.:\*\*\*:.\*\*\* ..\*\*: .: : : : .:\* .:\* \*.\*.\*\* \*...

Fruit\_fly LFNLEEALEGRTRQIPYDAILALDVVMRHLPSMTYTPVGRSFFSSPEGYYHPLGGGREVW

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 LHHLGLFLQGKQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW

Mung\_bean LHHLGLFLQGRQTDAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW

Peach LHHLGLFLQGRQADAPQEALQVLDIVLRELPTSRYCPVGRSFYAPDLGRRQSLGEGLESW

Castor\_bean LHHLGLFLQGRQADAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW

Cork\_oak LHHLGLFLQGRQADAPQEALQVLDIVLRELPTTRYCPVGRSFYSPDLGRRQPLGEGLESW

\*.:\* \*:\*. : \* :\*: .\*\*:\*:\* \*\*: \* \*\*\*\*\*\*::. \* :.\*\* \* \* \*

Fruit\_fly FGFHQSVRPSQWKMMLNIDVSATAFYKAQPVIDFMCEVLDIRDINEQRKPLTDSQRVKFT

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 KEIKGLKIEITHCGQMRRKYRVCNVTRRPAQMQSFPLQLENGQTVECTVAKYFLDKYRMK

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--SVVEYFYDTYGFV

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 IQHTQWPCLQVGNPQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPLDRERDI

Wild\_peanut IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Peanut IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Red\_clover IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPLDRERDI

Chickpea IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPLDRERDI

Velvet\_bean IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Jequirity\_bean IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Pigeon\_pea IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Soybean IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Tepary\_bean IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Mung\_bean IQHTQWPCLQVGNTQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPVERERDI

Peach IQHTQWPCLQVGNQQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPHDREQDI

Castor\_bean IQHTQWPCLQVGNQQRPNYLPMEVCKVVEGQRYSKRLNERQITALLKVTCQRPQERERDI

Cork\_oak IQHTQWPCLQVGNQQRPNYLPMEVCKIVEGQRYSKRLNERQITALLKVTCQRPQERERDI

:.:.: \*\*\*\*\*\*: :. .\*\*\*:\*\*\*::\* \*\*\* \*.\*.: \* ::::\*.\*.. . :\*\*.:\*

Fruit\_fly NNLVKRADFNNDSYVQEFGLTISNSMMEVRGRVLPPPKLQYGGRVSTGLTGQQLFPPQNK

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 MQTVEHNAYHNDPYAKEFGIKISERLASVEARILPAPWLKYH---DTGREKDCL------

. \* . : :\*.\*.:\*\*\*:.\*\*: : .\* .\*:\*\*.\*.\*:\* .\*\* : \*

Fruit\_fly VSLASPNQGVWDMRGKQFFTGVEIRIWAIACFAPQRTVREDALRNFTQQLQKISNDAGMP

Barrelclover -----PQVGQWNMMNKKMFNGGSVKYW--LCVNFSRTVQDSVARGFCYELAQMCYVSGME

Wild\_peanut -----PQVGQWNMMNKKMVNGGTVNHW--FCINFSRNVQDSVARTFCYELAQMCQVSGMA

Peanut -----PQVGQWNMMNKKMVNGGTVNHW--FCINFSRNVQDSVARTFCYELAQMCQVSGMA

Red\_clover -----PQVGQWNMMNKKMVNGGTVNNW--FCVNFSRNVQDSVARGFCSELAHMCYVSGMA

Chickpea -----PQVGQWNMMNKKMVNGGTVNNW--FCVNFSRNVQDSVARGFCCELAQMCYISGMA

Velvet\_bean -----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRNVQDSVARGFCYELAQMCYISGMA

Jequirity\_bean -----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRNVQDSVARGFCYELAQMCYISGMA

Pigeon\_pea -----PQVGQWNMMNKKMVNGGTVNNW--FCVNFSRNVQDTVARGFCYELAQMCYISGMA

Soybean -----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRNVQDSVARGFCYELAQMCYISGMA

Tepary\_bean -----PQVGQWNMMNKKMVNGGTVNNW--FCINFSRSVQDSVARGFCYELAQMCYISGMA

Mung\_bean -----PQVGQWNMMNKKMVNGGTVNHW--FCINFSRSVQDSVARGFCYELAQMCYISGMA

Peach -----PQVGQWNMMNKKMVNGGKVNNW--ICINFSRNVQDSVARGFCSELAQMCYISGMA

Castor\_bean -----PQVGQWNMMNKKMVNGGTVNNW--ICINFSRNVQDSVARGFCYELAQMCYISGMA

Cork\_oak -----PQVGQWNMMNKKMVNGGTVNNW--ICINFSRNVQDSVARGFCYELAQMCYISGMA

\*: \* \*:\* .\*::..\* :. \* \*. .\*.\*.: . \* \* :\* ::. :\*\*

Fruit\_fly IIGQPCFCKYATGPDQVEPMF--RY----LKIT--FPGLQLVVVVLPGKT-PVYAEVKRV

Barrelclover FNAEPVVPALTARPDQVEKVLKNRYHDAKSKMP-KDKELDLLIVILPDNNGSLYGDLKRI

Wild\_peanut FNLDPVVPPVSARPDQVEKVLKTRYYDAKNKLP--GKDLDLLIVILPDNNGSLYGDLKRI

Peanut FNLDPVVPPVSARPDQVEKVLKTRYYDAKNKLP--GKDLDLLIVILPDNNGSLYGDLKRI

Red\_clover FNPEPVVPPLSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI

Chickpea FNPEPVVPPLSARPDQVDKVLKTRYQDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI

Velvet\_bean FNPEPVVPPVSSRPDQVEKVLKTRYHDAKSKLQ--GRDLDLLIVILPDNNGSLYGDLKRI

Jequirity\_bean FNPEPVVPPLSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI

Pigeon\_pea FNPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI

Soybean FTPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ--GKELDLLIVILPDNNGSLYGDLKRI

Tepary\_bean FNPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI

Mung\_bean FNPEPVVPPVSARPDQVEKVLKTRYHDAKNKLQ--GRELDLLIVILPDNNGSLYGDLKRI

Peach FNPEPVLPPISARPDQVEKVLKTRYHDAMTKLRVQGKELDLLVVILPDNNGSLYGDLKRI

Castor\_bean FNPEPVLPPVSARPEQVEKVLKTRYHDAMTKLQ-QGKELDLLIVILPDNNGSLYGELKRI

Cork\_oak FNPEPVLPPLSARPDQVERVLKTRYHDAMTKIQPQGKELDLLIVILPDNNGSLYGDLKRI

: :\* . :: \*:\*\*: :: \*\* \*: \*:\*::\*:\*\*.:. .:\*.::\*\*:

Fruit\_fly GDTVLGMATQCVQAKNVNKTSPQTLSNLCLKINVKLGGINSILVPSIR---PKVFNEPVI

Barrelclover CETDLGVVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Wild\_peanut CETDLGLVSQCCLTKHVYKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Peanut CETDLGLVSQCCLTKHVYKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Red\_clover CETDLGVVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLIDALSRRIPLVSDRPTI

Chickpea CETELGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Velvet\_bean CETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Jequirity\_bean CETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDAISRRIPLVSDRPTI

Pigeon\_pea CETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Soybean CETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Tepary\_bean CETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVGDRPTI

Mung\_bean CETDLGLVSQCCLTKHVFKMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVGDRPTI

Peach CETDLGLVSQCCLTKHVFRMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Castor\_bean CETDLGLVSQCCLTKHVFRMNKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

Cork\_oak CETDLGLVSQCCLTKHVFRMSKQYLANVALKINVKVGGRNTVLVDALSRRIPLVSDRPTI

:\* \*\*:.:\*\* :\*:\* . . \* \*:\*:.\*\*\*\*\*\*:\*\* \*::\*: :: \* \* : \*.\*

Fruit\_fly FLGADVTHPPAGDNKKPSIAAVVGSMD-AHPSRYAATVRVQQHRQEIIQEL---------

Barrelclover IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Wild\_peanut IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQWQDPNR

Peanut IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQWQDPNR

Red\_clover IFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLFKQWQDPAR

Chickpea IFGADVTHPHPGEDMSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Velvet\_bean IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Jequirity\_bean IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Pigeon\_pea IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Soybean IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Tepary\_bean IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Mung\_bean IFGADVTHPHPGEDSSPSIAAVVASQDYPEITKYAGLVCAQAHRQELIQDLFKQWQDPVR

Peach IFGADVTHPHPGEDSSPSIAAVVASQDWPEITKYAGLVCAQAHRQELIQDLFKTWQDPAR

Castor\_bean IFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLFKEWQDPVR

Cork\_oak IFGADVTHPHPGEDSSPSIAAVVASQDWPEVTKYAGLVCAQAHRQELIQDLFKTWQDPVR

::\*\*\*\*\*\*\* .\*:: .\*\*\*\*\*\*\*.\* \* . :.\*\*. \* .\* \*\*\*\*:\*\*:\*

Fruit\_fly ----SSMVRELLIMFYKSTGGYKPHRIILYRDGVSEGQFPHVLQHELTAIREACIKLEPE

Barrelclover GTLTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Wild\_peanut GQVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Peanut GQVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Red\_clover GTLTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Chickpea GTLTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Velvet\_bean GTVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Jequirity\_bean GTLTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Pigeon\_pea GTVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Soybean GTVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Tepary\_bean GTVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Mung\_bean GTVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLFELDAIRKACASLEPN

Peach GTVSGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPN

Castor\_bean GRVTGGMIKELLISFRRATGQ-KPQRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPN

Cork\_oak GTVSGGMIKELLISFRRATGQ-KPKRIIFYRDGVSEGQFYQVLLYELDAIRKACASLEPN

..\*:.\*\*\*\* \* .:\*\* \*\*:\*\*\*:\*\*\*\*\*\*\*\*\*\* :\*\* .\*\* \*\*\*:\*\* .\*\*\*:

Fruit\_fly YRPGITFIVVQKRHHTRLFCAEKKEQSG--KSGNIPAGTTVDVGITHPTEFDFYLCSHQG

Barrelclover YQPPVTFVVVQKRHHTRLFASNHQDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Wild\_peanut YQPPVTFVVVQKRHHTRLFASNHHDRNSIDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Peanut YQPPVTFVVVQKRHHTRLFASNHHDRNSIDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Red\_clover YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Chickpea YQPPVTFVVVQKRHHTRLFASDHRDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Velvet\_bean YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Jequirity\_bean YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Pigeon\_pea YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Soybean YQPPVTFVVVQKRHHTRLFASNHHDKSSFDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Tepary\_bean YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Mung\_bean YQPPVTFVVVQKRHHTRLFASNHHDKSSVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Peach YQPPVTFVVVQKRHHTRLFANNHHDRNTVDRSGNILPGTVVDSKICHPTEFDFYLCSHAG

Castor\_bean YQPPVTFVVVQKRHHTRLFANNHNDRNAVDKSGNILPGTVVDSKICHPTEFDFYLCSHAG

Cork\_oak YQPPVTFVVVQKRHHTRLFANNHHDRNAVDKSGNILPGTVVDSKICHPTEFDFYLCSHAG

\*.\* :\*\*:\*\*\*\*\*\*\*\*\*\*\*. ::.:.. .\*\*\*\* .\*\*.\*\* \* \*\*\*\*\*\*\*\*\*\*\*\* \*

Fruit\_fly IQGTSRPSHYHVLWDDNHFDSDELQCLTYQLCHTYVRCTRSVSIPAPAYYAHLVAFRARY

Barrelclover IQGTSRPAHYHVLWDENNFSADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Wild\_peanut IQGTSRPAHYHVLWDENNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Peanut IQGTSRPAHYHVLWDENNFTADALQMLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Red\_clover IQGTSRPAHYHVLWDENNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Chickpea IQGTSRPAHYHVLWDENNFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Velvet\_bean IQGTSRPAHYHVLWDENNFTADGLQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Jequirity\_bean IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Pigeon\_pea IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Soybean IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Tepary\_bean IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Mung\_bean IQGTSRPAHYHVLWDENNFTADALQTLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Peach IQGTSRPAHYHVLWDENKFTADALQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Castor\_bean IQGTSRPAHYHVLWDENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

Cork\_oak IQGTSRPAHYHVLWDENKFTADGLQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRARF

\*\*\*\*\*\*\*:\*\*\*\*\*\*\*:\*:\* :\* \*\* \*\* :\*\*:\*\*.\*\*\*\*\*\*\*\* .\*\*\*\*\*\*\*.\*\*\*\*\*:

Fruit\_fly HLVEKEHDSGE--------------GSHQSGCSEDRTPGAMA--RAI-TVHADTKKVMYF

Barrelclover YMEPETSDSGSMTSGAVSRGG----AGAAVG-RSTRAPGANAAVRPLPALKDNVKKVMFY

Wild\_peanut YMEPETSDSGSMTSGAAAGRGM------GAG-RSTRAPGASAAVRPLPSLKDNVKRVMFY

Peanut YMEPETSDSGSMTSGAAAGRGM------GAG-RSTRAPGASAAVRPLPSLKDNVKRVMFY

Red\_clover YMEPETSDSGSMTSGAVSRGGM----GAAAG-RSTRAPGANAAVRPLPALKENVKRVMFY

Chickpea YMEPETSDSGSMTSGAVSGRGM----GGGVG-RSTRAPGANAAVRPLPALKENVKRVMFY

Velvet\_bean YMEPETSDSGSMTSGAI-GRGM---GGGGAG-RSTRAPGANAAVRPLPALKENVKRTSSY

Jequirity\_bean YMEPETSDSGSMTSGAVAGRGM---GGGGAG-RSTRAPGASAAVRPLPALKENVKRVMFY

Pigeon\_pea YMEPETSDSGSMTSGAVAGRGM--GGGGGLG-RSTRAPGASAAVRPLPALKENVKRVMFY

Soybean YMEPETSDSGSMTSGAVAGRGMGGGGGGGVG-RSTRAPGANAAVRPLPALKENVKRVMFY

Tepary\_bean YMEPETSDSGSMTSGAVAGRGM----XGGMG-RSTRAPGANAAVRPLPALKENVKRVMFY

Mung\_bean YMEPETSDSGSMTSGAVAGRGM----GGGIG-RSTRAPGANAAVRPLPALKENVKRVMFY

Peach YMEPETSDSGSMTSGAP--------GRGGMGARSTRAPGANAAVRPLPALKENVKRVMFY

Castor\_bean YMEPETSDSGSMTSGPVGGRGG---MGGGAGARSTRGPAASAAVRPLPALKENVKRVMFY

Cork\_oak YMEPETSDSGSMTSGAAAGRGM-----GGAGPRSTRPPGANAAVRPLPALKENVKRVMFY

:: : \*\*\*. \* . \* \*.\* \* \*.: ::: :.\*.. :

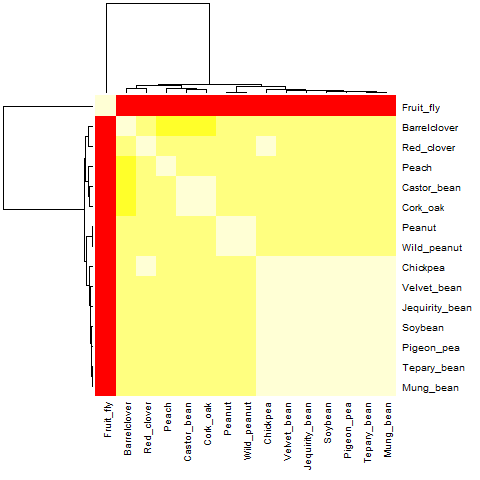
[Q6] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use “simple phylogeny” online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.

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.



[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).

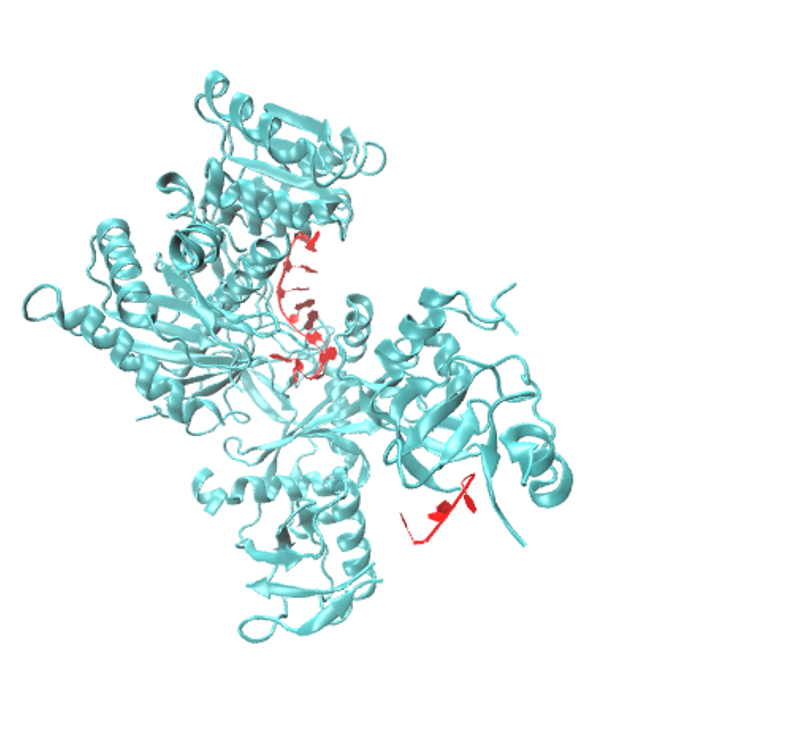
HINT: You can use a single sequence from your alignment or generate a consensus sequence from your alignment using the Bio3D function consensus(). The Bio3D functions blast.pdb(), plot.blast() and pdb.annotate() are likely to be of most relevance for completing this task. Note that the results of blast.pdb() 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 | *Homo sapiens* | 0 | 44.549 |
| 3VNA | X-RAY DIFFRACTION | 2 | *Arabidopsis thaliana* | 7.67E-84 | 84.459 |
| 5THE | X-RAY DIFFRACTION | 2.1 | *Vanderwaltozyma polyspora* | 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 ChEMBEL (<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>