ClustalW

1. What is ClustalW ?

   Is a tool to _____ two or more sequences together

2. True or False:

   - Aligning multiple sequences with ClustalW highlights areas of similarity that may be associated with specific features and these regions in turn can help classify sequences.

   - ClustalW accepts nucleic acid or potein sequences in fasta format.

HMMER

Go to the tutorials page www.vectorbase.org/tutorials and click on "Using HMMER versus BLAST to find homologs" to answer the following questions.

3. What can you use HMMER for?

4. If you use BLAST instead of HMMER, what is the best BLAST program to find homolog proteins?
   a. BLASTn
   b. BLASTp or translated blast (tBLASTn, tBLASTx and BLASTx)

5. Go to slide number 11 of the tutorial "Using HMMER versus BLAST to find homologs" and try the replicate the sample exercise. Use the sample file, *Anopheles gambaie* opsins, AgGPRop (located on Hmmer tutorial page). Stop on slide 16.

   a. ClustalW Job ID _____

   b. HMMER Job ID _____

6. You can analyze result directly on VectorBase page or you can click on "Download Raw Results" and open the file in a text editor such as Notepad++ (Windows) or TextWrangler (Mac).

   - In this case is easy because *Aedes aegypti* 1.4 and *Culex quinquefascaitus* 1.3 opsins have a description associated with them *i.e.*, metadata, but not all genes have this. However, we strongly suggest that you double check if you agree with the gene metadata and analyze the **e-value** and **score** for the top hits.

7. Repeat the exercise using both ClustalW and HMMER and the same sample file but on slide 13 click on *Glossina morsitans* 1.2*, Ixodes scapualris* 1.2 and *Pediculus humanus* 1.2*.*

    a. ClustalW Job ID _____

    b. HMMER Job ID _____

8. Write down the gene ID (*e.g.*, AGAP013149) and description (*e.g.*, hypothetical protein) of the hits that you believe are opsins for each one of these three species

Organim	**Gene ID**	**Gene description**
G. morsitans 1.2		
I. scapualris 1.2		
P. humanus 1.2		

9. What other piece of evidence can you use to be more confident in your results?

    These are the opsins relationships with its other G Protein–Coupled Receptors (GPCRs) family members:

    Class A: Rhodopsin-like
    Biogenic amine
    Glycoprotein hormone/ leucine-rich repeat (LGRs)
    Peptide
    Purine
    (Rhod)opsin
    Orphan
    Class B: Secretin-like
    Calcitonin

Diuretic insect hormone
Growth hormone releasing hormone
HE6 like
Latrophilin
Methuselah-like
Orphan

Class C: Metabotropicglutamate-like

Metabotropic glutamate
$GABA_B$

Class D: Atypical GPCRs

Frizzled/Smoothened
Boss
Orphan

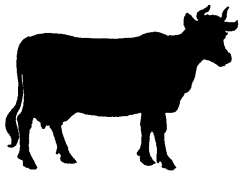
Opsins, like all proteins, are composed of multiple functional residues or domains. The bovine rhodopsin is the sequence of reference for opsin proteins and some of its functional important residues are highlighted in red in the next page.

- How is this information helpful to you to answer question 8?

_____

_____

_____



Bovine rhodopsin sequence
NCBI Reference Sequence: NP_001014890.1

```
      10         20         30         40         50
MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAEPWQFSML AAYMFLLIML
       60         70         80         90        100
GFPINFLTLY VTVQHKKLRT PLNYILLNLA VADLFMVFGG FTTTLYTSLH
      110        120        130        140        150
GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
      160        170        180        190        200
NHAIMGVAFT WVMALACAAP PLVGWSRYIP EGMQCSCGID YYTPHEETNN
      210        220        230        240        250
ESFVIYMFVV HFIIPLIVIF FCYGQLVFTV KEAAAQQQES ATTQKAEKEV
      260        270        280        290        300
TRMVIIMVIA FLICWLPYAG VAFYIFTHQG SDFGPIFMTI PAFFAKTSAV
      310        320        330        340      348
YNPVIYIMMN KQFRNCMVTT LCCGKNPLGD DEASTTVSKT ETSQVAPA
```

Based on the literature and following bovine rhodopsin nomenclature, the following are some of the functionally conserved amino acid residues for opsins:

- K296: the site of the Schiff base linkage to the chromophore
- N2 and N15: N-glycosylation sites
- C322 and C323: palmitoylation sites
- C110 and C187: disulfide bond sites
- E113: Schiff base counter-ion
- E134 and R135: sites important for transducing, binding and stabilizing the inactivated state of rhodopsin
- W126, W265, and Y268: sites involved in conformational changes of rhodopsin during retinal isomerization and formation of the retinal binding pocket
- A117, P267, and A292: sites affecting chromosome regeneration and activation of signal transduction
- H65, H152 and H211: sites important for conversion of metarhodopsin I to II and opsin activation-deactivation
- C140 and C185: sites involved in palmitoylation and phosphorylation
- E122: site involved in the stabilization of metarhodopsin II
- E134-C140, A241-K248, N310-Q312: three stretches of amino acids in the intracellular loops, I2 and I3, and 3’ C-terminus, crucial for G-protein interaction

10. Using the same sample file perform a BLASTp against the peptides of *Aedes aegypti* 1.4*, Culex quinquefascaitus* 1.3, *Glossina morsitans* 1.2*, Ixodes scapualris* 1.2 and *Pediculus humanus* 1.2*.*

    - Do you easily arrive to the same gene IDs?

    - What are the advantages and disadvantages of ClustalW and HMMER vs. BLAST?

Note: The answer key for these practice exercises will be available as a separate file next year. Remember that answers may change because VectorBase its updated every two months (www.vectorbase.org/releases).