HW11.

1. We want to know whether the human RBP4 (Retinol Binding Protein 4) has homologous proteins in insects.  Do a blastp protein search using RBP4 (NP\_006735), restricting Databases to Arthropoda (insects). Note how many homologous proteins found. Homologous proteins can be defined using an E value cutoff of 10-10. Next, do a blastn search using the RBP4 mRNA sequences (NM\_006744; select only the nucleotides corresponding to the coding region of the DNA and still restricting toe Arthropoda). Which search is more informative in identifying homologous proteins?  Try again use the DNA sequence  as query in a **blastx** search, which translates the DNA in all 6 frames and searches against a protein sequence.
2. The larges gene family is the olfactory receptor family. Do a BLAST search to evaluate how large the family is.

First identify an olfactory receptor protein from the human genome. You can search for gene first and then get the protein sequence from the gene page.  Use the protein sequence to run a blast search against human reference proteins. ( We need to change our Database to reference protein and Organisms to human (taxid:9606).

Answer 1:

1. For NP\_006735, on Arthropoda, we cannot find any homologous proteins using e-value cutoff of 10-10. The smallest e-value we can find is 2e-09.
2. For NM\_006744, on Arthropoda, we cannot find any homologous genes.

In this case, I suggest blastn search is more identifying. Because blastp search gives us several similar proteins, even though they are not so significant. In compare, blastn does not able to give us any similar sequence based on default e-value cut off.

1. Without specifying organism, I can find many highly related proteins:
2. PREDICTED: retinol-binding protein 4 isoform X1 [Chlorocebus sabaeus]

[XP\_007961780.1](https://www.ncbi.nlm.nih.gov/protein/XP_007961780.1?report=genbank&log$=protalign&blast_rank=1&RID=Z64DPK0H014)

1. Retinol-binding protein 4 isoform a precursor [Homo sapiens]

[NP\_006735.2](https://www.ncbi.nlm.nih.gov/protein/NP_006735.2?report=genbank&log$=protalign&blast_rank=2&RID=Z64DPK0H014)

1. Retinol binding protein 4, plasma [Homo sapiens]

[AAH20633.1](https://www.ncbi.nlm.nih.gov/protein/AAH20633.1?report=genbank&log$=protalign&blast_rank=3&RID=Z64DPK0H014)

1. retinol-binding protein 4 [Papio anubis]

[XP\_003904062.1](https://www.ncbi.nlm.nih.gov/protein/XP_003904062.1?report=genbank&log$=protalign&blast_rank=4&RID=Z64DPK0H014)

1. RBP4, partial [synthetic construct]

[AKI71630.1](https://www.ncbi.nlm.nih.gov/protein/AKI71630.1?report=genbank&log$=protalign&blast_rank=5&RID=Z64DPK0H014)

Answer 2:

OR1G1 olfactory receptor family 1 subfamily G member 1

mRNA: NM\_003555.1

ATGGAGGGGAAAAATCTGACCAGCATCTCAGAATGTTTCCTCCTGGGGTTCTCTGAGCAGCTGGAGGAGC

AGAAGCCCCTCTTTGGGTCCTTCCTGTTCATGTACTTGGTCACGGTGGCAGGCAACCTCCTCATCATTCT

AGTCATCATTACTGACACTCAACTCCATACCCCCATGTACTTCTTTCTAGCCAACCTCTCCCTTGCAGAT

GCCTGCTTTGTGTCCACCACAGTCCCTAAGATGCTGGCAAACATACAGATCCAGAGTCAGGCCATCTCCT

ACTCAGGGTGTCTACTACAGTTGTATTTTTTCATGTTATTTGTGATGCTGGAGGCATTCCTCTTGGCGGT

CATGGCCTATGACTGCTACGTGGCCATATGCCACCCACTTCATTACATTCTGATCATGAGCCCTGGGCTC

TGCATCTTCCTCGTGTCTGCATCCTGGATCATGAATGCCCTCCACTCCCTTCTACACACACTTCTGATGA

ACAGCCTGTCCTTCTGCGCAAACCATGAGATCCCACACTTCTTCTGTGACATCAATCCCCTCCTGAGTCT

GTCCTGCACAGACCCCTTCACCAATGAGCTGGTGATCTTCATCACTGGGGGTCTCACAGGACTCATTTGT

GTGCTTTGCCTGATTATCTCTTACACGAACGTTTTCTCGACCATCCTGAAGATCCCATCAGCTCAGGGGA

AGCGGAAAGCCTTTTCCACCTGCAGCTCTCATCTCTCCGTGGTCTCTCTCTTCTTTGGGACTTCTTTTTG

TGTTGATTTCAGTTCTCCCTCAACCCACTCGGCCCAGAAGGACACAGTTGCATCAGTGATGTACACAGTG

GTAACTCCAATGTTGAATCCCTTTATCTACAGTTTGAGGAACCAAGAAATAAAGTCTTCCCTGAGAAAGT

TAATCTGGGTTCGGAAAATTCATTCCCCTTAG

Protein: NP\_003546.1

MEGKNLTSISECFLLGFSEQLEEQKPLFGSFLFMYLVTVAGNLLIILVIITDTQLHTPMYFFLANLSLAD

ACFVSTTVPKMLANIQIQSQAISYSGCLLQLYFFMLFVMLEAFLLAVMAYDCYVAICHPLHYILIMSPGL

CIFLVSASWIMNALHSLLHTLLMNSLSFCANHEIPHFFCDINPLLSLSCTDPFTNELVIFITGGLTGLIC

VLCLIISYTNVFSTILKIPSAQGKRKAFSTCSSHLSVVSLFFGTSFCVDFSSPSTHSAQKDTVASVMYTV

VTPMLNPFIYSLRNQEIKSSLRKLIWVRKIHSP

DNA to Protein search (blastx):

1. Use default:

I found 100 matches (default max result). The highest e value is at 2e-100 level. So these result are very solid.

1. Increase search result size:

I found 1547 results, with e-value at or less than 1e-10 level.

Protein to Protein search (blast p):

1. Used default:

I found 100 matches with highest e-value at 2e-100 level.

1. Increase search result size:

I found 1547 results, with e-value at or less than 1e-10 level.

Conclusion:

1. The olfactory family are very large that default blast setting cannot find all family proteins.
2. In this case, using DNA and using protein give us almost same search result.