```
Project: find-a-Gene BGGN-213: FOUNDATIONS OF BIOINFORMATICS
Xuerui Huang (A91018769)
Q1
  Protein Name
  Accession
  Species
  Function known
Q2
  Blast Method
  Database
  Limits applied
  Search output list
  Alignment of choice
  Alignment details (Alignment of choice)
Q3
  Protein sequence
  Name
  Species
Q4
  Program run
  Result and Top alignment
Q5
  Sequence Collection
  Multiple sequence alignment by MUSCLE (version 3.8) at EBI
Q6
  Sequence clustering pattern
Q7
  Heatmap figure with 6 samples
Q8
  3 Unique hits info
Q9
  Structure Info
  Structure figure
Q9
  Link to ChEMBL
  ChEMBL Info
  Binding assay
```

Project: find-a-Gene BGGN-213: FOUNDATIONS OF BIOINFORMATICS

Xuerui Huang (A91018769)

Q₁

Protein Name

retinol-binding protein 4 isoform a precursor (RBP4)

Accession

NP 006735

Species

Homo sapiens

Function known

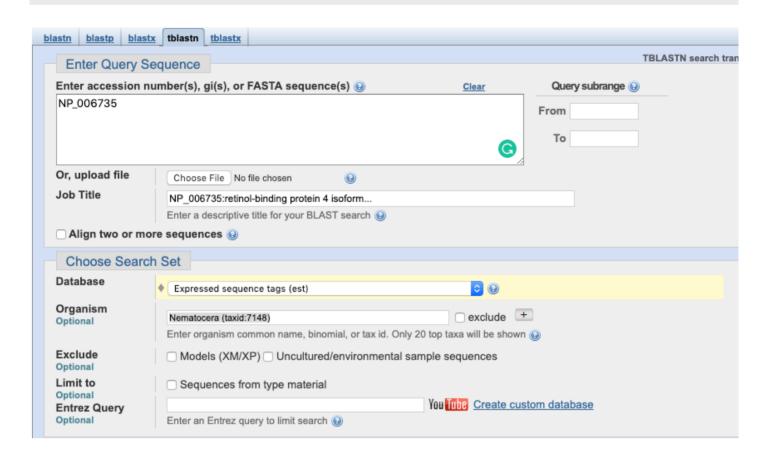
This protein belongs to the lipocalin family and is the specific carrier for retinol (vitamin A) in the blood. It delivers retinol from the liver stores to the peripheral tissues. Plasma concentration of Retinol Binding Protein 4 (RBP4) in relation to nutritional status and kidney function in older population of PolSeniorStudy

Q2

Blast Method

BLASTN 2.9.0+ search against Nematocera ESTs

BLAST ** * tblastn



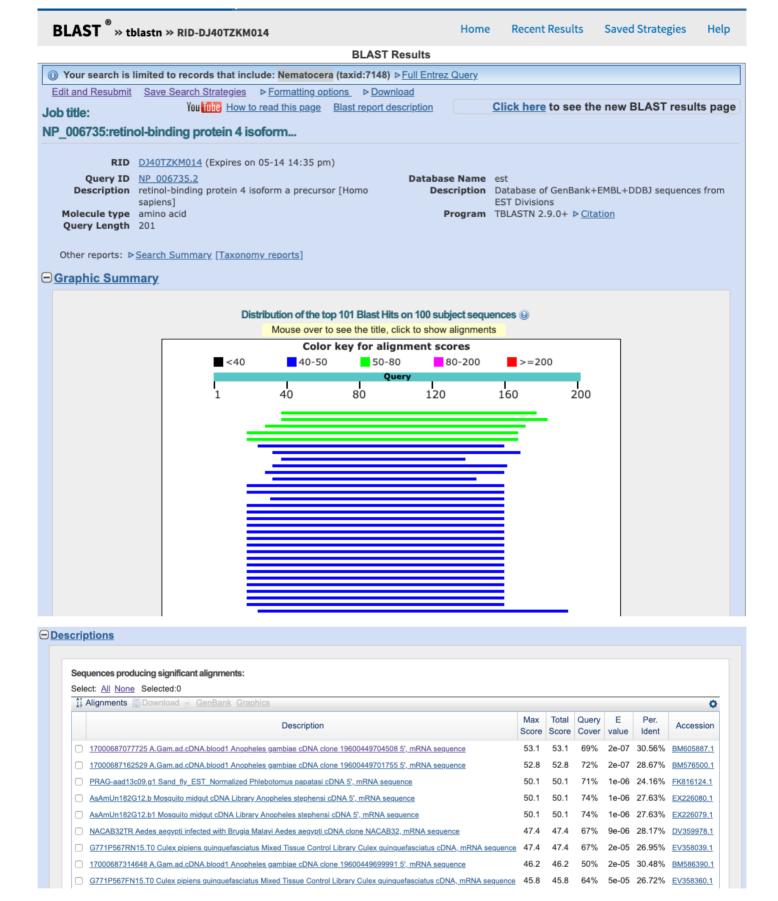
Database

Expressed Sequence Tags (est)

Limits applied

Organism: Nematocera (taxid:7148)

Search output list



Alignment of choice

Chosen match: Accesion BM605887.1, a 679 base pair clone from Anopheles gambiae

Alignment details (Alignment of choice)

score: 53.1 Evalue: 2e-07 Identities: 43/150(29%) Positives: 74/150(49%) Gaps: 15/150(10%)

Frame: +1

```
1 17000687077725 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone 19600449704508
   5', mRNA sequence
2 Sequence ID: BM605887.1Length: 679Number of Matches: 1
3 Related Information
4 UniGene-clustered expressed sequence tags
5 Range 1: 186 to 590GenBankGraphicsNext MatchPrevious Match
6 Alignment statistics for match #1
7 Score
          Expect Method Identities Positives
                                                   Gaps
8 Frame
9 53.1 bits(126) 2e-07 Compositional matrix adjust.
                                                         44/144(31%) 71/144(49%)
   13/144(9%)
10 +3
11 Query 38 FSGTWYAMAKKDPEGLFLQDNIVAEFSVDETGQMSATAK-GRVRLLNNWDVCADMVGTFT 96
12
               ++G WY + + + D +VAE+ +TG S + K G L NN V AD +
13 Sbjct 186 YAGRWYEIKRYEQFYEKDLDCVVAEYQ--KTGDNSISVKNGAFSLANNTRVVADGTAVVS 359
14
15 Query 97 ---DTEDPAKFKMKYWGVASFLQKGNDDHWIVDTDYDTYAVQYSCRLLNLDGTCADSYSF 153
                  DT PAK + ++G + ++W++DTDY ++AV +SC D + +
16
17
  Sbjct 360 YPDDTTHPAKLSVAFFGA----KADRSNYWVLDTDYTSFAVVWSCEPFFRDPSKNVLGFW 527
18
19 Query 154 VFSRDPNGLPPEAQKIVRQRQEEL 177
20
              +FSR+P EA V +R +EL
21 Sbjct 528 IFSRNPTFPTDEA---VVKRVDEL 590
```

■Download ∨ GenBank Graphics

17000687077725 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone 19600449704508 5', mRNA sequence Sequence ID: BM605887.1 Length: 679 Number of Matches: 1

Range 1: 186	to 590 GenBank Graphics						
Score	Expect Method	Identities	Positives	Gaps	Frame		
53.1 bits(12	2e-07 Compositional matrix	adjust. 44/144(31%)	71/144(49%)	13/144(9%)	+3		
Query 38	FSGTWYAMAKKDPEGLFLQDNIVAE		LLNNWDVCADMVG L NN V AD	TFT 96			
Sbjct 186				VVS 359			
Query 97	DTEDPAKFKMKYWGVASFLQKG	NDDHWIVDTDYDTYAVQYS		YSF 153			
Sbjct 360	YPDDTTHPAKLSVAFFGAKAD	RSNYWVLDTDYTSFAVVWS	CEPFFRDPSKNVL	GFW 527			
Query 154	VFSRDPNGLPPEAQKIVRQRQEEL +FSR+P EA V +R +EL	177					
Sbjct 528	IFSRNPTFPTDEAVVKRVDEL	590					

Q3

Protein sequence

- 1 > A. Apolipoprotein D precursor (sequence taken from BLAST result)
- 2 YAGRWYEIKRYEQFYEKDLDCVVAEYQKTGDNSISVKNGAFSLANNTRVVADGTAVVSYPDDTTHPAKLSVAFFGAKADRS NYWVLDTDYTSFAVVWSCEPFFRDPSKNVLGFWIFSRNPTFPTDEAVVKRVDEL

Name

Name: Apolipoprotein D precursor

Species

Anopheles gambiae (African malaria mosquito)

Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;

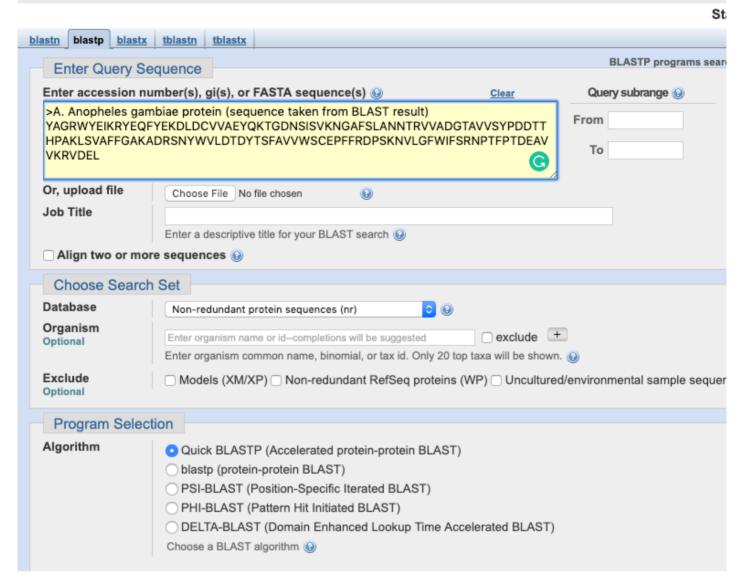
Pterygota; Neoptera; Holometabola; Diptera;

Nematocera; Culicoidea; Culicidae; Anophelinae; Anopheles.

Q4

Program run



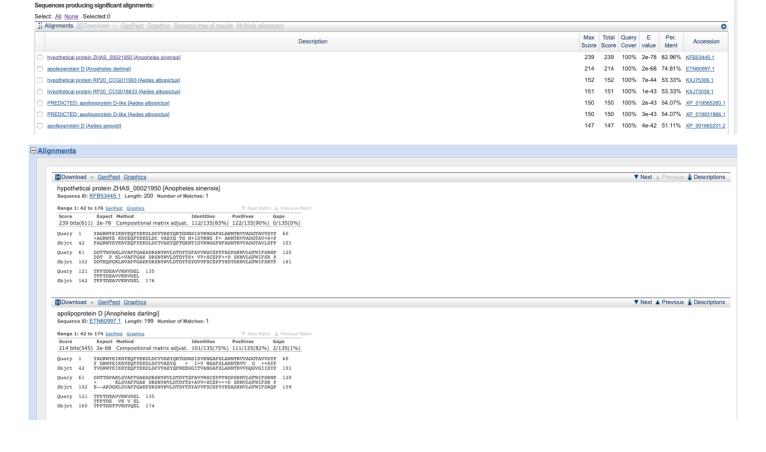


Result and Top alignment

Evalue: 2e-78

Per.Identity: 82.96%

The top result is to a protein from Anopheles sinensis, see second screen shot below for alignment details:



Sequence Collection

Collected in total of 6 sequences with different species

1 >Human_RBP4 gi|30497399|ref|NP_006735| retinol-binding protein 4 isoform a precursor [Homo sapiens] MKWVWALLLLAALGSGRAERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEGLFLQDNIVAEFSVDETGQ MSATAKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVASFL0KGNDDHWIVDTDYDTYAV0YSCRL LNLDGTCADSYSFVFSRDPNGLPPEA0KIVR0R0EELCLAR0YRLIVHNGYCDGRSERNLL 5 6 > Apolipoprotein_D_precursor (sequence taken from BLAST result) [Anopheles sinensis] YAGRWYEIKRYE0FYEKDLDCVVAEY0KTGDNSISVKNGAFSLANNTRVVADGTAVVSYPDDTTHPAKLS VAFFGAKADRSNYWVLDTDYTSFAVVWSCEPFFRDPSKNVLGFWIFSRNPTFPTDEAVVKRVDEL 10 >Riverine_mosquito gi|20920257|Accession|ETN60997|apolipoprotein D [Anopheles darling 11 MHSSLLAIVLLATVAATVTGLLTDSACPEGVKVKONFSLKDYVGRWYEIKRYE0FYEKDLDCVVAEYOPR EEGGITVANGAFSLANNTRVVGOGVGIISYPEAKDGKLSVAFFGAKPDRSNYWVLDTDYTSYAVVFSCEP YYKDASKNVLGFWIFSRQPTFPTDETTVKHVQELVKQYGDESKFEITNQSDERCPRSYV 13 14 15 >Southern_house_mosquito | Accession|XP_001866789.1| apolipoprotein D [Culex quinquefasciatus] 16 MYKSPOVFAAALLVLGLSSAVVYGVIYDRPCRTEISVVONFCLDRYLGKWYELORYEOPFOTKVDCTTAN 17 YGLLDSATVSVRNSAFSLINGTSSEAIGTAVLSFPEQEIVQAKLNVSFFGAPNDRSNYWVIDTDYENFAI

```
VWSCEPLPQDQSSEGFWFLSRERKFTDDKDANERAFGAIRKYIDQSEIRFTNQADERCPDF

VWSCEPLPQDQSSEGFWFLSRERKFTDDKDANERAFGAIRKYIDQSEIRFTNQADERCPDF

Atlantic_salmon Accession|XP_014045104.1|PREDICTED: apolipoprotein D-like [Salmo salar]

MQALHVLSLTLLSILTANAQTLRPGKCPQPPVQANFDAARYLGKWYEIKKLPVVFQKGECTTATYSLESP

GVVGVLNRELLADNTVSVITGYAKAKDPSEPAKLEVTFFEDSPPGNYWVLSTDYEGHSVVYSCTDILGTF

HADFAWILSRESTLSEEKLEELYNVFTSNGIDIDGMTVTNQSQELCADMPLWA

Chicken Accession|NP_990569.1|retinol-binding protein 4 precursor [Gallus gallus]

MAYTWRALLLLALAFLGSSMAERDCRVSSFKVKENFDKNRYSGTWYAMAKKDPEGLFLQDNVVAQFTVDE

NGQMSATAKGRVRLFNNWDVCADMIGSFTDTEDPAKFKMKYWGVASFLQKGNDDHWVVDTDYDTYALHYS

CRELNEDGTCADSYSFVFSRDPKGLPPEAQKIVRQRQIDLCLDRKYRVIVHNGFCS
```

Multiple sequence alignment by MUSCLE (version 3.8) at EBI

Result

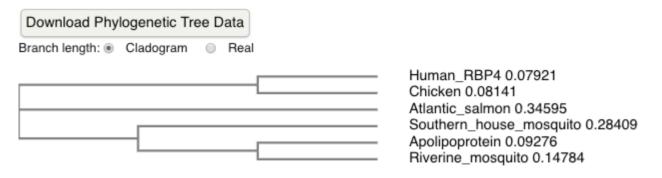
```
----MAYTWRALLLLALAFLGSS-
1 Chicken
  MAERDCRVSSFKVKENFDKNRYSGTWYAMAKKDPE
2 Human RBP4
                                ----MKWVWALLLLAALGSGR----
   AERDCRVSSFRVKENFDKARFSGTWYAMAKKDPE
3 Atlantic_salmon
                                -MQALHVLSLTLLSILTANAQT---LRPGKCP-QP-
   PVQANFDAARYLGKWYEIKKLP--
4 Southern_house_mosquito
                                MYKSPQVFAAALLVLGLSSAVVYGVIYDRPCR-
  TEISVVQNFCLDRYLGKWYELQRYE--
5 Apolipoprotein
  YAGRWYEIKRYE--
                                ----MHSSLLAIVLLATVAATVTGLLTDSACP-
6 Riverine_mosquito
   EGVKVKONFSLKDYVGRWYEIKRYE--
                                                                                ! * **
   : .
9 Chicken
                                GLFLQ---DNVVAQFTVDENGQMSATAKGRVRLFNNWDVCADMIGSFTDT--
   -EDPAKFK
10 Human_RBP4
                                GLFLQ---DNIVAEFSVDETGOMSATAKGRVRLLNNWDVCADMVGTFTDT--
   -EDPAKFK
11 Atlantic_salmon
                                -VVFQKG-ECTTATYSLESPGVVGV--LNRELLADNTVSVITGYAKAKDP--
   -SEPAKLE
12 | Southern_house_mosquito
                                -QPFQTKVDCTTANYGLLDSATVSV-
   RNSAFSLINGTSSEAIGTAVLSFPEQEIVQAKLN
13 Apolipoprotein
                                -QFYEKDLDCVVAEYQKTGDNSISV-
   KNGAFSLANNTRVVADGTAVVSYPDDTTHPAKLS
14 Riverine_mosquito
                                -QFYEKDLDCVVAEYQPREEGGITV-ANGAFSLANNTRVVGQGVGIISYPE-
   -AKDGKLS
15
      .*:.
```

16							
16							
17	Chicken						
	MKYWGVASFLQKGNDDHWVVDTDYDTYALHYSCRELNEDGTCADSYSFVFSRDPKGLPPE						
18	Human_RBP4						
	${\tt MKYWGVASFLQKGNDDHWIVDTDYDTYA}$	VQYSCRLLNLDGTCADSYSFVFSRDPNGLPPE					
19	Atlantic_salmon	VTFFEDSPPGNYWVLSTDYEGHSVVYSCTDIL					
	GTFHADFAWILSREST-LSEE						
20	Southern_house_mosquito	VSFFGAPNDRSNYWVIDTDYENFAIVWSCEPLPQDQS					
	SEGFWFLSRERK-FTDD						
21	Apolipoprotein	VAFFGAKA					
	DRSNYWVLDTDYTSFAVVWSCEPFFRDP	SKNVLGFWIFSRNPT-FPTD					
22	Riverine_mosquito	VAFFGAKP					
	DRSNYWVLDTDYTSYAVVFSCEPYYKDASKNVLGFWIFSRQPT-FPTD						
23		: ::					
	:::**: . :. :						
24							
25	Chicken	AQKIVRQRQIDLCLDRKYRVIVHNGFCS					
26	Human_RBP4	AQKIVRQRQEELCLARQYRLIVHNGYCDGRSERNLL					
27	Atlantic_salmon	KLEELYNVFTSNGIDIDGMTVTNQSQELCADMPLWA					
28	Southern_house_mosquito	KDANERAFGAIRKYI-DQSEIRFTNQADERCPDF					
29	Apolipoprotein	EAVVKRVDEL					
30	Riverine_mosquito	ETTVKHVQELVKQYG-DESKFEITNQSDERCPRSYV					
31							

Sequence clustering pattern

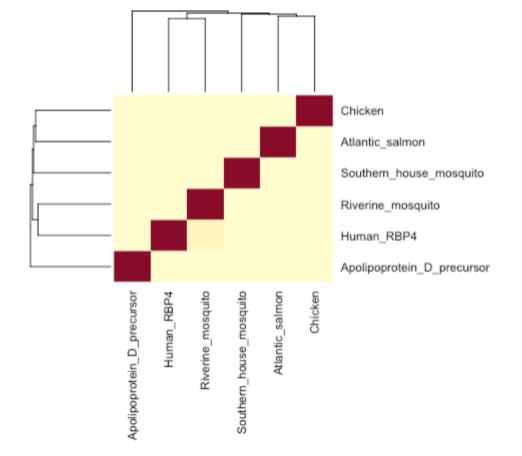
Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.



Q7

Heatmap figure with 6 samples



3 Unique hits info

Species: Sander vitreus, Homo sapiens, Escherichia coli

PDB_ID	Evalue	Identity	Technique	Resolution	Source
5F6Z	4.18E-15	41.026	X-RAY DIFFRACTION	2.25	Sander vitr
2HZQ	6.03E-12	32.558	X-RAY DIFFRACTION	1.8	Homo sapi
1QWD	1.91E-12	29.944	X-RAY DIFFRACTION	1.75	Escherichia

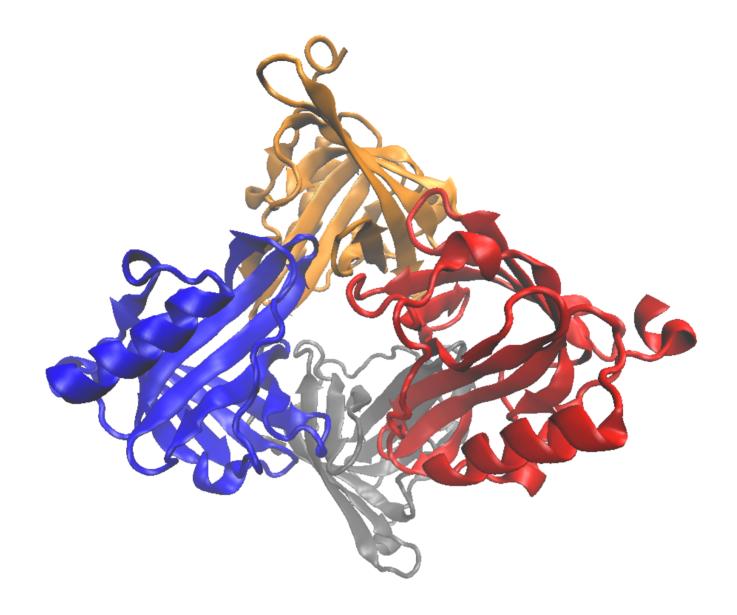
Q9

Structure Info

This structure is likely to be similar in structure to the Apolipoprotein D precursor given the medium sequence similarity (>40%).

In the figure below the Sandercyanin Fluorescent Protein chain A is colored blue and corresponds to the Apolipoprotein D precursor subject of this report.

Structure figure



Link to ChEMBL

https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL2364702/

ChEMBL Info

CHEMBL details 1 binding Assay (CHEMBL2364702) and 15 functional assays. No information of ligand efficiency and Associated Compounds.

Binding assay

Molecule ChEMBL \$ ID	Molecule 🚔 Name	Molecule 🕏	Max Phase	First \$\\\phi\$ Approval	USAN Stem	Mechanism $_{\hat{\oplus}}$ Of Action	Target ChEMBL ≑ ID	Target Name [⊕]	Mechanism $_{\stackrel{\triangle}{=}}$	Target Type [‡]	Target Organism
CHEMBL1201131	PIPERONYL BUTOXIDE	Small molecule	4	2000	No Data	Cytochrome P450 inhibitor	CHEMBL2364702	Cytochrome P450	INHIBITOR	PROTEIN FAMILY	Pediculus humanus capitis

An analogue of piperonyl butoxide facilitates the characterisation of metabolic resistance

Using examples of resistant insects with well-characterised resistance mechanisms, a combination of PBO and analogue allows identification

of the metabolic mechanism responsible for conferring resistance. The relative potency of PBO as both an esterase inhibitor and an oxidase

inhibitor is also discussed. Copyright © 2008 Society of Chemical Industry

Cite: Moores, Graham D., et al. "An analogue of piperonyl butoxide facilitates the characterisation of metabolic resistance." Pest Management Science: formerly Pesticide Science 65.2 (2009): 150-154. http://europepmc.org/abstract/MED/18951417