Eastern Biologist Supplemental File

Davis, A., H. Nguyen, and J. Qian, 2019. Zebrafish Embryos and Bioinformatics: Useful and Marketable Exercises for Students Enrolled in Upper-Level Undergraduate Courses. Eastern Biologist Special Issue 1:47–63.

SOP Supplemental file.

Supplementary Material
Zebrafish Embryos and Bioinformatics: Useful and Marketable Exercises for Students Enrolled in Upper-Level Undergraduate Courses

Adam Davis[1*], Hong Nguyen[1], and Jo Qian[1]

[1]Biology Department, University of North Georgia, Oakwood, GA 30566, USA. * Corresponding author: adam.davis@ung.edu

Standard Operating Procedures Supplemental File

Table of Contents:

SOP #	BIO-001
Title	Analysis of Hoxa2 amino acid sequences I: Extraction of Hoxa2 amino acid sequences from Genbank, Clustal Alignment, Formatting, and Color-Coding of these Sequences, and Generation of Percent Identity Matrices of Aligned Sequences
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

Objective:

To understand the function and evolution of *Hoxa2*, amino acid sequences, which make up the protein sequences, must be retrieved from the Genbank database over the National Center for Biotechnology Information (NCBI) website. Furthermore, amino acid sequences from evolutionarily divergent species must be analyzed to fully understand how proteins function. Amino acid sequences will be analyzed from several primate species, including Human (*Homo sapiens*), Chimpanzee (*Pan troglodytes*), Gorilla (*Gorilla gorilla*), and Orangutan (*Pongo abelii*) and several vertebrate biomedical models, including Zebrafish (*Danio rerio*), Chicken (*Gallus gallus*), and Mouse (*Mus musculus*). Sequence alignments between evolutionarily divergent species will help to reveal conserved and functional domains of proteins. Regions of amino acid sequences that are conserved over millions of years and between species indicate that there is functional importance of these sequences.

Relevant Terms and their Definitions:

Hoxa2 – Developmental regulatory gene that is evolutionarily conserved and functions to pattern the development of several cranial nerves and craniofacial skeletal elements.

National Center for Biotechnology Information (NCBI) – Web-based database containing genetic information submitted by scientists. Used for genetic, developmental, medical, ecological and evolutionary research analyses.

Genbank Accession Number – Identifying number for amino acid sequences

FASTA – Sequence format that must be obtained for amino acid sequences for downstream analyses.

Clustal – Multiple sequence alignment program for DNA or proteins.

Conservative mutations – Point mutation of a nucleotide that results in an amino acid change. However, the properties of the amino acid remain the same (e.g.: hydrophobic, hydrophyllic, etc.). These mutations generally are not detrimental to the organism as a whole.

Non-conservative mutations – Point mutation of a nucleotide that results in an amino acid change that has different properties from wild-type (phenotype of the typical form of a species as it occurs in nature). Non-conservative mutations may cause the protein to lose its function, which can result in a disorder in the organism (e.g.: sickle-cell anemia). This mutation may also cause a "gain of function" (or protein becoming abnormally activated) (e.g.: cancer).

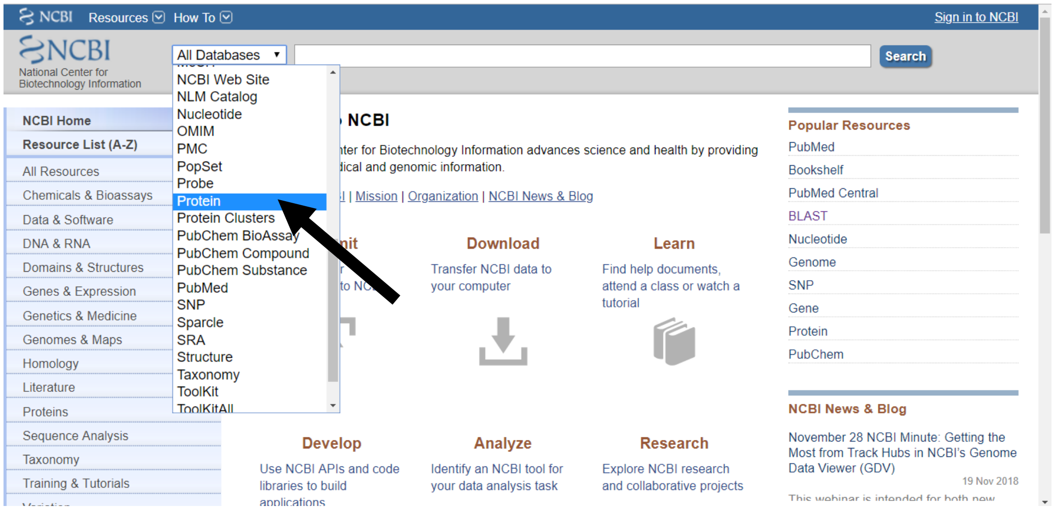
Indel mutations – Insertion or deletion of nucleotides resulting in the gain or loss of amino acids.

Procedure:

I. Retrieve all Hoxa2 amino acid sequences from the Genbank database in FASTA format and copy them to a Microsoft Word file.

A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), type in https://www.ncbi.nlm.nih.gov

B. Click on the drop down menu titled "All Databases" and select "Protein" (see image below).



Transfer NCBI data to your computer

Find help documents, attend a class or watch a tutorial

Use NCBI APIs and code libraries to build applications

Identify an NCBI tool for your data analysis task

Explore NCBI research and collaborative projects

November 28 NCBI Minute: Getting the Most from Track Hubs in NCBI's Genome Data Viewer (GDV)

C. Type in the appropriate Genbank accession number in the text box next to the Drop-down menu to retrieve the appropriate coding amino acid sequences and click on the "Search" button. The Genbank accession numbers for the species-specific *Hoxa2* amino acid sequences are listed in the table below.

Organism	***Hoxa2* Genbank Accession**
Human	NP006726
Chimpanzee	XP527697
Gorilla	XP004045263
Orangutan	XP002818153
Mouse	NP034581
Chicken	NP990481
Zebrafish	NP571181

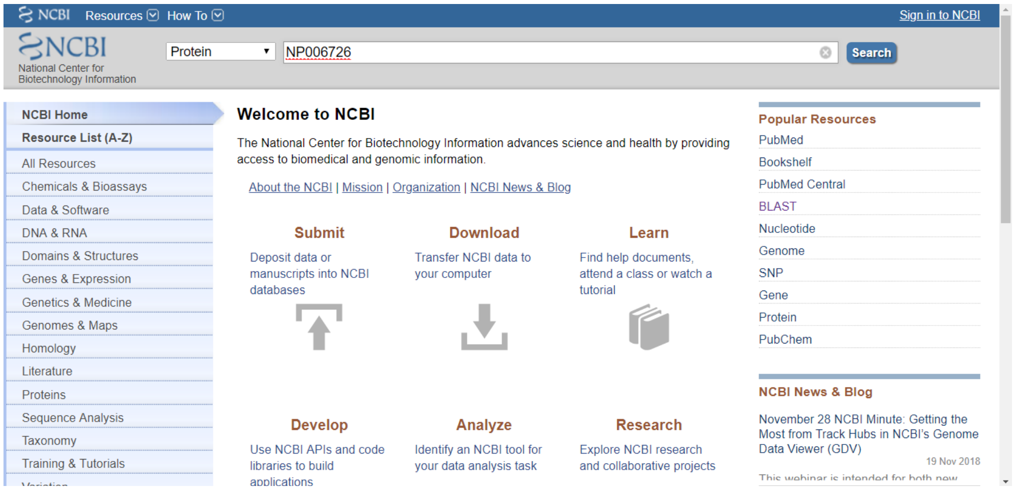
D. Once the Genbank sequence information is displayed, retrieve the Hoxa2 amino acid sequence in FASTA format.

   NOTE 1: FASTA format is necessary for future downstream analyses, including amino acid sequence alignment.

   NOTE 2: The Human Hoxa2 amino acid sequence (Accession #: NP006726) will be used as an example for retrieving amino acid sequences in FASTA format.

   1. Type "NP006726" in the text box in the to the right of the dropdown menu.

   2. Click on the "Search" button to the right of the textbox (see image below).



Protein NP006726

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

Submit: Deposit data or manuscripts into NCBI databases

Download: Transfer NCBI data to your computer

Learn: Find help documents, attend a class or watch a tutorial

Develop: Use NCBI APIs and code libraries to build applications

Analyze: Identify an NCBI tool for your data analysis task

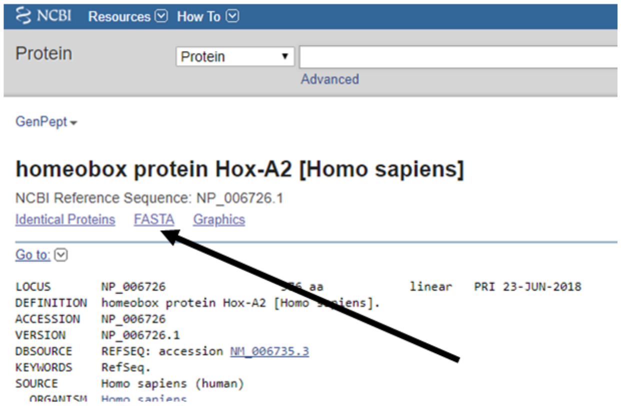
Research: Explore NCBI research and collaborative projects

NCBI News & Blog

November 28 NCBI Minute: Getting the Most from Track Hubs in NCBI's Genome Data Viewer (GDV)

19 Nov 2018

   3. The Human Hoxa2 amino acid sequence Genbank information will be displayed. "Click on FASTA" on the upper left corner of the screen. The FASTA sequence format removes all identifying information from the sequence file (see images below).



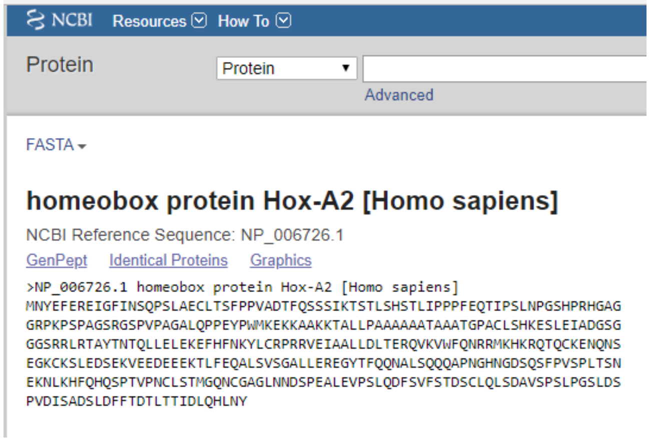
homeobox protein Hox-A2 [Homo sapiens]

NCBI Reference Sequence: NP_006726.1

Identical Proteins FASTA Graphics

Go to:

```
LOCUS       NP_006726                376 aa            linear   PRI 23-JUN-2018
DEFINITION  homeobox protein Hox-A2 [Homo sapiens].
ACCESSION   NP_006726
VERSION     NP_006726.1
DBSOURCE    REFSEQ: accession NM_006735.3
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```



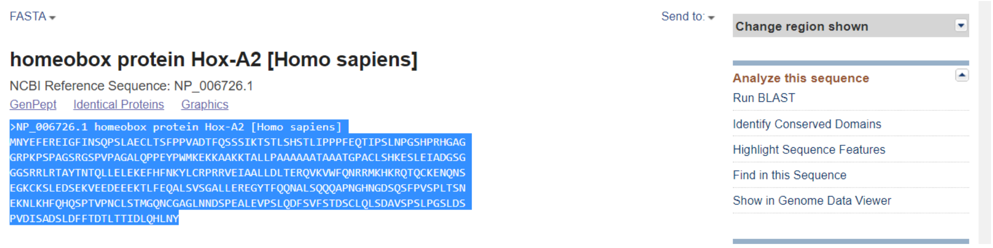
homeobox protein Hox-A2 [Homo sapiens]

NCBI Reference Sequence: NP_006726.1

GenPept Identical Proteins Graphics

```
>NP_006726.1 homeobox protein Hox-A2 [Homo sapiens]
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHGAG
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGSG
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNS
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSN
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDS
PVDISADSLDFFTDTLTTIDLQHLNY
```

4. Highlight the entire sequence including the sequence identification line, copy it, and paste it into a new Microsoft Word document (see image below).



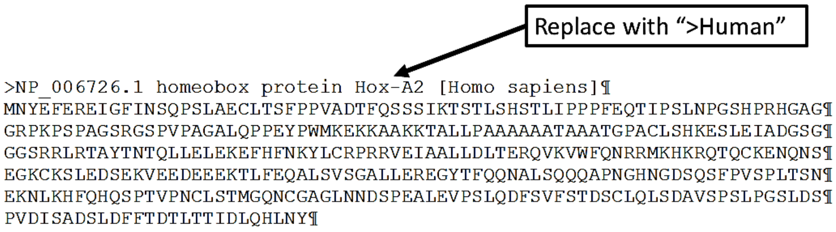
homeobox protein Hox-A2 [Homo sapiens]

NCBI Reference Sequence: NP_006726.1

GenPept Identical Proteins Graphics

```
>NP_006726.1 homeobox protein Hox-A2 [Homo sapiens]
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHGAG
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGSG
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNS
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSN
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDS
PVDISADSLDFFTDTLTTIDLQHLNY
```

5. In Microsoft Word, replace all information in the sequence identification line after the carrot symbol (>) with the word "Human"(see images below). **<u>Make sure to keep the ">" symbol!!!!</u>**



Replace with ">Human"

```
>NP_006726.1 homeobox protein Hox-A2 [Homo sapiens]¶
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHGAG¶
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGSG¶
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNS¶
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSN¶
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDS¶
PVDISADSLDFFTDTLTTIDLQHLNY¶
```

```
>Human
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHGAG
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGSG
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNS
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSN
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDS
PVDISADSLDFFTDTLTTIDLQHLNY
```

6. Repeat steps I.C.-I.D.5. for adding all other primate and model vertebrate Hoxa2 amino acid sequences using the accession numbers listed on the table on page 3 of this SOP.

7. Once all sequences have been added to the Microsoft Word file, save this file as “Hoxa2_Protein_Unaligned”.

NOTE 3: Several FASTA sequences are shown below as an example.

```
>Human
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHGAG
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGSG
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNS
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSN
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDS
PVDISADSLDFFTDTLTTIDLQHLNY
>Chimpanzee
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHGAG
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGSG
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNS
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSN
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDS
PVDISADSLDFFTDTLTTIDLQHLNY
>Gorilla
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHSAG
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAAAATGPACLSHKESLEIADGSGG
GSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNSE
GKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNE
KNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSP
VDISADSLDFFTDTLTTIDLQHLNY
```

II. Align the FASTA formatted Primate Hoxa2 amino acid sequences using the Clustal alignment software program.

A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), pull up the Clustal Omega website.

1. Type in https://www.ebi.ac.uk/Tools/msa/clustalo/

NOTE 4: Clustal produces biologically meaningful multiple sequence alignments of evolutionarily divergent DNA or protein sequences. It calculates the best match for the selected sequences and lines them up so that the identities, similarities and differences can be observed.

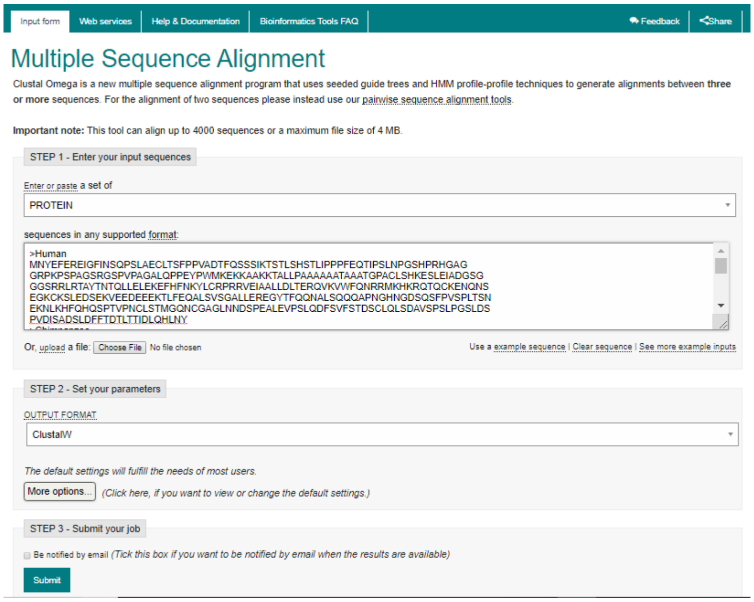
B. Click on the drop-down menu to select the appropriate sequence data to align.

C. Select “Protein” from the drop-down menu.

D. Copy all four primate Hoxa2 amino acid sequences (Human, Chimpanzee, Gorilla, and Organgutan), including the sequence identification lines with the “>” symbols, from the “Hoxa_Protein_Unaligned” file and paste these sequences in the textbox under the text, “sequences in any supported format”.

E. Click on the dropdown menu under the text, “OUTPUT FORMAT”, and select “ClustalW”.

F. Click on the “Submit” button on the bottom of the screen (see image below).



Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of: PROTEIN

sequences in any supported format:

```
>Human
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPRHGAG
GRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKESLEIADGSG
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKRQTQCKENQNS
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSN
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDS
PVDISADSLDFFTDTLTTIDLQHLNY
```

STEP 2 - Set your parameters

OUTPUT FORMAT: ClustalW

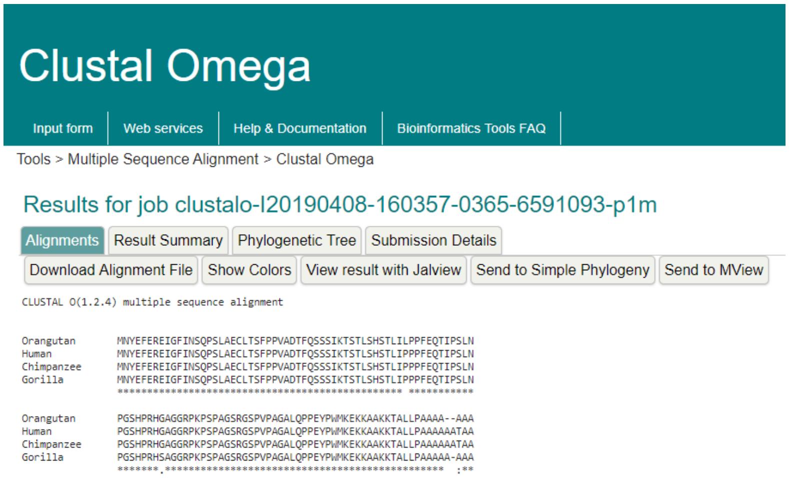
The default settings will fulfill the needs of most users.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

☐ Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

G. Once the results of the sequence alignment are complete, copy all aligned sequences and paste them into a new Microsoft Word file. Below is a subset of a sequence alignment for the primate Hoxa2 protein sequences.



Results for job clustalo-I20190408-160357-0365-6591093-p1m

```
CLUSTAL O(1.2.4) multiple sequence alignment


Orangutan       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Human           MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Gorilla         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
                ************************************************ ***********

Orangutan       PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Human           PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Chimpanzee      PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Gorilla         PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
                *******.***********************************************  :**
```

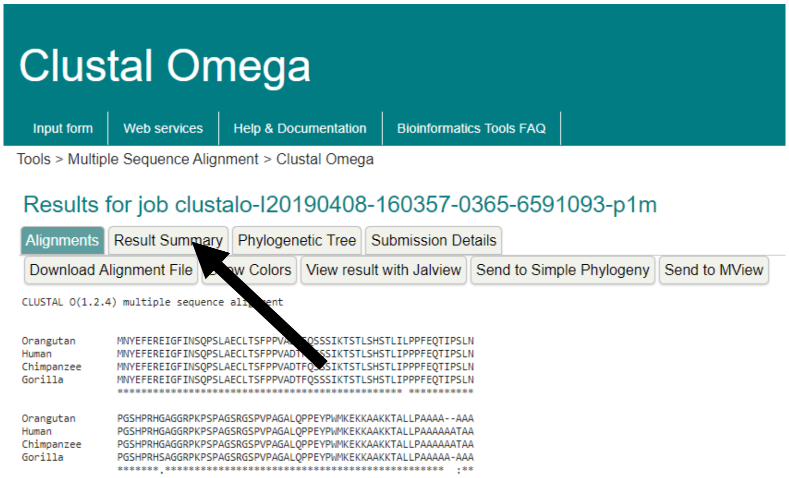
NOTE 5: Gaps are inserted between amino acid residues so that identical or similar characters are aligned in successive columns. Therefore, gaps represent insertion and deletion mutations.

NOTE 6: Asterisks under sequences denote amino acids that show 100% sequence identity across all sequences analyzed. Columns without asterisks indicate that at least one point mutation (conservative, non-conservative, insertion, or deletion) has occurred at that site within the coding DNA sequences.

H. Save the new Microsoft Word file as “Hoxa2_Primate_Protein_Aligned”

III. Obtain the Percent Identity Matrix for the Primate Hoxa2 Sequence Alignment from the Clustal Omega software program.

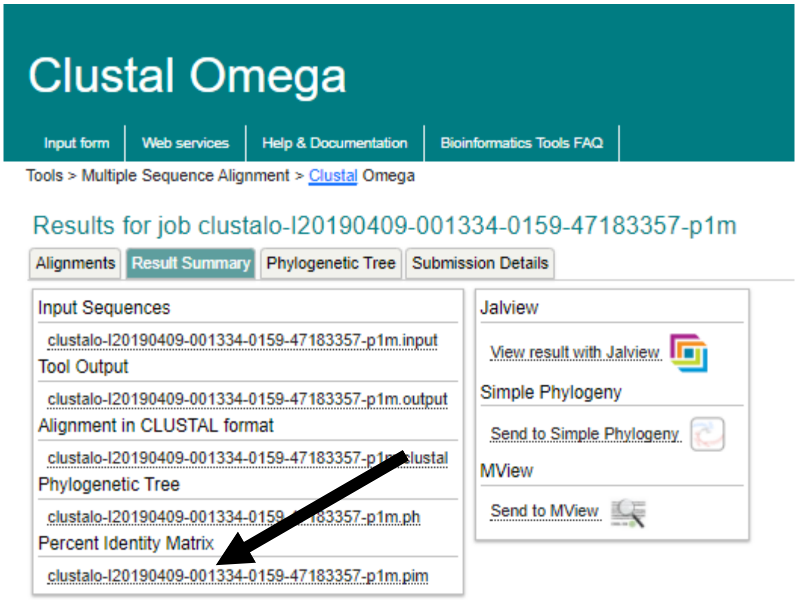
A. Click on the “Result Summary” button (see image below).



Results for job clustalo-I20190408-160357-0365-6591093-p1m

```
CLUSTAL O(1.2.4) multiple sequence ali[illegible]ent


Orangutan       MNYEFEREIGFINSQPSLAECLTSFPPVA[illegible]QSSSIKTSTLSHSTLILPPFEQTIPSLN
Human           MNYEFEREIGFINSQPSLAECLTSFPPVADT[illegible]SSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQ[illegible]SIKTSTLSHSTLIPPPFEQTIPSLN
Gorilla         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
                ************************************************ ***********

Orangutan       PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Human           PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Chimpanzee      PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Gorilla         PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
                *******.***********************************************  :**
```

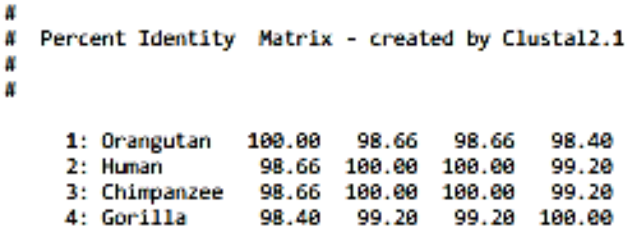
B. Click on the link for the Percent Identity Matrix on the next web page (see image below).



Results for job clustalo-I20190409-001334-0159-47183357-p1m

Input Sequences
clustalo-I20190409-001334-0159-47183357-p1m.input
Tool Output
clustalo-I20190409-001334-0159-47183357-p1m.output
Alignment in CLUSTAL format
clustalo-I20190409-001334-0159-47183357-p1m.clustal
Phylogenetic Tree
clustalo-I20190409-001334-0159-47183357-p1m.ph
Percent Identity Matrix
clustalo-I20190409-001334-0159-47183357-p1m.pim

C. A new page will appear showing the Percent Identity Matrix for all of the primates (see image below).

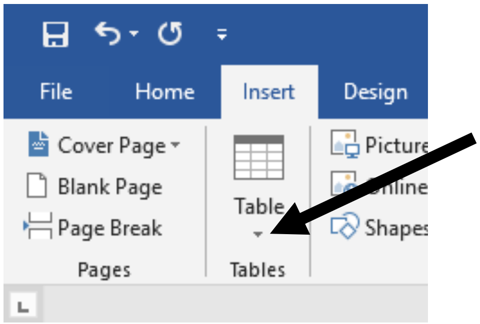


```
#
#  Percent Identity  Matrix - created by Clustal2.1
#
#

     1: Orangutan   100.00   98.66   98.66   98.40
     2: Human        98.66  100.00  100.00   99.20
     3: Chimpanzee   98.66  100.00  100.00   99.20
     4: Gorilla      98.40   99.20   99.20  100.00
```

NOTE 7: The percent identity matrix shows the percent similarity between Hoxa2 amino acid sequences of all sequence pairs. For instance, the Orangutan Hoxa2 is 100% similar to itself, 98.66% similar to Human, 98.66% similar to Chimpanzee, and 98.4% similar to Gorilla.

D. Open a new Microsoft Word file and make a 5 row X 5 column table.

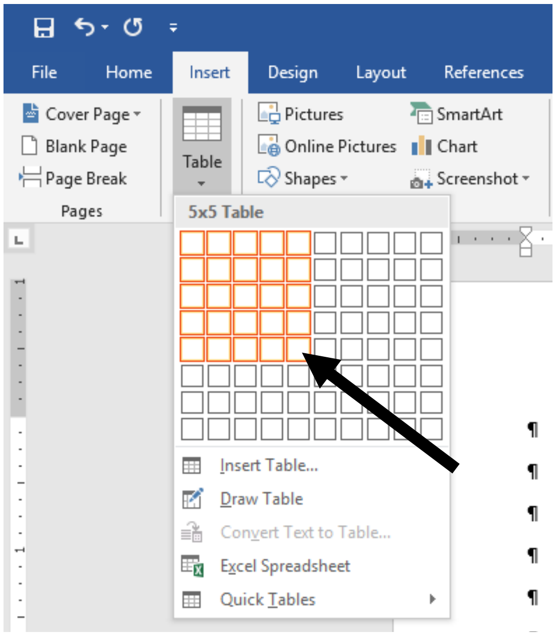
1. Under the “Insert” tab, click on the down arrow next underneath Table (see image below).



2. A grid representing the columns and tables will pop up below the arrow.

3. Move the cursor over the grid to highlight a 5x5 table.

4. Click the left mouse button on the square that represents the cell in the 5th row and 5th column. A 5x5 table will be appear in the Microsoft Word document (see image below).



E. Label the table and add in the Percent Identity values from the data obtained from the Clustal Omega software program.

1. Center all cells in Columns 2-5.

2. Leave the top-most and left-most cell blank but type in the primate names in the following order in the cells of the first row and first column: Human, Chimpanzee, Gorilla, Orangutan.

3. For cells that correspond to the primate Hoxa2 sequence being compared to itself (e.g.: Human vs Human), type "---" (see image below).

	Human	Chimpanzee	Gorilla	Orangutan
Human	---			
Chimpanzee		---		
Gorilla			---	
Orangutan				---

4. Type in the percent identity data for all other pairwise comparisons. These should be entered in twice for the table. For instance, Human is 100% similar to Chimpanzee, 99.2% similar to Gorilla, and 98.66% similar to Orangutan (see image below).

	Human	Chimpanzee	Gorilla	Orangutan
Human	---	100	99.2	98.66
Chimpanzee	100	---		
Gorilla	99.2		---	
Orangutan	98.66			---

F. Save the file as “Hoxa2_Primate_Protein_Matrix”.

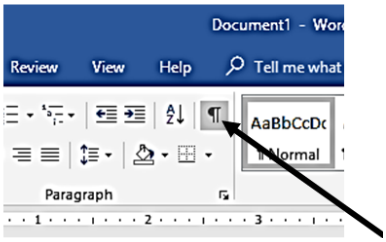
IV. Follow steps II.A.-II.G. to align the Human Hoxa2 amino acid sequence with the three model vertebrate amino acid sequences (Mouse, Chicken, and Zebrafish). Save the new Microsoft Word file as “Hoxa2_ModelVert_Protein_Aligned”.

V. Follow steps III.A.-III.E. to obtain the Percent Identity Matrix data for the model vertebrates. The order for the organisms in the first row and column should be Human, Mouse, Chicken, and Zebrafish. Save the new Microsoft Word file as “Hoxa2_ModelVert_Protein_Matrix”.

VI. Format the amino acid sequence alignment documents developed from steps II. and IV. above for amino acid color coding and eventual figure development.

NOTE 8: The “Hoxa2_Primate_Protein_Aligned” document will be used as an example in this SOP.

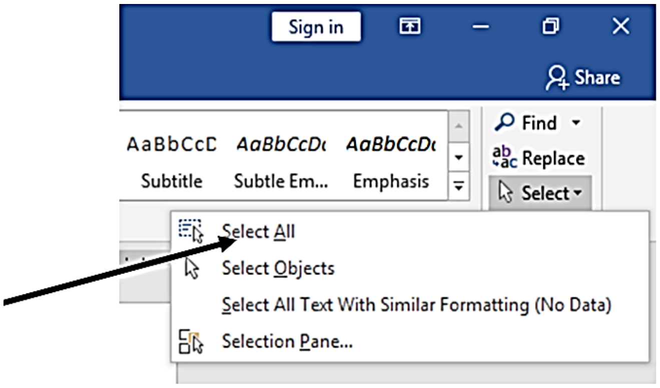
A. Under the Home tab in Microsoft Word, click on the paragraph button (¶). This will allow all spaces and paragraph marks to be viewed. Spaces appear as dots (see image below).



B. If the data is not already at a Font style of “Courier New” and “9”, change the Font style to “Courier New” and the Font size to “9”.

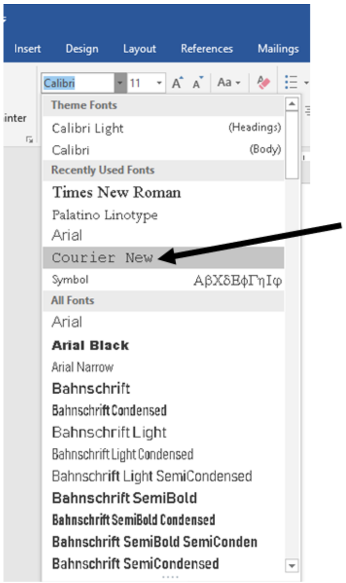
1. Under the “Home” tab, click on the down arrow next to the “Select” button at the top-right portion of the screen.

2. Click on “Select All”. All data in the word document should now be highlighted (see image below).



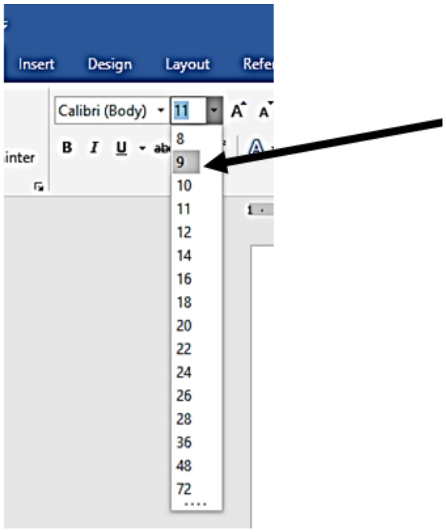
3. While all data is highlighted, under the “Home” tab, click on the down arrow in the window that reads the font style

4. A dropdown menu will appear showing all font styles. If “Courier New” is not already the font in place, click on it from the dropdown menu (see image below).

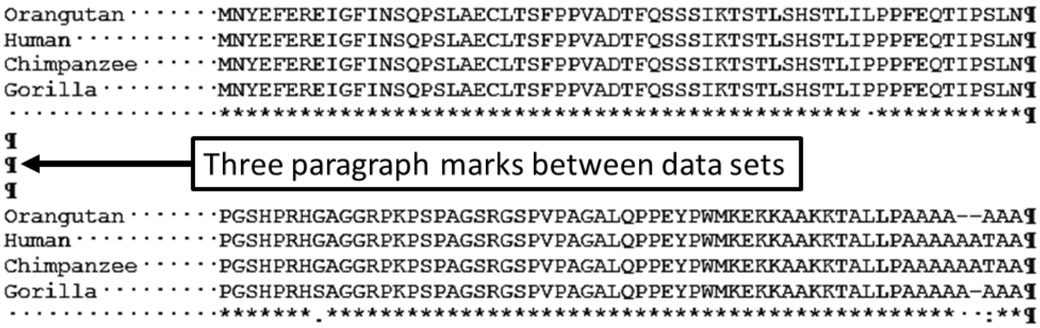


5. While all data is still highlighted, under the “Home” tab, click on the down arrow in the window that reads the font size.

6. Click on “9”. All data should now be in “Courier New” font at a size of “9” (see image below).



C. After each set of data, add two paragraph (¶) marks so that each set of data (including the line with the asterisks) is separated by three paragraph marks. This will be necessary when labeling specific regions of the protein sequence (see example below).



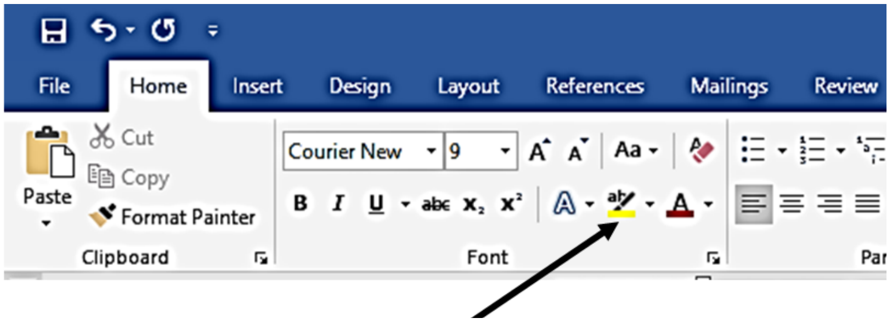
```
Orangutan······MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN¶
Human··········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Chimpanzee·····MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Gorilla········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
···············************************************************·***********¶
¶
¶
¶
Orangutan······PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA¶
Human··········PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Chimpanzee·····PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Gorilla········PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA¶
···············*******.***********************************************··:**¶
```

D. Save the file.

E. Follow steps VI.A.-VI.D. for the “Hoxa2_ModelVert_Protein_Aligned” document.

VII. Color code the amino acid data to highlight conserved protein regions.

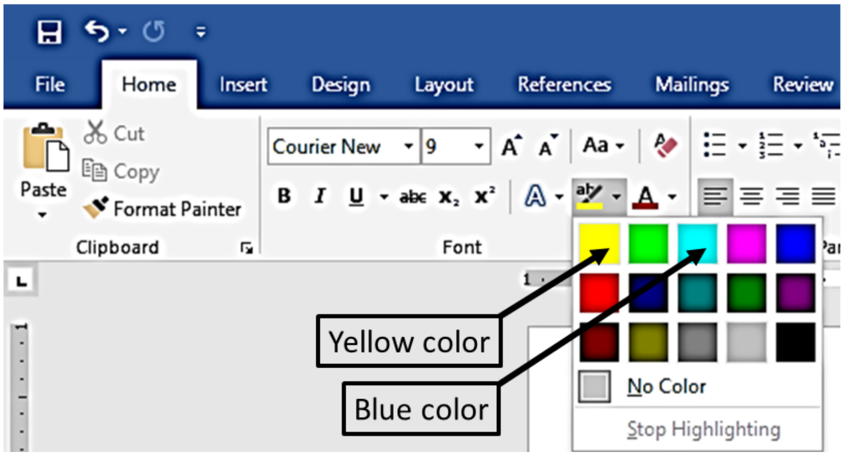
A. In Microsoft Word, use the Text Highlight Color to color-code the amino acid data (see image below).



1. Amino acid residues in columns that show 100% sequence identity across all organisms assayed must be colored yellow.

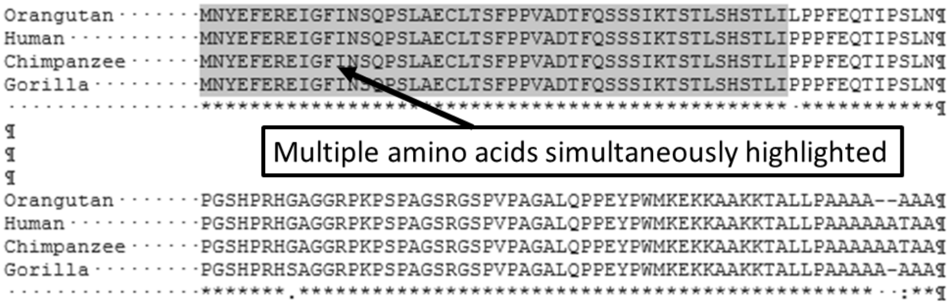
   a. Highlight amino acid sites that are 100% identical. Under the “Home” tab, click the button with the “ab” text and paintbrush symbol to highlight the text.

   b. Press the “yellow” color to color the amino acid yellow (see below).



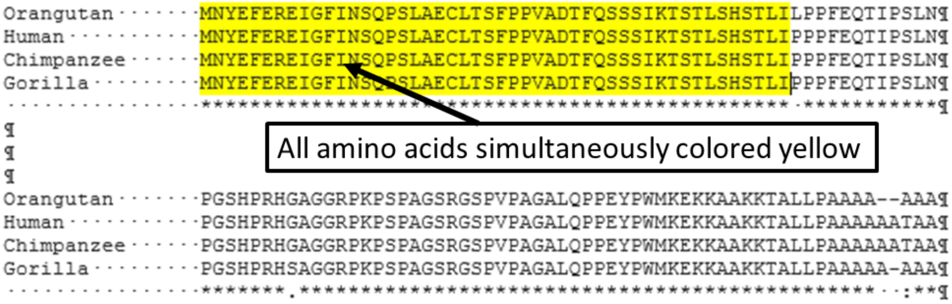
NOTE 9: Asterisks underneath sequence alignments denote 100% sequence identity at specific sites. Therefore, asterisks can be used as a guide to color-code amino acid sites with 100% sequence identity.

NOTE 10: When there are two or more columns of data that have similar sequences, the “Alt” button on the keyboard and the left button on the mouse can be held down to highlight large segments of sequence. Once all desired amino acid characters are highlighted, let go of both the “Alt” button on the keyboard and the left mouse button. All highlighted characters can now be color-coded (see images below).



```
Orangutan·······MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN¶
Human···········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Chimpanzee······MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Gorilla·········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
················************************************************·***********¶
¶
¶
¶
Orangutan·······PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA¶
Human···········PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Chimpanzee······PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Gorilla·········PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA¶
················*******.***********************************************··:**¶
```

Multiple amino acids simultaneously highlighted



```
Orangutan·······MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN¶
Human···········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Chimpanzee······MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Gorilla·········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
················************************************************·***********¶
¶
¶
¶
Orangutan·······PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA¶
Human···········PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Chimpanzee······PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Gorilla·········PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA¶
················*******.***********************************************··:**¶
```

All amino acids simultaneously colored yellow

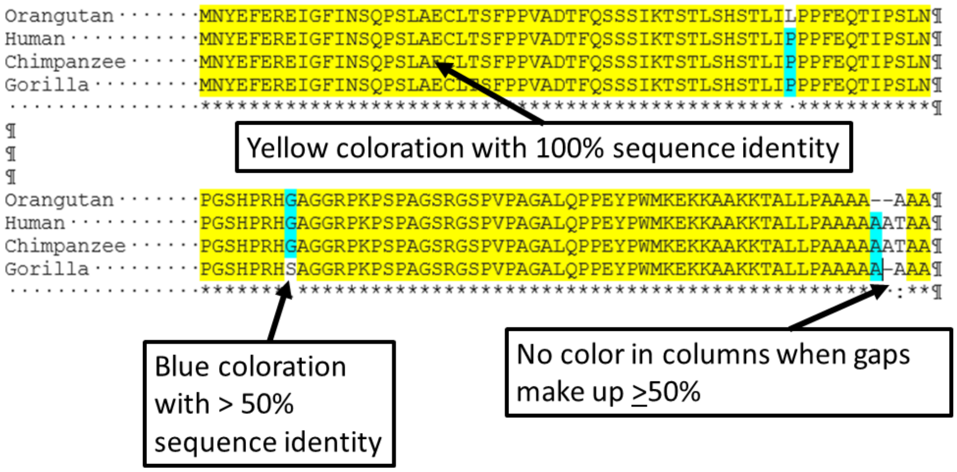
2. Amino acid residues in columns that make up more than 50% but less than 100% of the residues in columns across all organisms assayed must be colored blue. Since all sequence alignment files contain four sequences, columns must have 3 residues that are identical when using the blue coloration.

   a. Highlight amino acid sites that make up more than 50% of the columns of data. Under the "Home" tab, click the button with the "ab" text and paintbrush symbol to highlight the text.

   b. Press the "blue" color to color the amino acid blue.

   NOTE 11: **Gaps, which represent indel mutations, are not given any colors, even if there are three gaps per one site in the sequence alignment. Sites with three gaps should not be given any color**.

   NOTE 12: See the examples for the *Hoxa2* amino acid data below. Regions of aligned sequences that show increased yellow and blue coloring indicate functional significance and overall selective constraint.



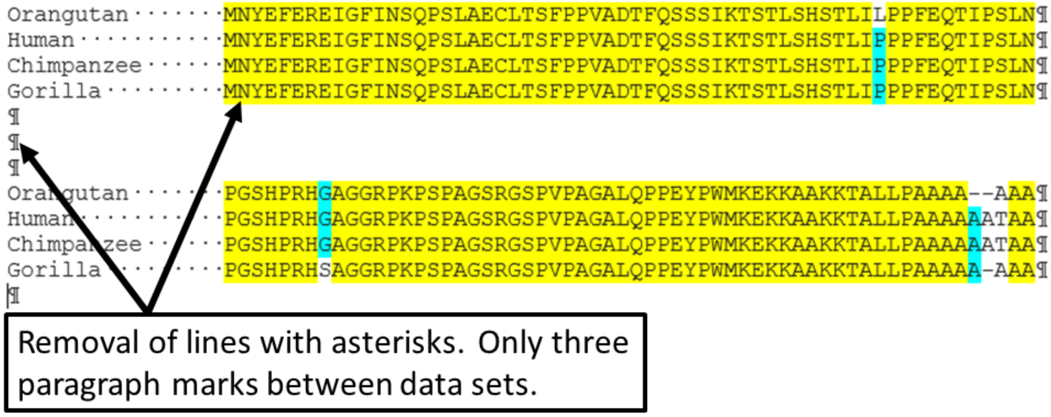
```
Orangutan······MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN¶
Human··········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Chimpanzee·····MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Gorilla········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
···············************************************************·***********¶
¶
¶
¶
Orangutan······PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA¶
Human··········PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Chimpanzee·····PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Gorilla········PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA¶
···············*******·***********************************************·:**¶
```

Yellow coloration with 100% sequence identity

Blue coloration with > 50% sequence identity

No color in columns when gaps make up ≥50%

B. Once all amino acid color coding is finished, remove all lines containing asterisks and dots underneath each set of amino acid sequence alignment. Therefore, each data set of amino acid sequence should be separated by just three paragraph marks (see example below).



```
Orangutan······MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN¶
Human··········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Chimpanzee·····MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
Gorilla········MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN¶
¶
¶
¶
Orangutan······PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA¶
Human··········PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Chimpanzee·····PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA¶
Gorilla········PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA¶
¶
```

Removal of lines with asterisks. Only three paragraph marks between data sets.

C. Save the file.

D. Follow steps VII.A.-VII.C. for the "Hoxa2_ModelVert_Protein_Aligned" document.

VIII. Import the formatted and color-coded amino acid sequence alignment data from the "Hoxa2_Primate_Protein_Aligned" document as a figure into Microsoft Powerpoint.

A. Copy the entire color-coded Hoxa2 Primate sequence alignment.

B. Paste the aligned data in Microsoft Powerpoint as an image.

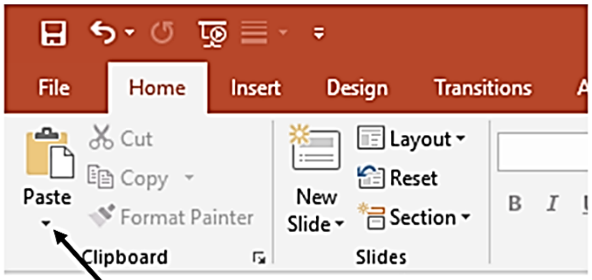
1. Open Microsoft Powerpoint and select "Blank Presentation".

2. Remove the text boxes that are present in the Powerpoint file.

3. Paste the amino acid sequence alignment text from the Word file into Powerpoint as an image.

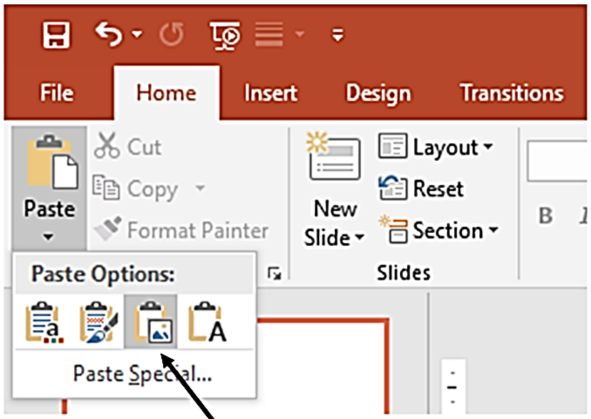
   a. Under the Home tab in Powerpoint, click on the small, black arrowhead that is pointing down and located underneath the Paste clipboard icon (shown below).

   b. Under the Home tab in Powerpoint, click on the small, black arrowhead that is pointing down and located underneath the Paste clipboard icon (shown below).



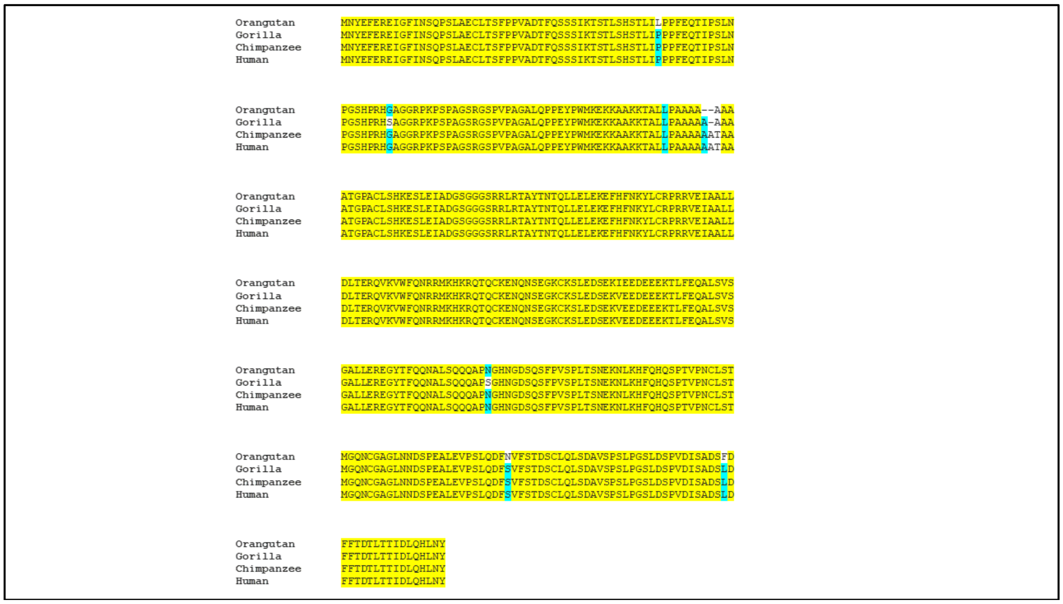
Arrow under clipboard

   c. A new pop-up window will open just below the paste clipboard icon that is labeled "Paste Options". Click on the third option from the left. When the mouse cursor rests on this image, a small text box labeled "Picture (U)" should appear (see example below).



Icon to click on to convert text to figure

   d. The sequence alignment will appear as an image in the center of the PowerPoint slide (see image below).



```
Orangutan     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN


Orangutan     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla       PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human         PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA


Orangutan     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL


Orangutan     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS


Orangutan     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla       GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human         GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST


Orangutan     MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla       MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee    MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD


Orangutan     FFTDTLTTIDLQHLNY
Gorilla       FFTDTLTTIDLQHLNY
Chimpanzee    FFTDTLTTIDLQHLNY
Human         FFTDTLTTIDLQHLNY
```

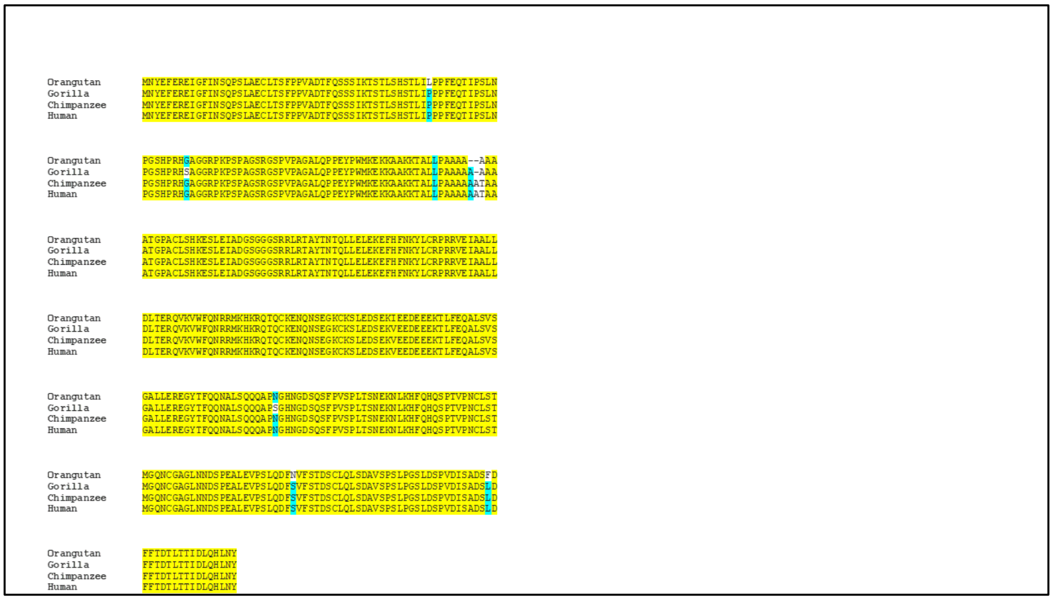
C. Move the Primate sequence alignment data to the left of the screen by left clicking on the image, holding the mouse button down, and dragging the image to the left (see image below).

```
Orangutan     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN


Orangutan     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla       PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human         PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA


Orangutan     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL


Orangutan     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS


Orangutan     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla       GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human         GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST


Orangutan     MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla       MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee    MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD


Orangutan     FFTDTLTTIDLQHLNY
Gorilla       FFTDTLTTIDLQHLNY
Chimpanzee    FFTDTLTTIDLQHLNY
Human         FFTDTLTTIDLQHLNY
```

D. Add a title to the Primate amino acid sequence alignment data.

1. Shrink the Primates alignment figure in size slightly so that a title can be added above the figure.

2. Click and hold one of the upper corners of one of the alignment images and drag downward and inward to shrink the image (see image of slightly shrunken sequence alignment below).

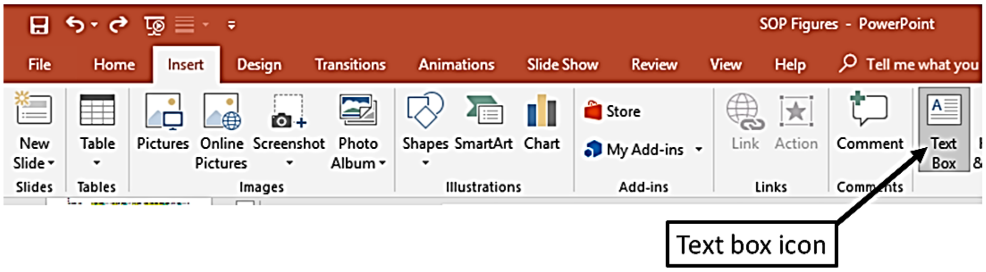


```
Orangutan       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human           MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan       PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla         PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee      PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human           PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee      ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human           ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee      DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human           DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan       GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla         GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee      GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human           GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan       MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee      MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human           MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan       FFTDTLTTIDLQHLNY
Gorilla         FFTDTLTTIDLQHLNY
Chimpanzee      FFTDTLTTIDLQHLNY
Human           FFTDTLTTIDLQHLNY
```

3. Add a title over the alignment data.

   a. Click on the “Insert” tab.

   b. Under the “Insert” tab, click on “Text Box” (See image below).



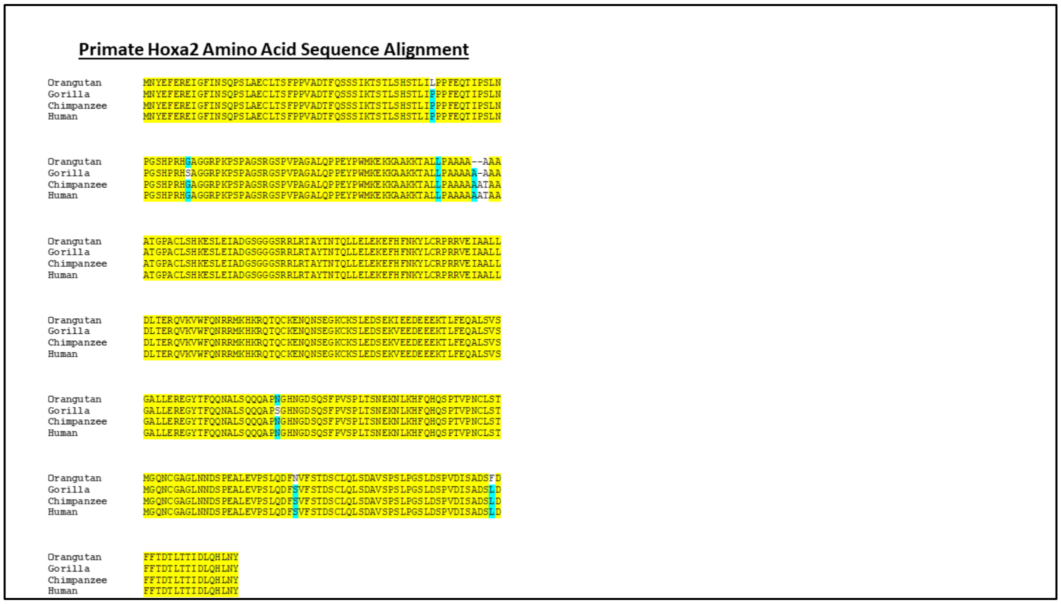
   c. The mouse icon will appear as a cursor. Move the cursor above the Primates sequence data and left click the mouse button. A text box will automatically appear where the mouse was clicked.

   d. Type “Primate Hoxa2 Amino Acid Sequence Alignment”.

   e. Keep the Font type and size as is or change it to your liking.

   f. Bold and Underline all Text.

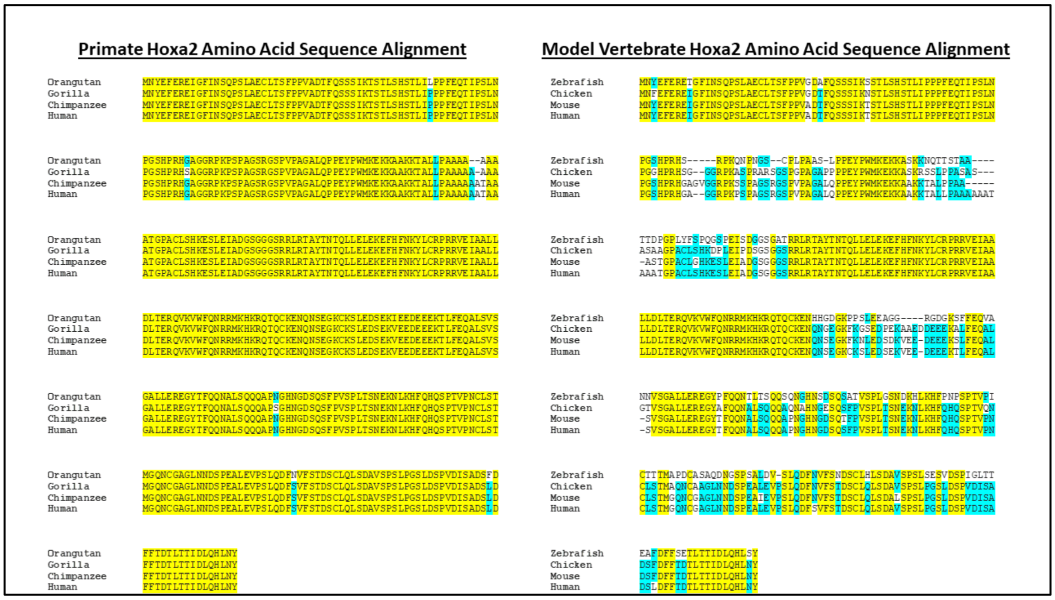
g. Position the text box so that it is centered over the Primate data set (see image below).



Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human           MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN


Orangutan       PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla         PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee      PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human           PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA


Orangutan       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee      ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human           ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL


Orangutan       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee      DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human           DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS


Orangutan       GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla         GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee      GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human           GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST


Orangutan       MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee      MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human           MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD


Orangutan       FFTDTLTTIDLQHLNY
Gorilla         FFTDTLTTIDLQHLNY
Chimpanzee      FFTDTLTTIDLQHLNY
Human           FFTDTLTTIDLQHLNY
```

IX. Follow steps VIII.A.-VIII.D.3.f. for importing the formatted and color-coded amino acid sequence alignment data from the "Hoxa2_ModelVert_Protein_Aligned" document as a figure into Microsoft Powerpoint. However, this image should be (1) on the right side of the Microsoft PowerPoint slide, (2) resized and positioned so that it is the same size and at the same level along the Y axis of the slide as the Primate Hoxa2 amino acid sequence alignment and at the same position, and (3) given the title, "Model Vertebrate Hoxa2 Amino Acid Sequence Alignment", that is in the same font type and size as that for the Primate sequence alignment, centered above the Model Vertebrate amino acid sequence alignment and positioned at the same level along the Y axis of the PowerPoint slide as the title for the Primate Hoxa2 amino acid sequence alignment (see image below).



Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla       PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human         PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla       GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human         GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan     MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla       MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee    MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan     FFTDTLTTIDLQHLNY
Gorilla       FFTDTLTTIDLQHLNY
Chimpanzee    FFTDTLTTIDLQHLNY
Human         FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

```
Zebrafish     MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken       MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish     PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken       PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse         PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human         PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish     TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken       ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse         -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human         AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish     LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken       LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish     NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken       GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish     CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken       CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse         CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human         CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish     EAFDFFSETLTTIDLQHLSY
Chicken       DSFDFFTDTLTTIDLQHLNY
Mouse         DSFDFFTDTLTTIDLQHLNY
Human         DSLDFFTDTLTTIDLQHLNY
```

X. Save the Microsoft PowerPoint file as “Hoxa2_Figures_Powerpoint”. More figures from later BIO SOPs will be made in this file.

XI. Turn in all three documents listed below according to the deadline set by the instructor:

1. Hoxa2_Figures_Powerpoint
2. Hoxa2_Primate_Protein_Matrix
3. Hoxa2_ModelVert_Protein_Matrix

References:

Gendron-Maguire, M., M. Mallo, M. Zhang, T. Gridley. 1993. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75(7):1317-1331.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947-2948.

Rijli, F.M., M. Mark, S. Lakkaraju, A. Dierich, P. Dolle, and P. Chambon. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75(7):1333-1349.

SOP #	BIO-001
Title	Analysis of Hoxa2 amino acid sequences I: Extraction of Hoxa2 amino acid sequences from Genbank, Clustal Alignment, Formatting, and Color-Coding of these Sequences and Generation of Percent Identity Matrices of Aligned Sequences
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

SOP Assessment

1. True or False: A conservative mutation gives rise to an amino acid with different properties from the wild-type amino acid.

2. NCBI is short for:

    a. National Center for Bigoted Idiots
    b. National Center for Biological Instagrams
    c. National Center for Biotechnology Information
    d. National Center for Biochemical Issues

3. Amino acid sequence formats that do not contain all sequence identifying information and are necessary for downstream analyses are:

    a. Genbank
    b. FASTA
    c. Graphical
    d. Biotech

4. Gaps in sequence alignments represent:

    a. Conservative mutations
    b. Non-conservative mutations
    c. Insertion or Deletion (Indel) mutations
    d. Clustal mutations

Date: ____

Name (Print): ____

Signature: ____

SOP #	BIO-002
Title	Analysis of Hoxa2 amino acid sequences II: Identification and labeling of Functional Hoxa2 Protein Domains.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

Objective:

To understand the function and evolution of *Hoxa2*, the functional domains of the Hoxa2 protein must be identified. The Hoxa2 Amino acid sequence retrieved from the NCBI website from BIO-001 SOP will be used in the ProSite software to identify the functional domains. Further, these domains will be labeled using Microsoft PowerPoint.

Relevant Terms and their Definitions

Hoxa2 – Developmental regulatory gene that is evolutionarily conserved and functions to pattern the development of several cranial nerves and craniofacial skeletal elements.

ProSite – online software package and database of protein domains and families.

Hexapeptide – Functional domain of Hoxa2 that binds with co-transcription factors, such as Pbx and Meis

Homeodomain – DNA binding domain of Hoxa2.

Procedure

I. Identify the functional domains of Hoxa2 by using the PROSITE database.

  A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), pull up the PROSITE website.

    1. Type in https://prosite.expasy.org/prosite.html

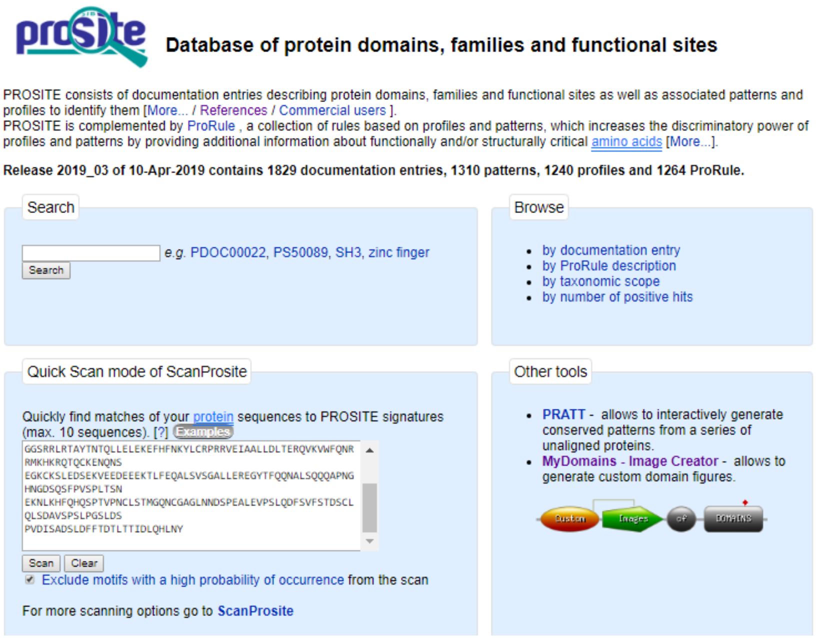
  B. Open the "Hoxa2_Protein_Unaligned" file from the BIOI-001 SOP.

  C. Copy the entire Human Hoxa2 amino acid sequence but **NOT** the sequence identification line (>Human).

  D. Paste the sequence in the textbox underneath the title "Quick Scan mode of ScanProSite".

  E. If it is not already checked, check the box for the "Exclude motifs with a high probability of occurrence from the scan" option.

  F. Click on the "Scan" button (see image below).



Database of protein domains, families and functional sites

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More... / References / Commercial users].
PROSITE is complemented by ProRule , a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [More...].

Release 2019_03 of 10-Apr-2019 contains 1829 documentation entries, 1310 patterns, 1240 profiles and 1264 ProRule.

Search

e.g. PDOC00022, PS50089, SH3, zinc finger

Browse

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hits

Quick Scan mode of ScanProsite

Quickly find matches of your protein sequences to PROSITE signatures (max. 10 sequences). [?]

```
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNR
RMKHKRQTQCKENQNS
EGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREGYTFQQNALSQQQAPNG
HNGDSQSFPVSPLTSN
EKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCL
QLSDAVSPSLPGSLDS
PVDISADSLDFFTDTLTTIDLQHLNY
```

☑ Exclude motifs with a high probability of occurrence from the scan

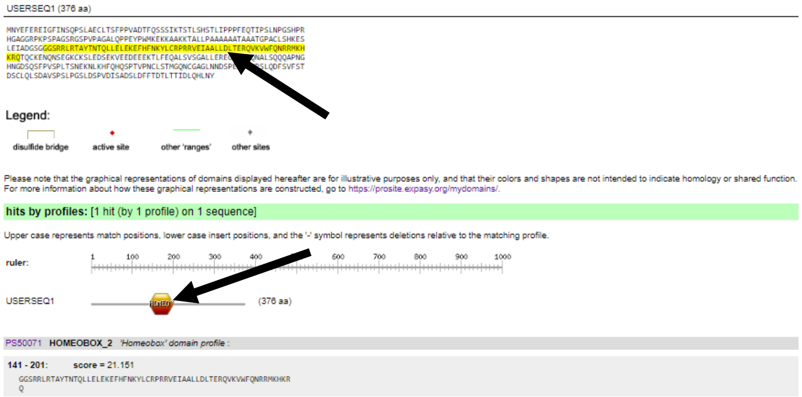
For more scanning options go to **ScanProsite**

Other tools

- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins.
- **MyDomains - Image Creator** - allows to generate custom domain figures.

G. The results page will appear with the Hoxa2 amino acid sequence and hits (or matches) of functional domains of Hoxa2 with those of other proteins within the database.

1. Under "hits by profiles", move your cursor over the orange hexagon labeled "HOMEO". The amino acids making up this domain will be highlighted in yellow (see image below).



USERSEQ1 (376 aa)

```
MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPR
HGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKES
LEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKH
KRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVSGALLEREG     QNALSQQQAPNG
HNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLSTMGQNCGAGLNNDSPE     PSLQDFSVFST
DSCLQLSDAVSPSLPGSLDSPVDISADSLDFFTDTLTTIDLQHLNY
```

Legend:

disulfide bridge
active site
other 'ranges'
other sites

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes are not intended to indicate homology or shared function.
For more information about how these graphical representations are constructed, go to https://prosite.expasy.org/mydomains/.

hits by profiles: [1 hit (by 1 profile) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.

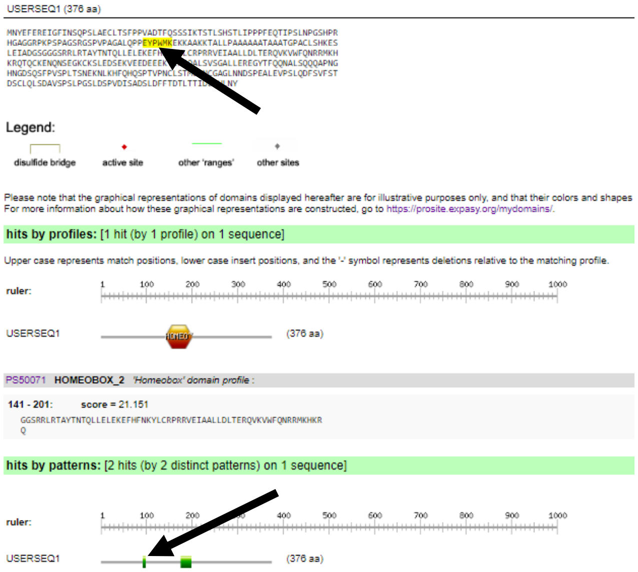
PS50071 **HOMEOBOX_2** *'Homeobox' domain profile :*

141 - 201: score = 21.151

```
GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKR
Q
```

<u>NOTE 1</u>: The amino acid sequence shown above corresponds to the homeodomain, a 60 amino acid sequence that functions as the DNA binding domain of *Hoxa2* (see references below).

2. Under “hits by profiles”, move your cursor over the first green line rectangle. The amino acids making up this domain will be highlighted in yellow (see image below).


USERSEQ1 (376 aa)

MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLNPGSHPR
HGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAAATGPACLSHKES
LEIADGSGGGSRRLRTAYTNTQLLELEKEFH[illegible]LCRPRRVEIAALLDLTERQVKVWFQNRRMKH
KRQTQCKENQNSEGKCKSLEDSEKVEEDEEEK[illegible]QALSVSGALLEREGYTFQQNALSQQQAPNG
HNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLSTM[illegible]MCGAGLNNDSPEALEVPSLQDFSVFST
DSCLQLSDAVSPSLPGSLDSPVDISADSLDFFTDTLTTID[illegible]LNY

Please note that the graphical representations of domains displayed hereafter are for illustrative purposes only, and that their colors and shapes
For more information about how these graphical representations are constructed, go to https://prosite.expasy.org/mydomains/.

hits by profiles: [1 hit (by 1 profile) on 1 sequence]

Upper case represents match positions, lower case insert positions, and the '-' symbol represents deletions relative to the matching profile.

PS50071 **HOMEOBOX_2** *'Homeobox' domain profile :*

141 - 201: score = 21.151

GGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALLDLTERQVKVWFQNRRMKHKR
Q

hits by patterns: [2 hits (by 2 distinct patterns) on 1 sequence]

NOTE 2: The amino acid sequence shown above corresponds to the hexapeptide, a 6 amino acid sequence that functions in binding to co-transcription factors, such as *Pbx* (see references below).

II. Based on the ProSite data above, develop a figure that labels the homeodomain and hexapeptide domains in the amino acid sequence alignments generated from the BIO-001 SOP.

A. Open the “Hoxa2_Figures_Powerpoint” file developed from BIO-001.

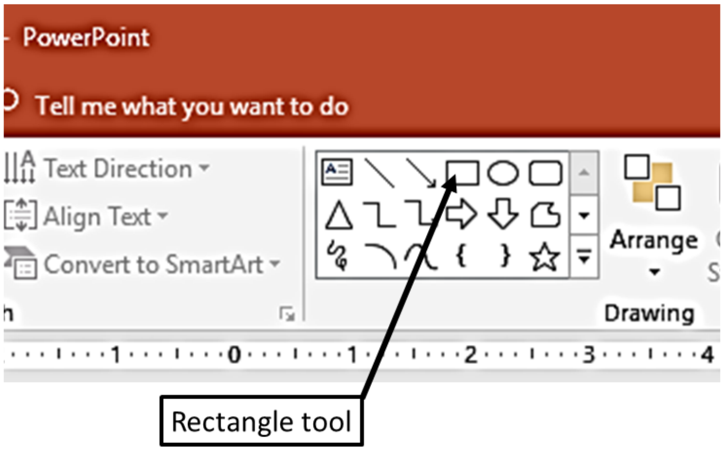
B. Label the Homeodomain region in the Model Vertebrates sequence alignment data.

NOTE 3: Since the Primate amino acid sequence alignment data shows virtually identical sequence identity among all organisms, the homeodomain and hexapeptide are not easily recognizable. They are more recognizable in the Model Vertebrates sequence alignment data because these organisms share a much more distant common ancestor.

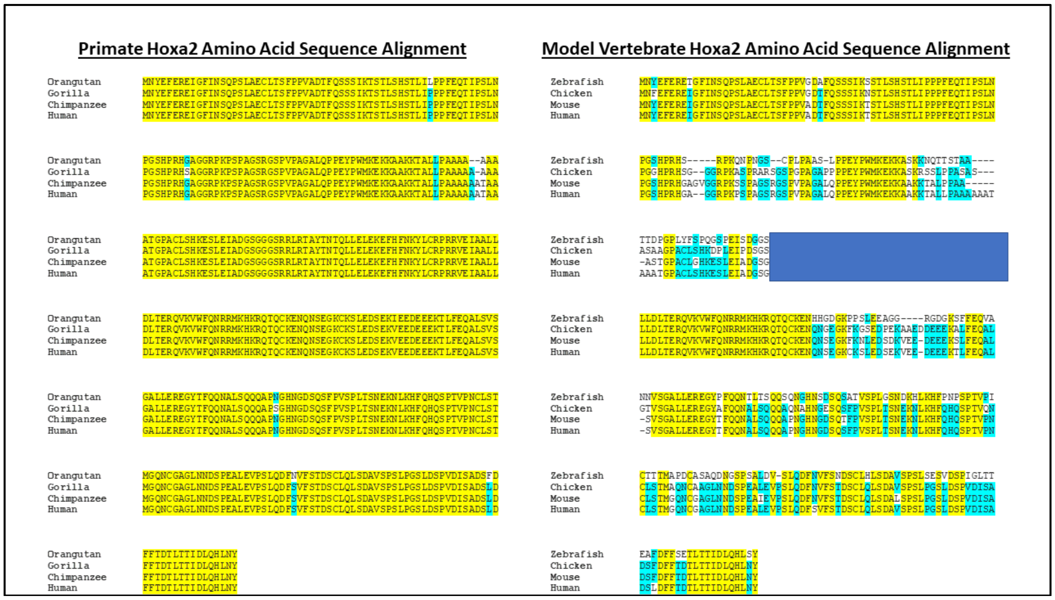
1. Label the homeodomain in the 3rd set of data within the Model Vertebrates sequence alignment data.

   a. Locate the conserved homeodomain amino acid sequence. This is roughly 60 amino acids long, starts with the amino acids, “GGSRRL” for Human, Mouse, and Chicken, and ends with the characters, “MKHKRQ” for all model vertebrates. Most of the homeodomain sequence shows 100% sequence identity among all model vertebrate organisms, except for amino acids 2 and 3 for zebrafish. Also, the homeodomain can be seen on both the end-half of the 3rd data set and the beginning of the 4th data set.

b. Under the “Home” tab, click on the rectangle tool in the “Drawing” items window. The rectangle tool is adjacent to and to the right of the arrow tool (see image below).



c. Move the mouse cursor to just above the first amino acid of the homeodomain of the first organism, click and hold the left mouse button, and drag the mouse to the right and down so that the entire portion of the homeodomain within the 3rd data set is covered with a rectangle (see image below).



Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla       PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human         PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla       GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human         GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan     MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla       MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee    MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan     FFTDTLTTIDLQHLNY
Gorilla       FFTDTLTTIDLQHLNY
Chimpanzee    FFTDTLTTIDLQHLNY
Human         FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

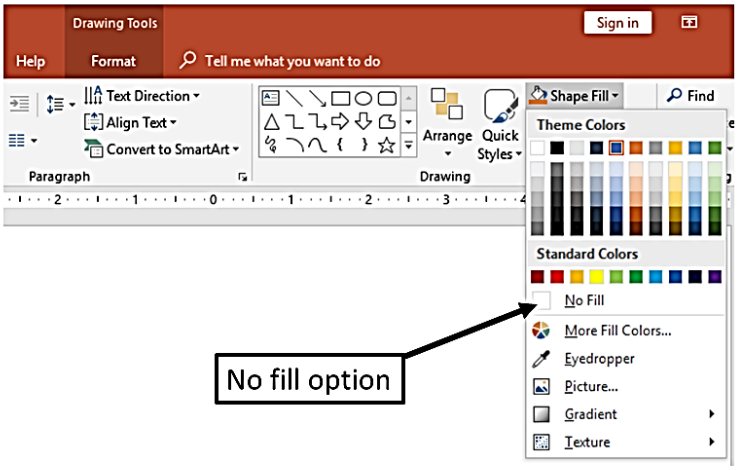
```
Zebrafish     MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken       MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish     PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken       PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse         PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human         PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish     TTDPGPLYFSPQGSPEISDGGS
Chicken       ASAAGPACLSHKDPLEIPDSGS
Mouse         -ASTGPACLGHKESLEIADGSG
Human         AAATGPACLSHKESLEIADGSG

Zebrafish     LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken       LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish     NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken       GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish     CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken       CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse         CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human         CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish     EAFDFFSETLTTIDLQHLSY
Chicken       DSFDFFTDTLTTIDLQHLNY
Mouse         DSFDFFTDTLTTIDLQHLNY
Human         DSLDFFTDTLTTIDLQHLNY
```

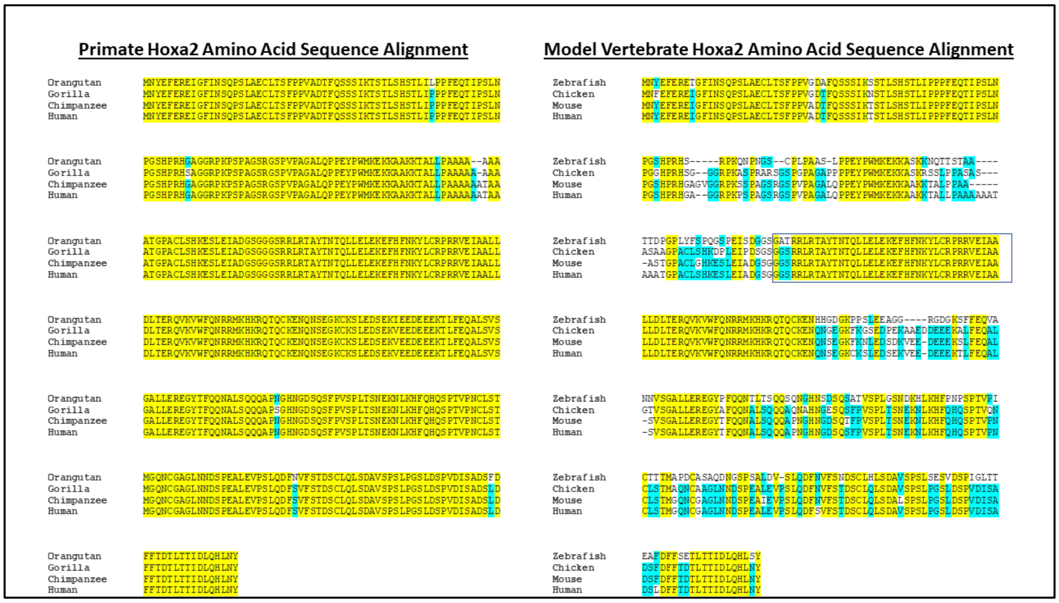
d. Change the rectangle over the homeodomain into a transparent window.

  i. Click on the rectangle. A new tab labeled "Format" will appear to the right of the tab labeled "Help".

  ii. Under the "Format" tab, click on the "Shape Fill" button. A dropdown menu with several choices will appear.

  iii. Click on the "No Fill" choice. The solid rectangle has been converted into a transparent window (see images below).





Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla       PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human         PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla       GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human         GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan     MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla       MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee    MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan     FFTDTLTTIDLQHLNY
Gorilla       FFTDTLTTIDLQHLNY
Chimpanzee    FFTDTLTTIDLQHLNY
Human         FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

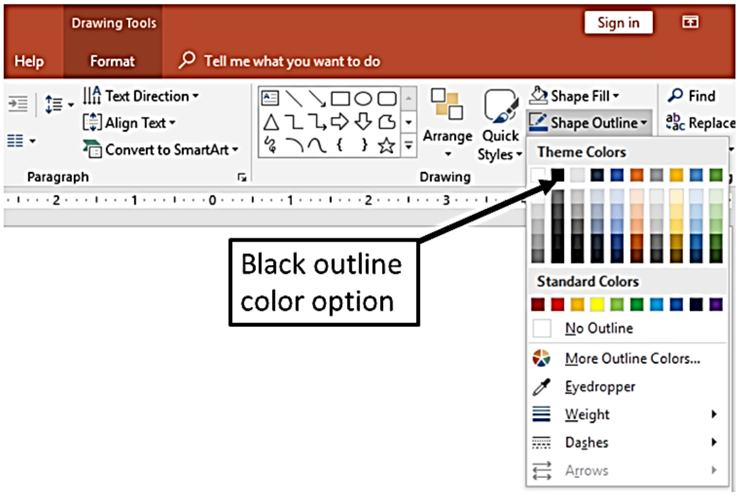
```
Zebrafish     MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken       MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish     PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken       PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse         PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human         PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish     TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken       ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse         -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human         AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish     LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken       LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish     NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken       GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish     CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken       CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse         CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human         CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish     EAFDFFSETLTTIDLQHLSY
Chicken       DSFDFFTDTLTTIDLQHLNY
Mouse         DSFDFFTDTLTTIDLQHLNY
Human         DSLDFFTDTLTTIDLQHLNY
```

e. Change the border of the transparent window so that it is colored black.

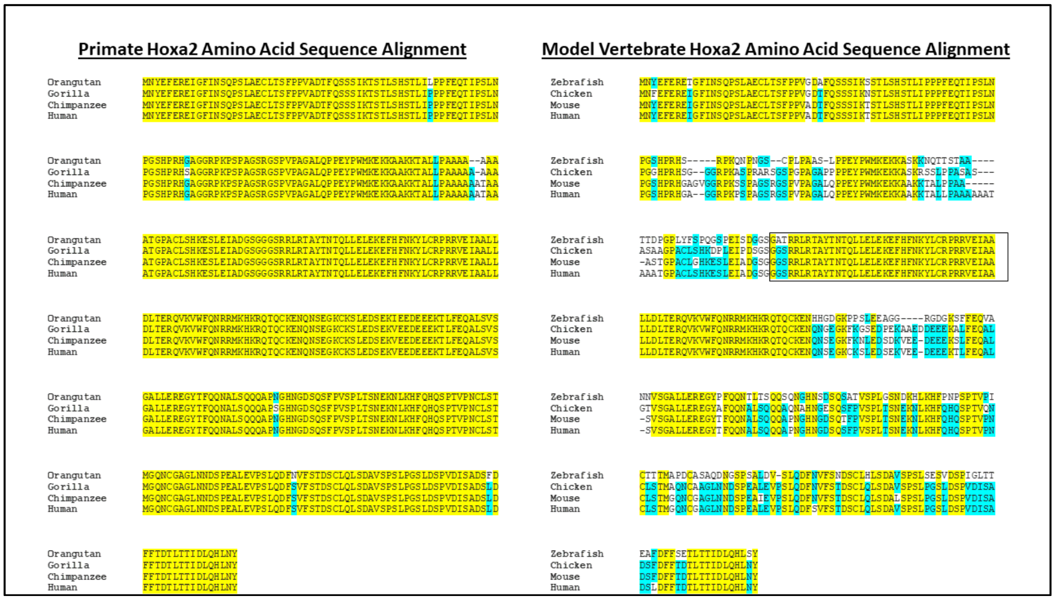
i. Click on the border of the transparent window. A new tab labeled "Format" will appear to the right of the tab labeled "Help".

ii. Under the "Format" tab, click on the "Shape Outline" button. A dropdown menu with several choices will appear.

iii. Under the "Theme colors", click on the upper-most Black box. A text box labeled "Black, Text 1" will appear when the mouse cursor is placed over this box (see images below).



Black outline color option



Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla       PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human         PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla       GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human         GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan     MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla       MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee    MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan     FFTDTLTTIDLQHLNY
Gorilla       FFTDTLTTIDLQHLNY
Chimpanzee    FFTDTLTTIDLQHLNY
Human         FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

```
Zebrafish     MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken       MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish     PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken       PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse         PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human         PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish     TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken       ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse         -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human         AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish     LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken       LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish     NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken       GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish     CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken       CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse         CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human         CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish     EAFDFFSETLTTIDLQHLSY
Chicken       DSFDFFTDTLTTIDLQHLNY
Mouse         DSFDFFTDTLTTIDLQHLNY
Human         DSLDFFTDTLTTIDLQHLNY
```

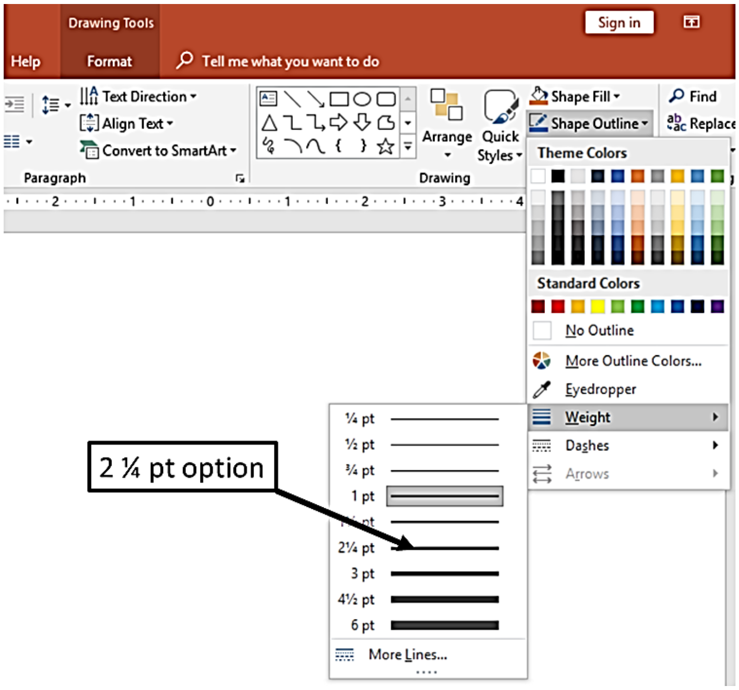
f. Change the border of the transparent window so that it has a weight of 2 ¼ pt.

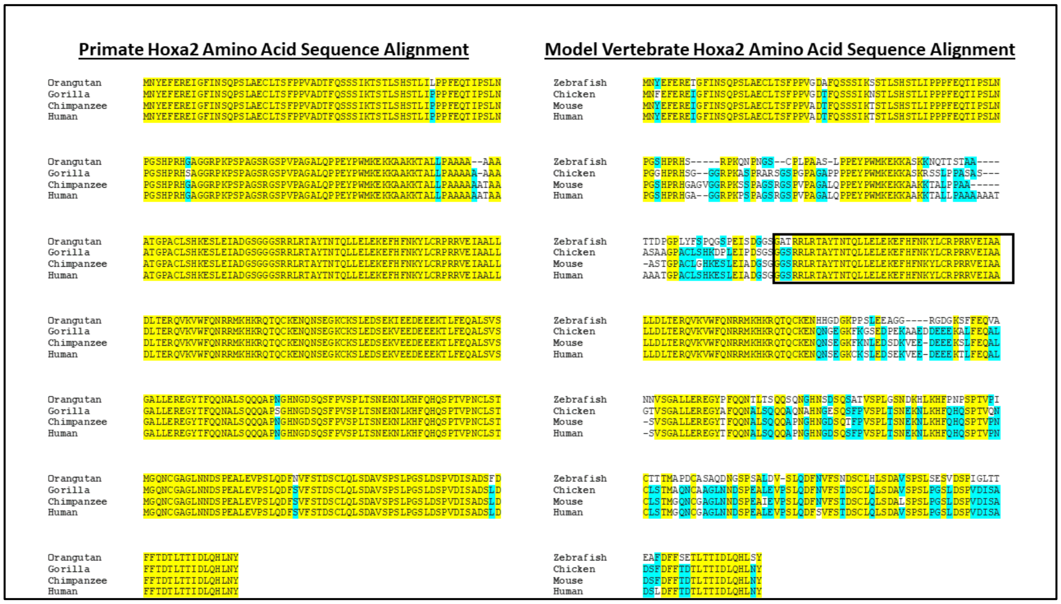
i. Click on the border of the transparent window. A new tab labeled "Format" will appear to the right of the tab labeled "Help".

ii. Under the “Format” tab, click on the “Shape Outline” button. A dropdown menu with several choices will appear.

iii. Place the mouse cursor over the “Weight” label. A popup window will automatically appear on the right showing several different lines of different weights. The “1 pt.” line should already be highlighted.

iv. Click on the button with the line that is at “2 ¼ pt” (see images below).





Primate Hoxa2 Amino Acid Sequence Alignment

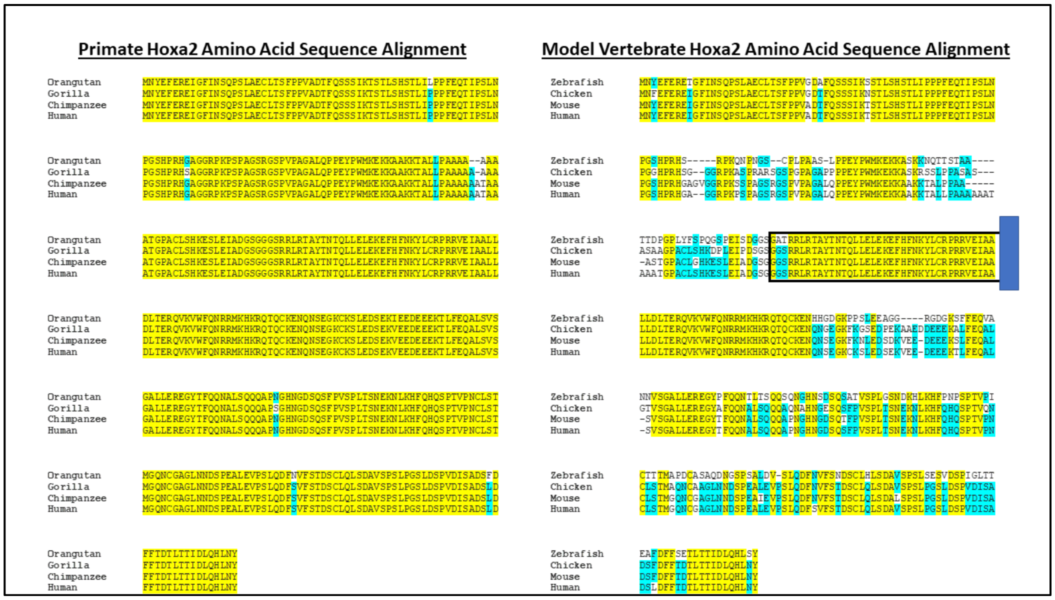
```
Orangutan    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee   MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla      PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee   PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human        PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla      ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee   ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human        ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla      DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee   DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human        DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla      GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee   GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human        GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan    MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla      MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee   MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human        MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan    FFTDTLTTIDLQHLNY
Gorilla      FFTDTLTTIDLQHLNY
Chimpanzee   FFTDTLTTIDLQHLNY
Human        FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

```
Zebrafish    MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken      MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish    PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken      PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse        PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human        PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish    TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken      ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse        -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human        AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish    LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken      LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse        LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human        LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish    NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken      GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse        -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human        -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish    CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken      CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse        CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human        CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish    EAFDFFSETLTTIDLQHLSY
Chicken      DSFDFFTDTLTTIDLQHLNY
Mouse        DSFDFFTDTLTTIDLQHLNY
Human        DSLDFFTDTLTTIDLQHLNY
```

g. Remove the right-most border of the rectangle so as to show that the homeodomain encompasses both the 3rd and 4th data sets within the model vertebrates sequence alignment.

   i. Make a rectangle following the directions in II.B.1.b.-II.B.1.c. but position and size it so that it only covers the right border of the transparent rectangle (see image below).



Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee   MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla      PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee   PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human        PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla      ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee   ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human        ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla      DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee   DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human        DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla      GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee   GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human        GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan    MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla      MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee   MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human        MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan    FFTDTLTTIDLQHLNY
Gorilla      FFTDTLTTIDLQHLNY
Chimpanzee   FFTDTLTTIDLQHLNY
Human        FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

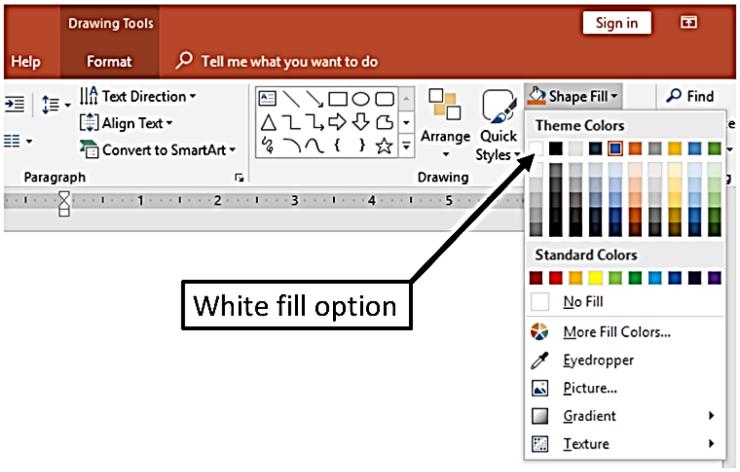
```
Zebrafish    MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken      MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish    PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken      PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse        PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human        PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish    TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken      ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse        -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human        AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish    LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken      LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse        LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human        LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish    NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken      GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse        -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human        -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish    CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken      CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse        CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human        CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish    EAFDFFSETLTTIDLQHLSY
Chicken      DSFDFFTDTLTTIDLQHLNY
Mouse        DSFDFFTDTLTTIDLQHLNY
Human        DSLDFFTDTLTTIDLQHLNY
```

   ii. Make this new rectangle a solid white color with a white border to cover the right black border of the original rectangle.

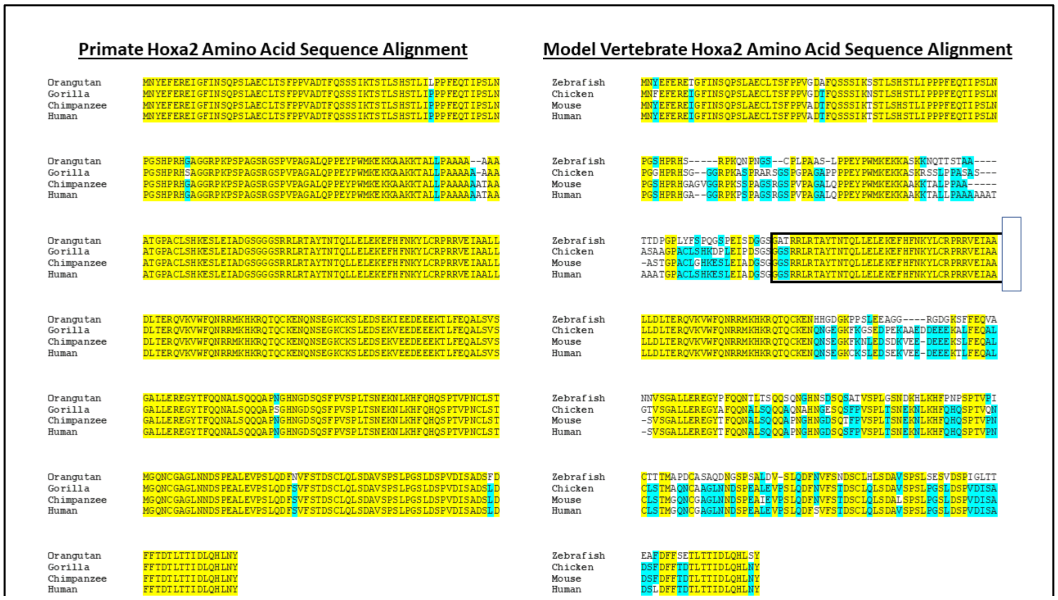
      a. Click on the rectangle. A new tab labeled “Format” will appear to the right of the tab labeled “Help”.

b. Under the “Format” tab, click on the “Shape Fill” button. A dropdown menu with several choices will appear.

c. Click on the White color choice under the “Theme Colors” menu. The solid blue rectangle now has a solid white center (see images below).



White fill option



Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human          MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan      PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla        PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human          PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan      ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla        ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human          ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan      DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla        DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human          DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan      GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla        GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human          GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan      MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla        MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee     MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human          MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan      FFTDTLTTIDLQHLNY
Gorilla        FFTDTLTTIDLQHLNY
Chimpanzee     FFTDTLTTIDLQHLNY
Human          FFTDTLTTIDLQHLNY
```

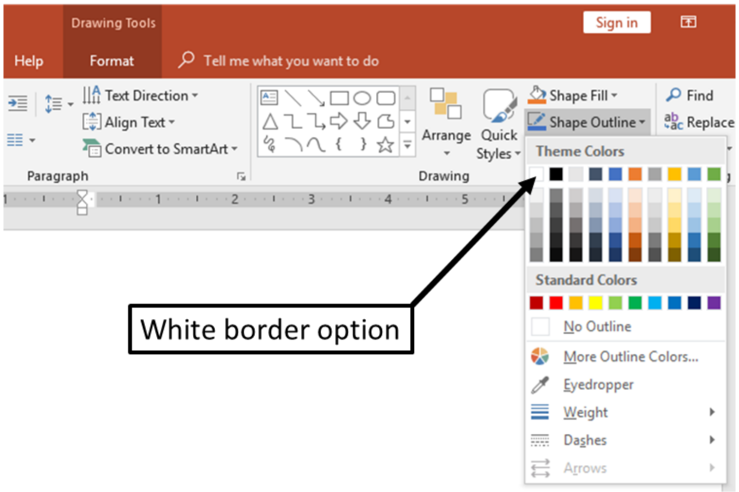
Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

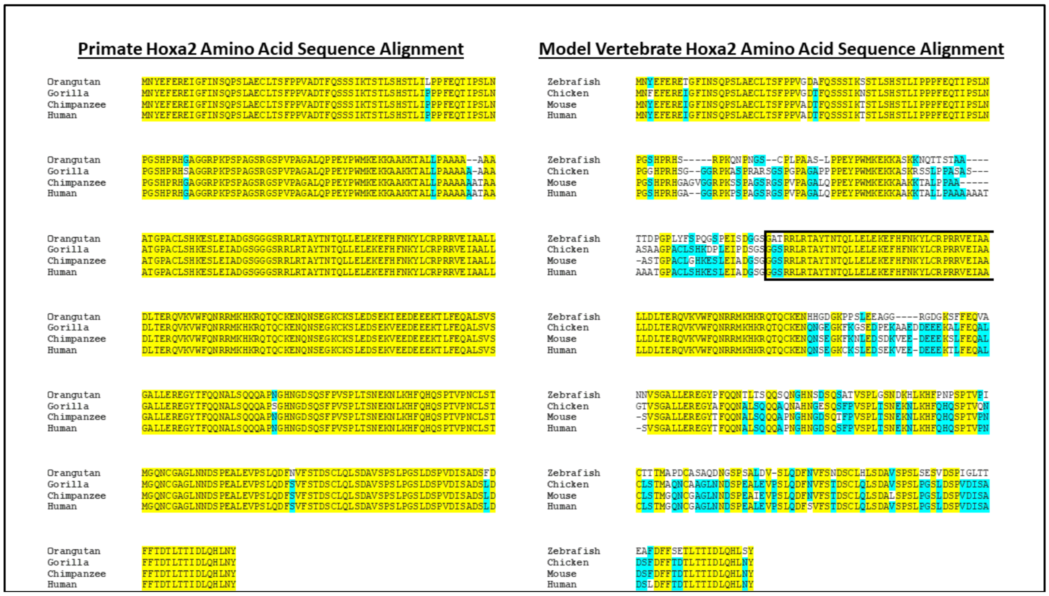
```
Zebrafish      MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken        MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse          MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human          MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish      PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken        PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse          PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human          PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish      TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken        ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse          -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human          AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish      LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken        LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse          LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human          LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish      NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken        GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse          -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human          -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish      CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken        CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse          CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human          CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish      EAFDFFSETLTTIDLQHLSY
Chicken        DSFDFFTDTLTTIDLQHLNY
Mouse          DSFDFFTDTLTTIDLQHLNY
Human          DSLDFFTDTLTTIDLQHLNY
```

d. Click on the rectangle. A new tab labeled “Format” will appear to the right of the tab labeled “Help”.

e. Under the “Format” tab, click on the “Shape Outline” button. A dropdown menu with several choices will appear.

f. Click on the White color choice under the “Theme Colors” menu. The blue border of the rectangle is now white (see images below).





Primate Hoxa2 Amino Acid Sequence Alignment

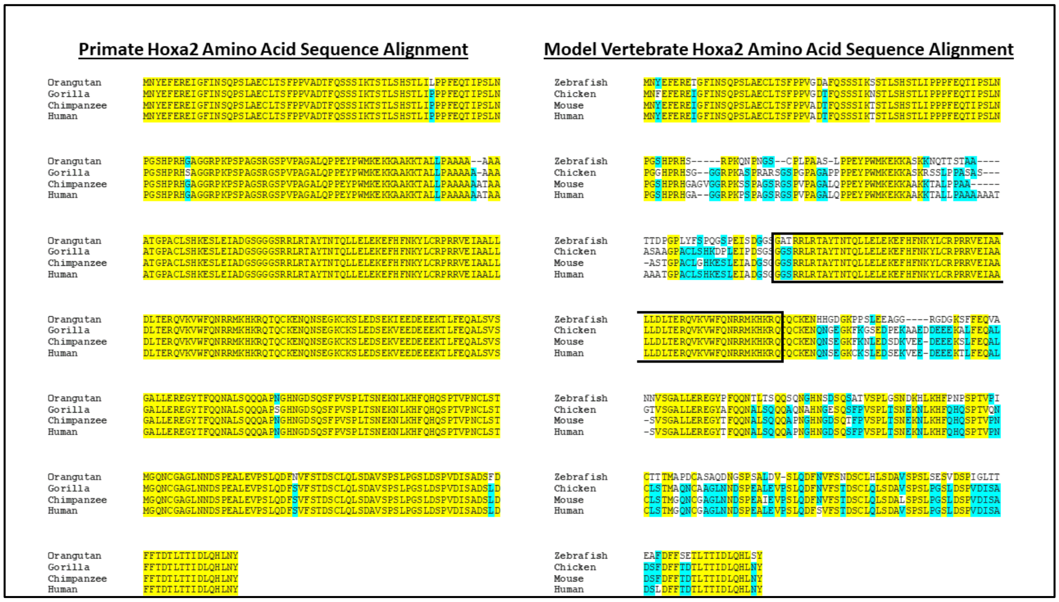
```
Orangutan    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla      MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee   MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla      PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee   PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human        PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla      ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee   ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human        ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla      DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee   DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human        DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla      GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee   GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human        GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan    MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla      MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee   MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human        MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan    FFTDTLTTIDLQHLNY
Gorilla      FFTDTLTTIDLQHLNY
Chimpanzee   FFTDTLTTIDLQHLNY
Human        FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

```
Zebrafish    MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken      MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human        MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish    PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken      PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse        PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human        PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish    TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken      ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse        -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human        AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish    LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken      LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse        LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human        LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish    NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken      GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse        -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human        -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish    CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken      CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse        CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human        CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish    EAFDFFSETLTTIDLQHLSY
Chicken      DSFDFFTDTLTTIDLQHLNY
Mouse        DSFDFFTDTLTTIDLQHLNY
Human        DSLDFFTDTLTTIDLQHLNY
```

2. Label the homeodomain in the 4th set of data within the Model vertebrates sequence alignment data.

   a. Follow steps in II.B.1.-II.B.1.g.ii.f., but take away the left-most border of the homeodomain labeling rectangle (see image below).



Primate Hoxa2 Amino Acid Sequence Alignment

```
Orangutan     MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLILPPFEQTIPSLN
Gorilla       MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Chimpanzee    MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Orangutan     PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAA--AAA
Gorilla       PGSHPRHSAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAA-AAA
Chimpanzee    PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA
Human         PGSHPRHGAGGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAATAA

Orangutan     ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Gorilla       ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Chimpanzee    ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL
Human         ATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAALL

Orangutan     DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKIEEDEEEKTLFEQALSVS
Gorilla       DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Chimpanzee    DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS
Human         DLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEEDEEEKTLFEQALSVS

Orangutan     GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Gorilla       GALLEREGYTFQQNALSQQQAPSGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Chimpanzee    GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST
Human         GALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPNCLST

Orangutan     MGQNCGAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSFD
Gorilla       MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Chimpanzee    MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD
Human         MGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISADSLD

Orangutan     FFTDTLTTIDLQHLNY
Gorilla       FFTDTLTTIDLQHLNY
Chimpanzee    FFTDTLTTIDLQHLNY
Human         FFTDTLTTIDLQHLNY
```

Model Vertebrate Hoxa2 Amino Acid Sequence Alignment

```
Zebrafish     MNYEFERETGFINSQPSLAECLTSFPPVGDAFQSSSIKSSTLSHSTLIPPPFEQTIPSLN
Chicken       MNFEFEREIGFINSQPSLAECLTSFPPVGDTFQSSSIKNSTLSHSTLIPPPFEQTIPSLN
Mouse         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN
Human         MNYEFEREIGFINSQPSLAECLTSFPPVADTFQSSSIKTSTLSHSTLIPPPFEQTIPSLN

Zebrafish     PGSHPRHS-----RPKQNPNGS--CPLPAAS-LPPEYPWMKEKKASKKNQTTSTAA----
Chicken       PGGHPRHSG--GGRPKASPRARSGSPGPAGAPPPPEYPWMKEKKASKRSSLPPASAS---
Mouse         PGSHPRHGAGVGGRPKSSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALPPAA-----
Human         PGSHPRHGA--GGRPKPSPAGSRGSPVPAGALQPPEYPWMKEKKAAKKTALLPAAAAAAT

Zebrafish     TTDPGPLYFSPQGSPEISDGGSGATRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Chicken       ASAAGPACLSHKDPLEIPDSGSGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Mouse         -ASTGPACLGHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA
Human         AAATGPACLSHKESLEIADGSGGGSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRVEIAA

Zebrafish     LLDLTERQVKVWFQNRRMKHKRQTQCKENHHGDGKPPSLEEAGG----RGDGKSFFEQVA
Chicken       LLDLTERQVKVWFQNRRMKHKRQTQCKENQNGEGKFKGSEDPEKAAEDDEEEKALFEQAL
Mouse         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKFKNLEDSDKVEE-DEEEKSLFEQAL
Human         LLDLTERQVKVWFQNRRMKHKRQTQCKENQNSEGKCKSLEDSEKVEE-DEEEKTLFEQAL

Zebrafish     NNVSGALLEREGYPFQQNTLTSQQSQNGHNSDSQSATVSPLGSNDKHLKHFPNPSPTVPI
Chicken       GTVSGALLEREGYAFQQNALSQQQAQNAHNGESQSFPVSPLTSNEKNLKHFQHQSPTVQN
Mouse         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQTFPVSPLTSNEKNLKHFQHQSPTVPN
Human         -SVSGALLEREGYTFQQNALSQQQAPNGHNGDSQSFPVSPLTSNEKNLKHFQHQSPTVPN

Zebrafish     CTTTMAPDCASAQDNGSPSALDV-SLQDFNVFSNDSCLHLSDAVSPSLSESVDSPIGLTT
Chicken       CLSTMAQNCAAGLNNDSPEALEVPSLQDFNVFSTDSCLQLSDAVSPSLPGSLDSPVDISA
Mouse         CLSTMGQNCGAGLNNDSPEAIEVPSLQDFNVFSTDSCLQLSDALSPSLPGSLDSPVDISA
Human         CLSTMGQNCGAGLNNDSPEALEVPSLQDFSVFSTDSCLQLSDAVSPSLPGSLDSPVDISA

Zebrafish     EAFDFFSETLTTIDLQHLSY
Chicken       DSFDFFTDTLTTIDLQHLNY
Mouse         DSFDFFTDTLTTIDLQHLNY
Human         DSLDFFTDTLTTIDLQHLNY
```

3. Label the homeodomain region.

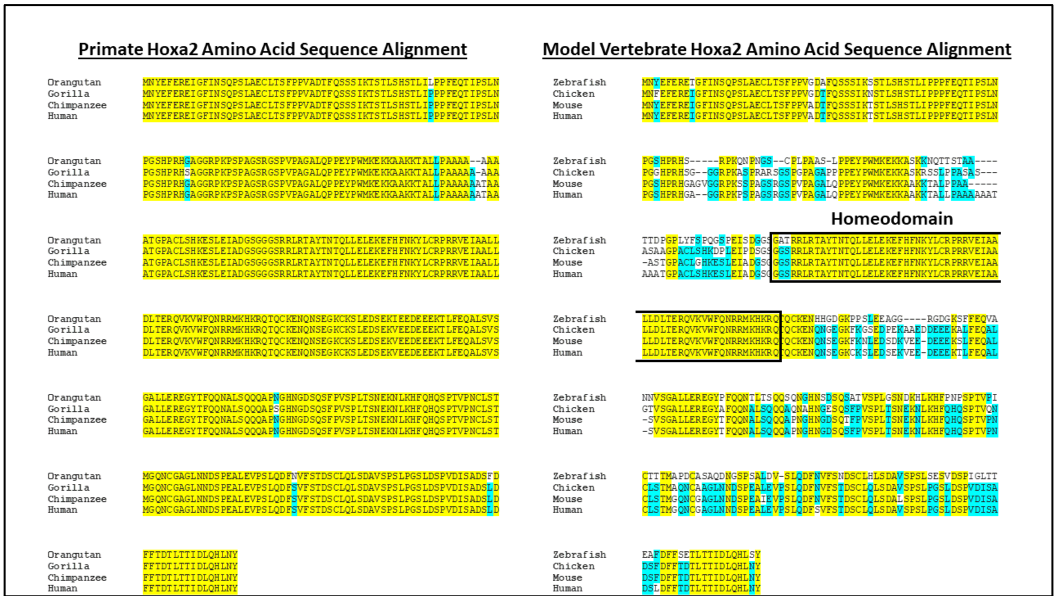
    a. Follow steps VIII.D.3.a.-VIII.D.3.b. from the BIO-001 SOP to create a text box.

    b. Move the cursor above the transparent window surrounding the homeodomain region in the 3rd data set and left click the mouse button. A text box will automatically appear where the mouse was clicked.

    c. Type "Homeodomain" and position the text above the center of the transparent window.

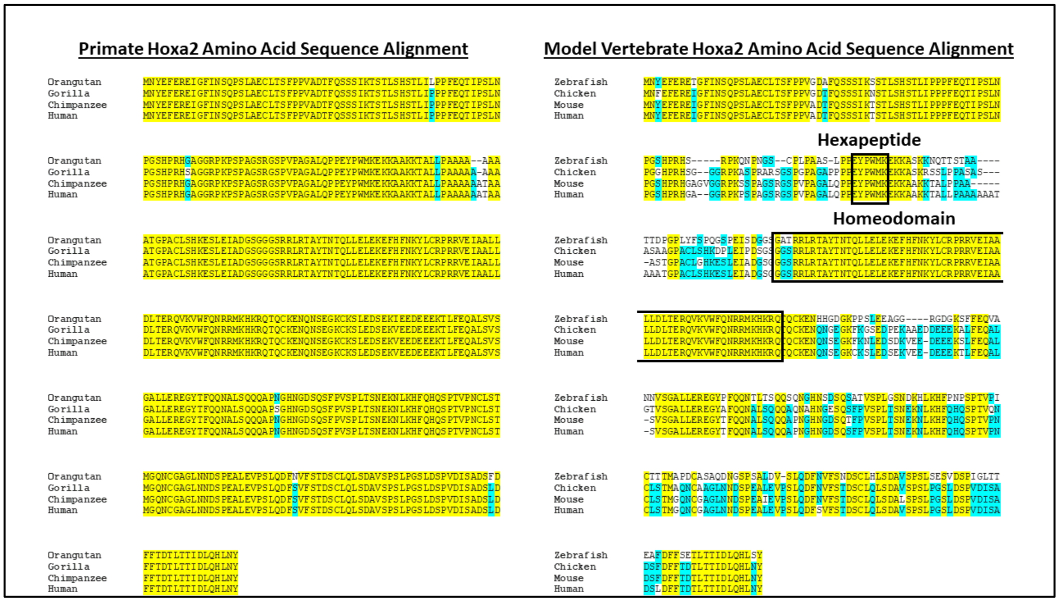
    d. Keep the Font type and size as is or change it to your liking.

    e. Bold all Text (see image below).



C. Label the Hexapeptide domain in the Model vertebrates sequence alignment data.

1. Locate the conserved hexapeptide amino acid sequence. This is a 6 amino acid long domain in the following sequence “EYPWMK”. This sequence is embedded in a larger sequence just upstream of the homeodomain and shows 100% identity among all vertebrate model organisms.
2. Follow steps II.B.1.b-II.B.1.f.iv. to make a transparent rectangle.
3. Follow steps II.B.3.a.-II.B.3.e. to make the Hexapeptide label (see image below).



III. Save the Microsoft PowerPoint file as “Hoxa2_Figures_Powerpoint”. More figures from later BIO SOPs will be made in this file.

IV. Convert the sequence alignment data into a TIFF image file.

A. Click the “File” tab.

B. Under the “File” tab, click on “Save As”.

C. Once a destination folder is chosen, click on the dropdown menu to the right of the text, “Save as type:”.

D. Choose “TIFF Tag Image File Format”.

NOTE 4: TIFF files offer high resolution data for figures.

E. Label the file “Hoxa2_Figure_1”.

F. A new window will pop up asking, “Which slides do you want to export?”. Click on the button labeled, “Just This One”.

V. Develop a Word file with the figure created from this SOP and other later SOPs.

A. Open a new Microsoft Word document.

B. Click on the “Hoxa2_Figure_1” TIFF file and drag it into the Word document.

C. Underneath the figure, write a legend that describes:

1. The layout of the figure (describe the alignments on each side of the figure)
2. The methods used to develop this figure (e.g.: what software programs were used to obtain amino acid sequences, align amino acid sequences, identify the functional domains, etc.)
3. The yellow and blue color-coding process of the amino acid sequence alignments.

D. Save the Word file as “Hoxa2_Figures_Word”. More figures from other BIO SOPs will be made in this file.

VI. Turn in the “Hoxa2_Figures_Word” file according to the deadline set by the instructor.

References:

Chang, C.P., L. Brocchieri, W.F. Shen, C. Largman, and M.L. Cleary. 1996. Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Mol Cell Biol 167(4): 1734-1745.

Gendron-Maguire, M., M. Mallo, M. Zhang, T. Gridley. 1993. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75(7):1317-1331.

LaRonde-LeBlanc, N.A. and C. Wolberger. 2003. Structure of Hoxa9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. Genes Dev 17(16):2060-2072.

Piper, D.E., A.H. Batchelor, C.P. Chang, M.L. Cleary, and C. Wolberger. 1999. Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell 96(4):587-597.

Rijli, F.M., M. Mark, S. Lakkaraju, A. Dierich, P. Dolle, and P. Chambon. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75(7):1333-1349.

Sigrist, C.J.A., L. Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch, and P. Bucher. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform 3:265-274.

SOP #	BIO-002
Title	Analysis of Hoxa2 amino acid sequences II: Identification and labeling of Functional Hoxa2 Protein Domains.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

SOP Assessment

1. The homeodomain functions to:

    a. Bind to co-activating transcription factors
    b. Bind to *cis*-regulatory elements (CREs) to activate or repress translation of downstream genes
    c. Bind to *cis*-regulatory elements (CREs) to activate or repress transcription of downstream genes
    d. Bind to splicing elements for RNA processing

2. The hexapeptide functions to:

    a. Bind to co-activating transcription factors
    b. Bind to *cis*-regulatory elements (CREs) to activate or repress translation of downstream genes
    c. Bind to *cis*-regulatory elements (CREs) to activate or repress transcription of downstream genes
    d. Bind to splicing elements for RNA processing

3. The homeodomain and hexapeptide can be visualized more easily in the Model Vertebrates amino acid alignment than the Primates alignment because:

    a. The model vertebrates share a more distant common ancestor than the primates
    b. The model vertebrates share a more recent common ancestor than the primates
    c. The primates share a more distant common ancestor than the model vertebraes
    d. The model vertebrates have bigger, meaner domains in their proteins than the primates

Date: ______________________________

Name (Print): ______________________________

Signature: ______________________________

SOP #	BIO-003
Title	Analysis of *Hoxa2* Genomic DNA Sequences I: Use of mVISTA, Microsoft Word, and Microsoft PowerPoint Software to Analyze Conserved Coding and Intergenic DNA.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

Objective:

To understand the function and evolution of *Hoxa2*, the analysis of conserved coding and intergenic DNA must be performed. Sequence alignments between evolutionarily divergent species will help to reveal conserved and functional *cis*-regulatory elements (CREs). CREs within genomic DNA that are conserved for over millions of years and between species indicate that there is functional importance with these sequences. Such regions can be tested to understand the overall functions of genes and how medical disorders develop. Genomic DNA sequences containing the *Hoxa2* exons, intron, and 3000 bp upstream of *Hoxa2* will be retrieved from several primate species, including Human (*Homo sapiens*), Chimpanzee (*Pan troglodytes*), Gorilla (*Gorilla gorilla*), and Orangutan (*Pongo abelii*) and several vertebrate biomedical models, including Zebrafish (*Danio rerio*), Chicken (*Gallus gallus*), and Mouse (*Mus musculus*).

Relevant Terms and their Definitions

Hoxa2 – Developmental regulatory gene that is evolutionarily conserved and functions to pattern the development of the facial nerve.

Exons – Portions of genes that are transcribed and give rise to amino acids

Introns – Portions of genes that are transcribed but are excised from the transcript, and do not give rise to amino acids.

Cis-Regulatory Elements (CREs) – Genomic DNA sequences that function to direct the spatial and temporal expression of genes.

National Center for Biotechnology Information (NCBI) – Web-based database containing genetic information submitted by scientists. Used for genetic, developmental, medical, ecological and evolutionary research analyses.

Genbank Accession number – Identifying number for amino acid sequences

FASTA – Sequence format that must be obtained for nucleotide sequences for downstream analyses.

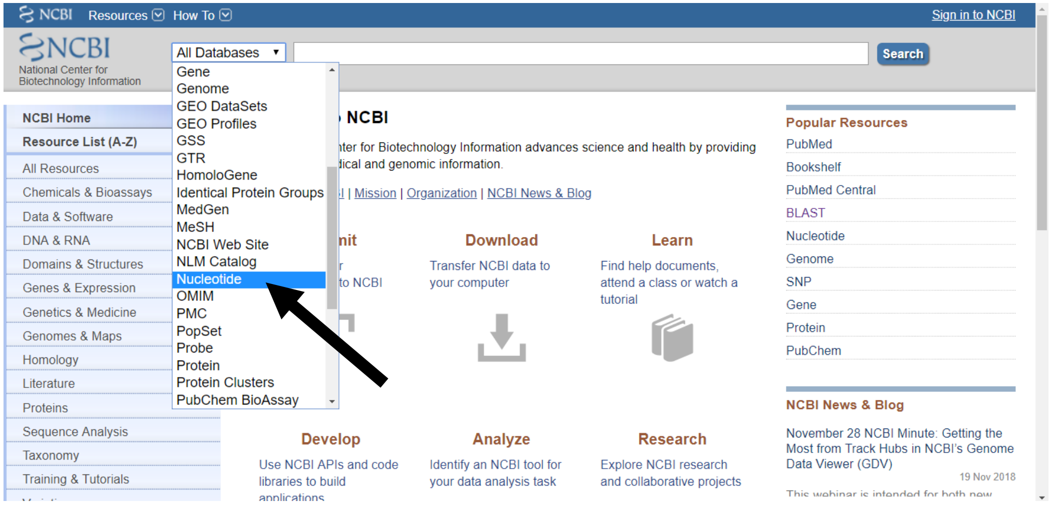
mVISTA - Software program that generates global and graphical views of sequence comparisons between multiple species.

Procedure

I. Retrieve primate and model vertebrate *Hoxa2* genomic DNA sequences from the Genbank Database.

A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), type in https://www.ncbi.nlm.nih.gov

B. Click on the drop down menu titled “All Databases” and select “Nucleotide” (see image below).



C. Type in the appropriate Genbank accession number in the text box next to the Drop-down menu to retrieve the appropriate genomic DNA sequences and click on the “Search” button. The Genbank accession numbers for the species-specific *Hoxa2* genomic DNA sequences are listed in the table below.

Organism	*Hoxa2* Genbank Accession	Sequence Begin Position	Sequence End Position	Reverse Complement Necessary
Human	AC004079	83121	87895	Yes
Chimpanzee	NC_036886	27217325	27222099	Yes
Gorilla	NC_018431	27155539	27160306	Yes
Orangutan	NC_036910	46400034	46404802	No
Mouse	CH466597	4372386	4377144	Yes
Chicken	NC_006089	32582800	32587527	Yes
Zebrafish	AL645795	56638	61314	Yes

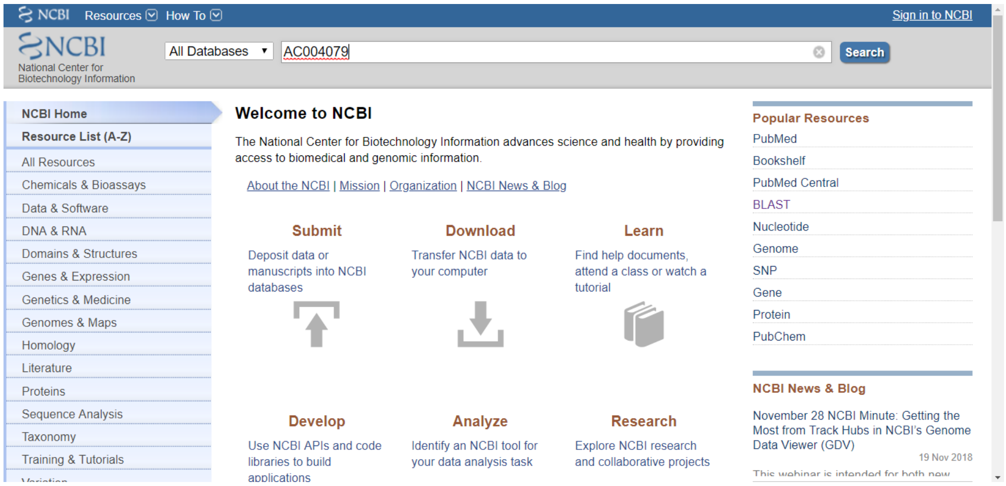
D. Once the Genbank sequence information is displayed, retrieve the designated region of genomic DNA associated with *Hoxa2* in FASTA format and in the 5’-3’ orientation.

NOTE 1: FASTA format is necessary for future downstream analyses, including DNA sequence alignment.

NOTE 2: The Human *Hoxa2* sequence (Accession #: AC004079) will be used as an example for the remainder of this SOP. This genomic sequence is from the entire chromosome 17, which is 102,717 bp in length. Due to the extreme length, it is computationally too intensive to show the entire sequence. For this reason, sequence start and end positions are listed in the table on page 2 of this SOP.

1. Type "AC004079" in the text box to the right of the dropdown menu.

2. Click on the "Search" button to the right of the textbox (see image below).



All Databases: AC004079

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

Deposit data or manuscripts into NCBI databases

Transfer NCBI data to your computer

Find help documents, attend a class or watch a tutorial

Use NCBI APIs and code libraries to build applications

Identify an NCBI tool for your data analysis task

Explore NCBI research and collaborative projects

November 28 NCBI Minute: Getting the Most from Track Hubs in NCBI's Genome Data Viewer (GDV)

19 Nov 2018

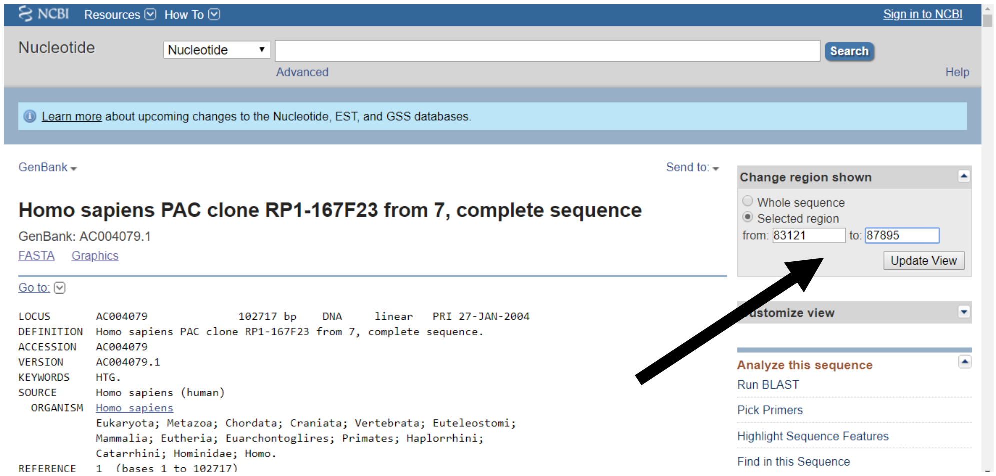
3. The Human *Hoxa2* Genomic DNA sequence Genbank information will be displayed. Click on the Down arrow next to the words "Change Region Shown" on the right side of the screen.

4. Click on "Selected region".

5. Type in the sequence begin position (for Human – 83121) from the table above in the box labeled "begin".

6. Type in the sequence end position (for Human – 87895) from the table above in the box labeled "end".

7. Click on the "Update View" button (see image below).



Learn more about upcoming changes to the Nucleotide, EST, and GSS databases.

Change region shown

○ Whole sequence
● Selected region
from: 83121 to: 87895

Homo sapiens PAC clone RP1-167F23 from 7, complete sequence

GenBank: AC004079.1

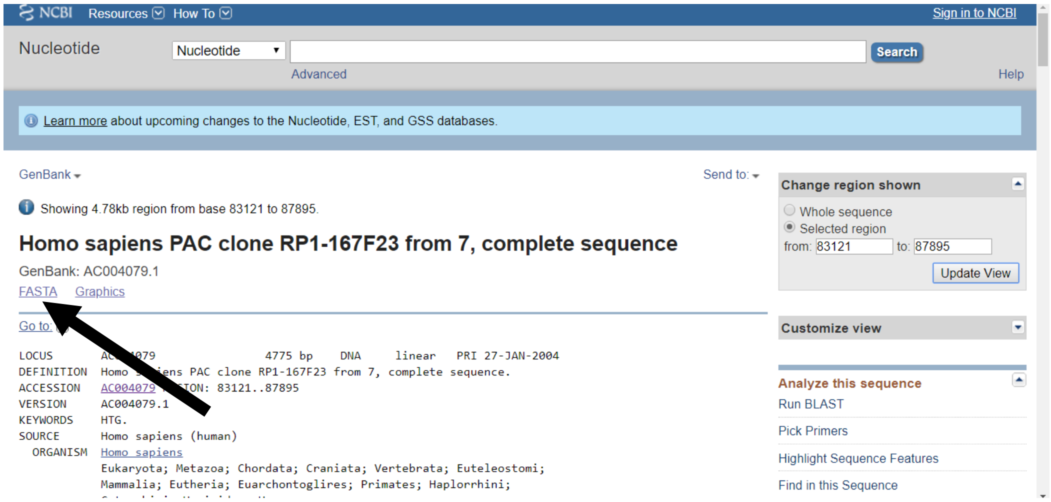
FASTA Graphics

Go to:

```
LOCUS       AC004079               102717 bp    DNA     linear   PRI 27-JAN-2004
DEFINITION  Homo sapiens PAC clone RP1-167F23 from 7, complete sequence.
ACCESSION   AC004079
VERSION     AC004079.1
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102717)
```

NOTE 3: A new screen will be displayed showing just the length of sequence specified (for Human it is 4775 bp). For this SOP, only a portion of the specified genomic DNA corresponding to *Hoxa2* will be displayed.

E. Click on "FASTA" on the upper left corner of the screen. The FASTA sequence format removes all identifying information from the sequence file (see images below).



Learn more about upcoming changes to the Nucleotide, EST, and GSS databases.

Showing 4.78kb region from base 83121 to 87895.

Change region shown

○ Whole sequence
● Selected region
from: 83121 to: 87895

Homo sapiens PAC clone RP1-167F23 from 7, complete sequence

GenBank: AC004079.1

FASTA Graphics

Go to:

```
LOCUS       AC  079                  4775 bp    DNA     linear   PRI 27-JAN-2004
DEFINITION  Homo   iens PAC clone RP1-167F23 from 7, complete sequence.
ACCESSION   AC004079   ION: 83121..87895
VERSION     AC004079.1
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
```

F. If necessary, change the orientation of the sequence so that it is displayed in the 5'-3' orientation.

NOTE 4: Since DNA is double-stranded and both strands are complementary to each other, separate genes can be located on either strand. Depending on how the genomic DNA sequence (normally an entire chromosomal DNA sequence) was loaded into Genbank, the sequence of interest could be in either orientation. In order to perform sequence alignments, all sequences retrieved for this analysis must be in the 5'-3' orientation. **Reverse complements of the genomic DNA**

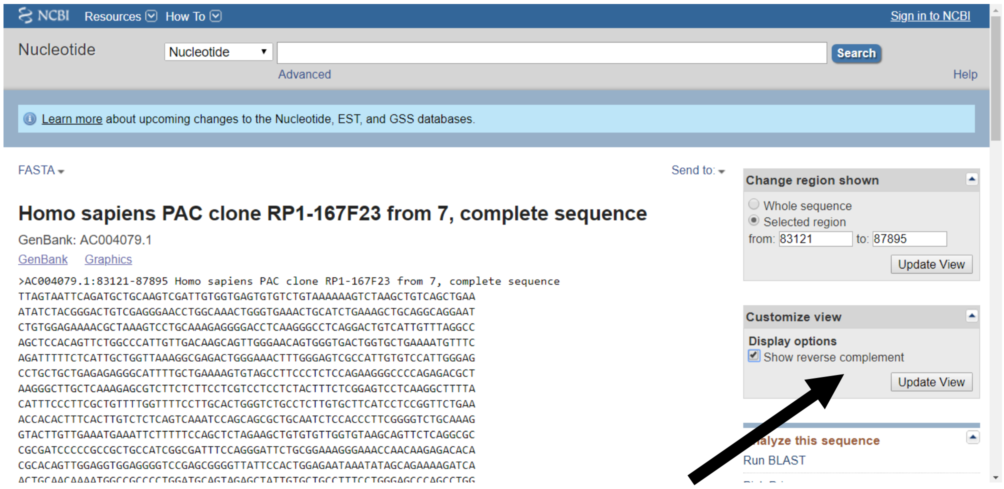
corresponding to *Hoxa2* must be obtained for ALL organisms except for Orangutan (see Table above).

NOTE 5: The human sequence starts with the bases "TTA…", which is the Stop codon for *Hoxa2* in reverse complement orientation. This must be viewed as "…TAA".

1. Click on the Down arrow next to the words "Customize view" on the right side of the screen (see image below).

2. Under "Display options", click on "Show reverse complement".

3. Click on the "Update View" button (See image below).



Learn more about upcoming changes to the Nucleotide, EST, and GSS databases.

Homo sapiens PAC clone RP1-167F23 from 7, complete sequence

GenBank: AC004079.1

GenBank Graphics

```
>AC004079.1:83121-87895 Homo sapiens PAC clone RP1-167F23 from 7, complete sequence
TTAGTAATTCAGATGCTGCAAGTCGATTGTGGTGAGTGTGTCTGTAAAAAAGTCTAAGCTGTCAGCTGAA
ATATCTACGGGACTGTCGAGGGAACCTGGCAAACTGGGTGAAACTGCATCTGAAAGCTGCAGGCAGGAAT
CTGTGGAGAAAACGCTAAAGTCCTGCAAAGAGGGGACCTCAAGGGCCTCAGGACTGTCATTGTTTAGGCC
AGCTCCACAGTTCTGGCCCATTGTTGACAAGCAGTTGGGAACAGTGGGTGACTGGTGCTGAAAATGTTTC
AGATTTTTCTCATTGCTGGTTAAAGGCGAGACTGGGAAACTTTGGGAGTCGCCATTGTGTCCATTGGGAG
CCTGCTGCTGAGAGAGGGCATTTTGCTGAAAAGTGTAGCCTTCCCTCTCCAGAAGGGCCCCAGAGACGCT
AAGGGCTTGCTCAAAGAGCGTCTTCTCTTCCTCGTCCTCCTCTACTTTCTCGGAGTCCTCAAGGCTTTTA
CATTTCCCTTCGCTGTTTTGGTTTTCCTTGCACTGGGTCTGCCTCTTGTGCTTCATCCTCCGGTTCTGAA
ACCACACTTTCACTTGTCTCTCAGTCAAATCCAGCAGCGCTGCAATCTCCACCCTTCGGGGTCTGCAAAG
GTACTTGTTGAAATGAAATTCTTTTTCCAGCTCTAGAAGCTGTGTGTTGGTGTAAGCAGTTCTCAGGCGC
CGCGATCCCCCGCCGCTGCCATCGGCGATTTCCAGGGATTCTGCGGAAAGGGAAACCAACAAGAGACACA
CGCACAGTTGGAGGTGGAGGGGTCCGAGCGGGGTTATTCCACTGGAGAATAAATATAGCAGAAAAGATCA
ACTGCAACAAAATGGCCGCCCCTGGATGCAGTAGAGCTATTGTGCTGCCTTTCCTGGGAGCCCAGCCTGG
```

Change region shown

○ Whole sequence
● Selected region
from: 83121 to: 87895
Update View

Customize view

Display options
☑ Show reverse complement
Update View

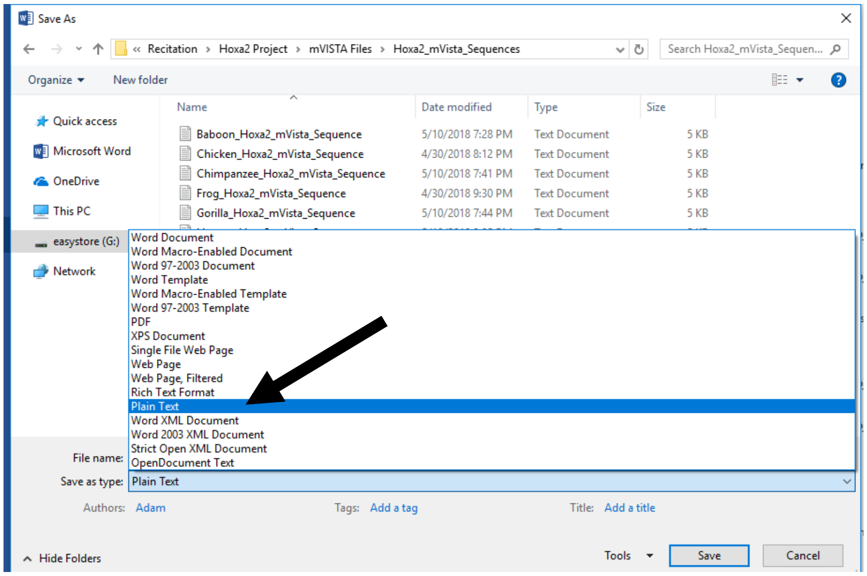
Analyze this sequence

G. Highlight the entire sequence, copy it, paste it into a new word document, and save it as a separate Plain Text document.

NOTE 6: Plain text format is necessary for the mVISTA software for reading genomic DNA sequences.

1. Save the file as "Human_Hoxa2_mVista_Sequence".

2. Develop a new folder on your computer and name it "Hoxa2_mVista_Sequences". All mVista sequence files for each organism will be saved to this folder (see image below).

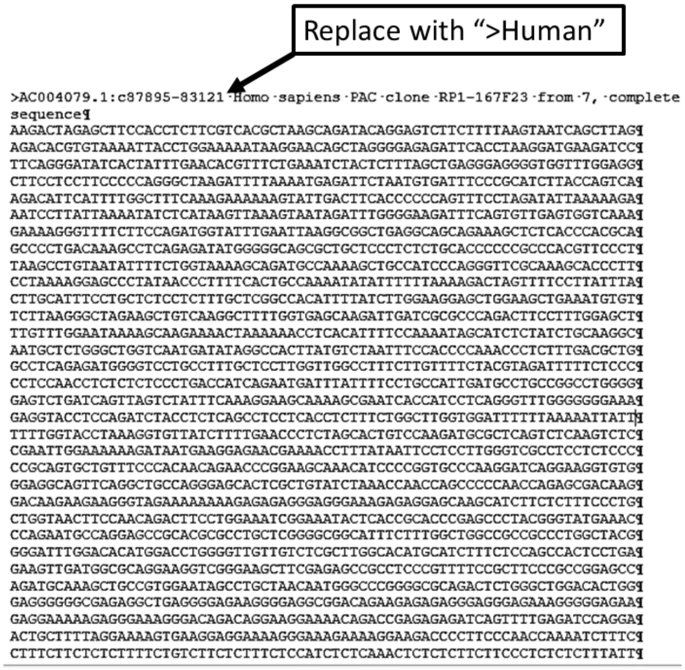


Name	Date modified	Type	Size
Baboon_Hoxa2_mVista_Sequence	5/10/2018 7:28 PM	Text Document	5 KB
Chicken_Hoxa2_mVista_Sequence	4/30/2018 8:12 PM	Text Document	5 KB
Chimpanzee_Hoxa2_mVista_Sequence	5/10/2018 7:41 PM	Text Document	5 KB
Frog_Hoxa2_mVista_Sequence	4/30/2018 9:30 PM	Text Document	5 KB
Gorilla_Hoxa2_mVista_Sequence	5/10/2018 7:44 PM	Text Document	5 KB

File name:
Save as type: Plain Text
Authors: Adam
Tags: Add a tag
Title: Add a title

II. Format the "Human_Hoxa2_mVista_Sequence" so that it can be read and processed by the mVISTA software program.

A. Replace all information after the carrot symbol (>) in the sequence identification line with the word "Human" (see image below).



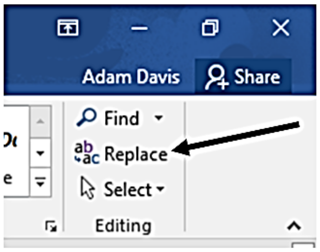
```
>AC004079.1:c87895-83121·Homo·sapiens·PAC·clone·RP1-167F23·from·7,·complete
sequence¶
AAGACTAGAGCTTCCACCTCTTCGTCACGCTAAGCAGATACAGGAGTCTTCTTTTAAGTAATCAGCTTAG¶
AGACACGTGTAAAATTACCTGGAAAAATAAGGAACAGCTAGGGGAGAGATTCACCTAAGGATGAAGATCC¶
TTCAGGGATATCACTATTTGAACACGTTTCTGAAATCTACTCTTTAGCTGAGGGAGGGGTGGTTTGGAGG¶
CTTCCTCCTTCCCCCAGGGCTAAGATTTTAAAATGAGATTCTAATGTGATTTCCCGCATCTTACCAGTCA¶
AGACATTCATTTTGGCTTTCAAAGAAAAAAGTATTGACTTCACCCCCCAGTTTCCTAGATATTAAAAAGA¶
AATCCTTATTAAAATATCTCATAAGTTAAAGTAATAGATTTGGGGAAGATTTCAGTGTTGAGTGGTCAAA¶
GAAAAGGGTTTTCTTCCAGATGGTATTTGAATTAAGGCGGCTGAGGCAGCAGAAAGCTCTCACCCACGCA¶
GCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGCACCCCCCGCCCACGTTCCCT¶
TAAGCCTGTAATATTTTCTGGTAAAAGCAGATGCCAAAAGCTGCCATCCCAGGGTTCGCAAAGCACCCTT¶
CCTAAAAGGAGCCCTATAACCCTTTTCACTGCCAAAATATATTTTTTAAAAGACTAGTTTTCCTTATTTA¶
CTTGCATTTCCTGCTCTCCTCTTTGCTCGGCCACATTTTATCTTGGAAGGAGCTGGAAGCTGAAATGTGT¶
TCTTAAGGGCTAGAAGCTGTCAAGGCTTTTGGTGAGCAAGATTGATCGCGCCCAGACTTCCTTTGGAGCT¶
TTGTTTGGAATAAAAGCAAGAAAACTAAAAAACCTCACATTTTCCAAAATAGCATCTCTATCTGCAAGGC¶
AATGCTCTGGGCTGGTCAATGATATAGGCCACTTATGTCTAATTTCCACCCCAAACCCTCTTTGACGCTG¶
GCCTCAGAGATGGGGTCCTGCCTTTGCTCCTTGGTTGGCCTTTCTTGTTTTCTACGTAGATTTTTCTCCC¶
CCTCCAACCTCTCTCTCCCTGACCATCAGAATGATTTATTTTCCTGCCATTGATGCCTGCCGGCCTGGGG¶
GAGTCTGATCAGTTAGTCTATTTCAAAGGAAGCAAAAGCGAATCACCATCCTCAGGGTTTGGGGGGGAAA¶
GAGGTACCTCCAGATCTACCTCTCAGCCTCCTCACCTCTTTCTGGCTTGGTGGATTTTTTAAAAATTATT¶
TTTTGGTACCTAAAGGTGTTATCTTTTGAACCCTCTAGCACTGTCCAAGATGCGCTCAGTCTCAAGTCTC¶
CGAATTGGAAAAAAGATAATGAAGGAGAACGAAAACCTTTATAATTCCTCCTTGGGTCGCCTCCTCTCCC¶
CCGCAGTGCTGTTTCCCACAACAGAACCCGGAAGCAAACATCCCCGGTGCCCAAGGATCAGGAAGGTGTG¶
GGAGGCAGTTCAGGCTGCCAGGGAGCACTCGCTGTATCTAAACCAACCAGCCCCCAACCAGAGCGACAAG¶
GACAAGAAGAAGGGTAGAAAAAAAAAGAGAGAGGGAGGGAAAGAGAGGAGCAAGCATCTTCTCTTTCCCTG¶
CTGGTAACTTCCAACAGACTTCCTGGAAATCGGAAATACTCACCGCACCCGAGCCCTACGGGTATGAAAC¶
CCAGAATGCCAGGAGCCGCACGCGCCTGCTCGGGGCGGCATTTCTTTGGCTGGCCGCCGCCCTGGCTACG¶
GGGATTTGGACACATGGACCTGGGGTTGTTGTCTCGCTTGGCACATGCATCTTTCTCCAGCCACTCCTGA¶
GAAGTTGATGGCGCAGGAAGGTCGGGAAGCTTCGAGAGCCGCCTCCCGTTTTCCGCTTCCCGCCGGAGCC¶
AGATGCAAAGCTGCCGTGGAATAGCCTGCTAACAATGGGCCCGGGGCGCAGACTCTGGGCTGGACACTGG¶
GAGGGGGGCGAGAGGCTGAGGGGAGAAGGGGAGGCGGACAGAAGAGAGAGGGAGGGAGAAAGGGGGAGAA¶
GAGGAAAAAGAGGGAAAGGGACAGACAGGAAGGAAAACAGACCGAGAGAGATCAGTTTTGAGATCCAGGA¶
ACTGCTTTTAGGAAAAGTGAAGGAGGAAAAGGGAAAGAAAAGGAAGACCCCTTCCCAACCAAAATCTTTC¶
CTTTCTTCTCTCTTTTCTGTCTTCTCTTTCTCCATCTCTCAAACTCTCTCTTCTTCCCTCTCTCTTTATT¶
```

B. Remove all paragraph symbols (¶) from the sequence to have a full continuous sequence.

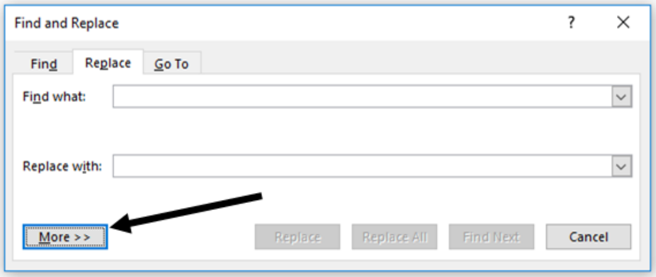
NOTE 7: Continuous, unbroken sequences are necessary for downstream analyses.

1. Highlight the entire sequence, excluding “>Human”.

2. Under the “Home” tab, click on “Replace” at the right top portion of the window (see image below).

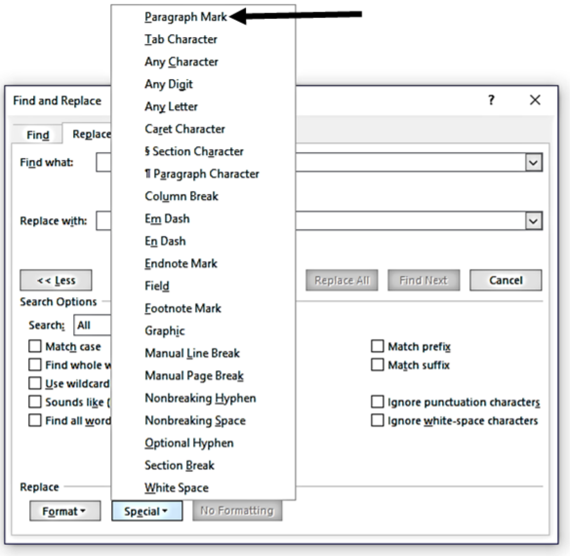


3. A new “Find and Replace” window will pop up. In the “Find and Replace” box, click on the “More” button. The window will expand to allow for more choices (see image below).



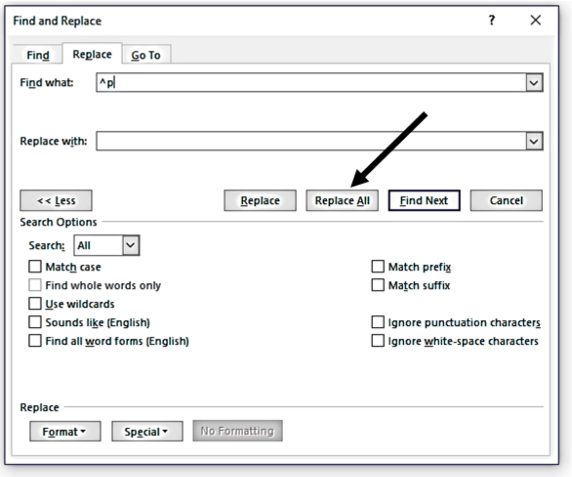
4. Click on the “Special” button. A drop-down menu will appear.

5. Click on “Paragraph” at the top of the list. “^p” will be entered into the “Find what” entry (see image below).



6. Leave the "Replace with" entry blank. This will just remove all paragraph symbols from the sequence.

7. Click the "Replace all" button (see image below).


Find what: ^p

Search: All

8. A new window will appear specifying how many paragraph symbols have been replaced. When asked "Do you want to search the remainder of the document?", click on the "No" button. The sequence will now be devoid of paragraph marks (see image below).

```
>Human¶
AAGACTAGAGCTTCCACCTCTTCGTCACGCTAAGCAGATACAGGAGTCTTCTTTTAAGTAATCAGCTTAGAGACACG
TGTAAAATTACCTGGAAAAATAAGGAACAGCTAGGGGAGAGATTCACCTAAGGATGAAGATCCTTCAGGGATATCAC
TATTTGAACACGTTTCTGAAATCTACTCTTTAGCTGAGGGAGGGGTGGTTTGGAGGCTTCCTCCTTCCCCCAGGGCT
AAGATTTTAAAATGAGATTCTAATGTGATTTCCCGCATCTTACCAGTCAAGACATTCATTTTGGCTTTCAAAGAAAA
AAGTATTGACTTCACCCCCCAGTTTCCTAGATATTAAAAAGAAATCCTTATTAAAATATCTCATAAGTTAAAGTAAT
AGATTTGGGGAAGATTTCAGTGTTGAGTGGTCAAAGAAAAGGGTTTTCTTCCAGATGGTATTTGAATTAAGGCGGCT
GAGGCAGCAGAAAGCTCTCACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTG
CACCCCCCGCCCACGTTCCCTTAAGCCTGTAATATTTTCTGGTAAAAGCAGATGCCAAAAGCTGCCATCCCAGGGTT
CGCAAAGCACCCTTCCTAAAAGGAGCCCTATAACCCTTTTCACTGCCAAAATATATTTTTTAAAAGACTAGTTTTCC
TTATTTACTTGCATTTCCTGCTCTCCTCTTTGCTCGGCCACATTTTATCTTGGAAGGAGCTGGAAGCTGAAATGTGT
¶
```

C. Save the file.

III. Follow steps I.-II.C. for all other sequences. Save all sequences as individual Plain Text files in the "Hoxa2_mVista_Sequences" folder. Be sure to label each file by its correct organismal name ("Chimpanzee, Orangutan, etc.").

NOTE 8: At this point, there will be a total of 7 individual FASTA genomic sequence files in the "Hoxa2_mVista_Sequences" folder.

IV. Generate Annotation files for each *Hoxa2* Genomic DNA sequence.

NOTE 9: The annotation files are made to define the genomic DNA coordinates of the upstream noncoding DNA, exon 1, intron, and exon 2 of *Hoxa2* for each organism analyzed.

A. Develop a new folder on your computer and name it "Hoxa2_mVista_Annotations".

B. Open a new word document.

C. Enter in the coordinates in FASTA format for the beginning of exon 1, end of exon1, beginning of exon 2, and end of exon 2. All coordinates for each organism are shown in the table below.

Organism	**Exon 1 Begin**	**Exon 1 End**	**Exon 2 Begin**	**Exon 2 End**
Human	3001	3391	4036	4775
Chimpanzee	3001	3391	4036	4775
Gorilla	3001	3388	4029	4768
Orangutan	3001	3385	4030	4769
Mouse	3001	3379	4020	4759
Chicken	3001	3382	3983	4728
Zebrafish	3001	3361	3947	4677

NOTE 10: The human coordinates will be used as an example for constructing the annotations files.

1. For line 1, type "> 3001 4775 Hoxa2". Be sure to add a space between ">", "3001", "4775", and "Hoxa2".

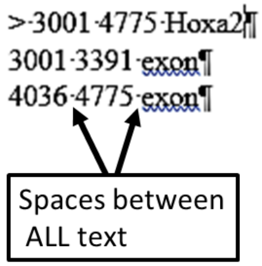
   NOTE 11: The first line denotes that the *Hoxa2* gene is located at bp positions 3001 to 4775 and that there are 3000 bp of noncoding DNA upstream of the *Hoxa2* gene.

2. Fore line 2, type "3001 3391 exon". Be sure to add a space between "3001", "3391", and "exon".

   NOTE 12: The second line denotes that the first exon of *Hoxa2* is located at bp positions 3001-3391.

3. For line 3, type "4036 4775 exon". Be sure to add a space between "4036", "4775", and "exon" (see image below).

   NOTE 13: The third line denotes that the second exon of *Hoxa2* is located at bp positions 4036-4775. All text altogether also denote bp positions 3392-4035 as noncoding DNA, otherwise known as the intron of *Hoxa2* (see image below).

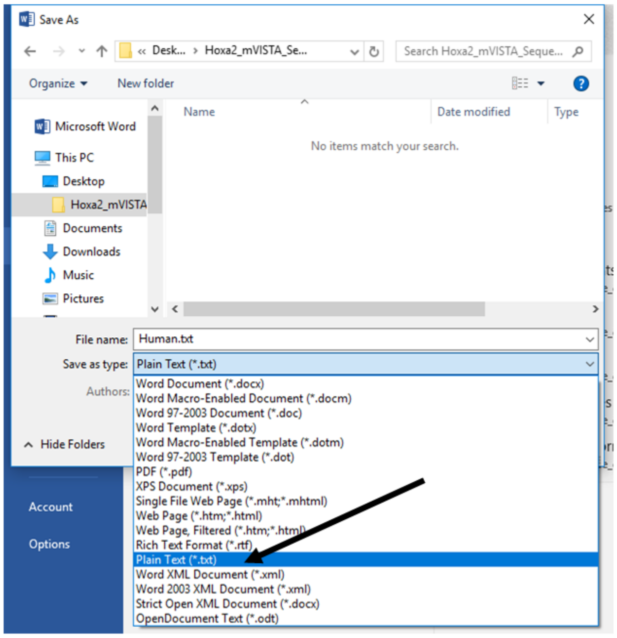


```
>·3001·4775·Hoxa2¶
3001·3391·exon¶
4036·4775·exon¶
```

D. Save the file as a Plain Text file.

1. Under the "File" tab, Click on "Save As".
2. Choose the "Hoxa2_mVista_Annotations" folder to save the file.
3. Name the file, "Human_Hoxa2_mVista_Annotation".
4. Choose the "Plain text" option (see image below).

NOTE 14: The Plain text file format must be used for the mVista program.



File name: Human.txt

Save as type: Plain Text (*.txt)

E. Follow steps IV.B.-IV.D.4. to generate all other sequence annotation files. Save all sequences as individual Plain Text files in the "Hoxa2_mVista_Annotations" folder. Be sure to label each file by its correct organismal name ("Chimpanzee, Orangutan, etc.").

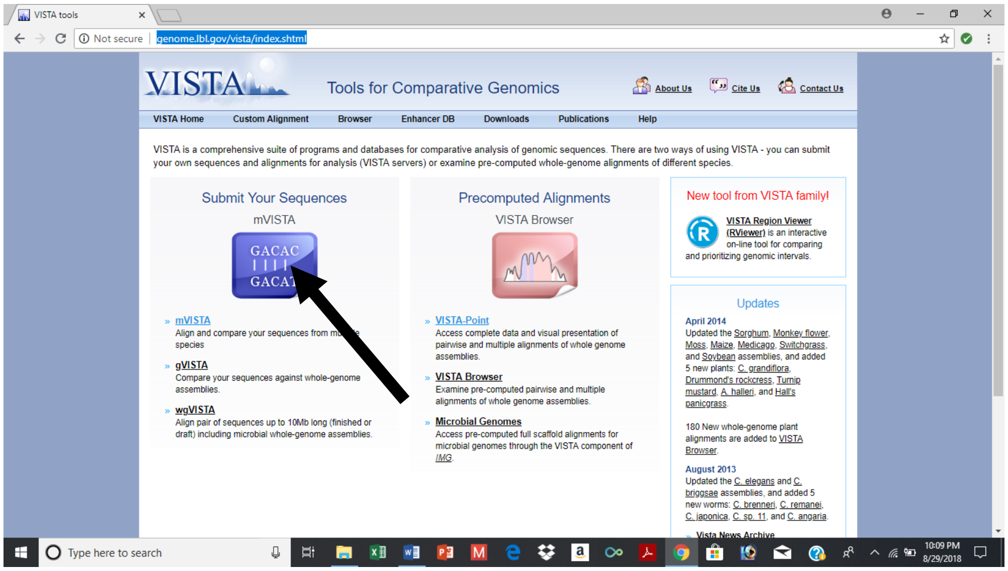
NOTE 15: At this point, there will be a total of 7 individual sequence annotation files in the "Hoxa2_mVista_Annotations" folder.

V. Construct a mVISTA graphical representation of the exonic, intronic, and intergenic genomic DNAs of all of the organisms listed in this SOP.

A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), pull up the VISTA web page.

1. Type in http://genome.lbl.gov/vista/index.shtml

2. Click on the mVISTA button toward the left side of the screen (see image below).



VISTA is a comprehensive suite of programs and databases for comparative analysis of genomic sequences. There are two ways of using VISTA - you can submit your own sequences and alignments for analysis (VISTA servers) or examine pre-computed whole-genome alignments of different species.

Submit Your Sequences

mVISTA

» mVISTA
Align and compare your sequences from multiple species

» gVISTA
Compare your sequences against whole-genome assemblies.

» wgVISTA
Align pair of sequences up to 10Mb long (finished or draft) including microbial whole-genome assemblies.

Precomputed Alignments

VISTA Browser

» VISTA-Point
Access complete data and visual presentation of pairwise and multiple alignments of whole genome assemblies.

» VISTA Browser
Examine pre-computed pairwise and multiple alignments of whole genome assemblies.

» Microbial Genomes
Access pre-computed full scaffold alignments for microbial genomes through the VISTA component of IMG.

New tool from VISTA family!

VISTA Region Viewer (RViewer) is an interactive on-line tool for comparing and prioritizing genomic intervals.

Updates

April 2014
Updated the Sorghum, Monkey flower, Moss, Maize, Medicago, Switchgrass, and Soybean assemblies, and added 5 new plants: C. grandiflora, Drummond's rockcress, Turnip mustard, A. halleri, and Hall's panicgrass.

180 New whole-genome plant alignments are added to VISTA Browser.

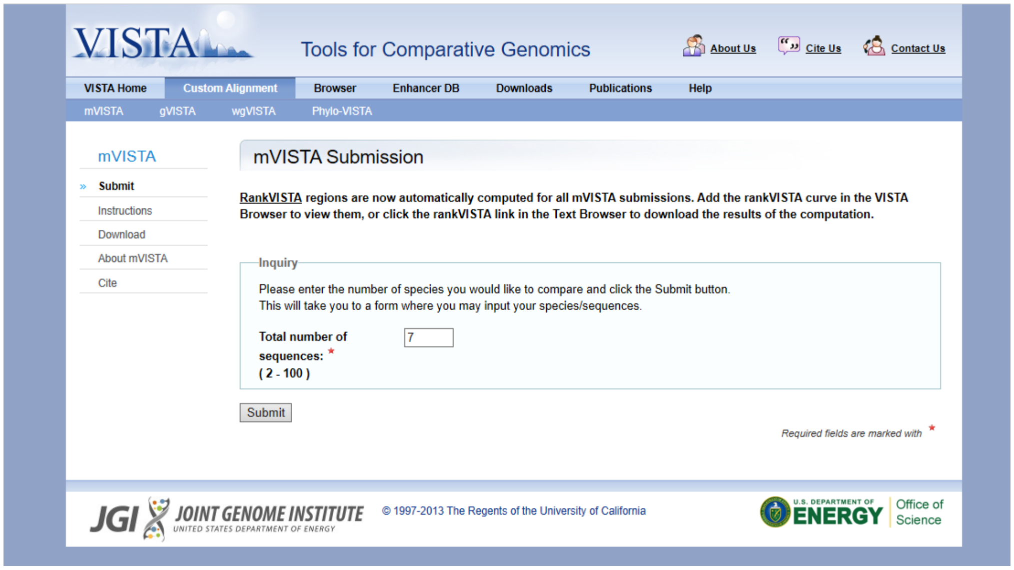
August 2013
Updated the C. elegans and C. briggsae assemblies, and added 5 new worms: C. brenneri, C. remanei, C. japonica, C. sp. 11, and C. angaria.

» Vista News Archive

B. In the mVISTA screen, submit the number of sequences to analyze.

1. Type in the number "7" in the text box to the right of the text, "Total number of sequences". The seven sequences retrieved from NCBI from steps I.-III. of this SOP will be used for this analysis.

2. Click on the "Submit" button (see image below).



mVISTA Submission

RankVISTA regions are now automatically computed for all mVISTA submissions. Add the rankVISTA curve in the VISTA Browser to view them, or click the rankVISTA link in the Text Browser to download the results of the computation.

Inquiry

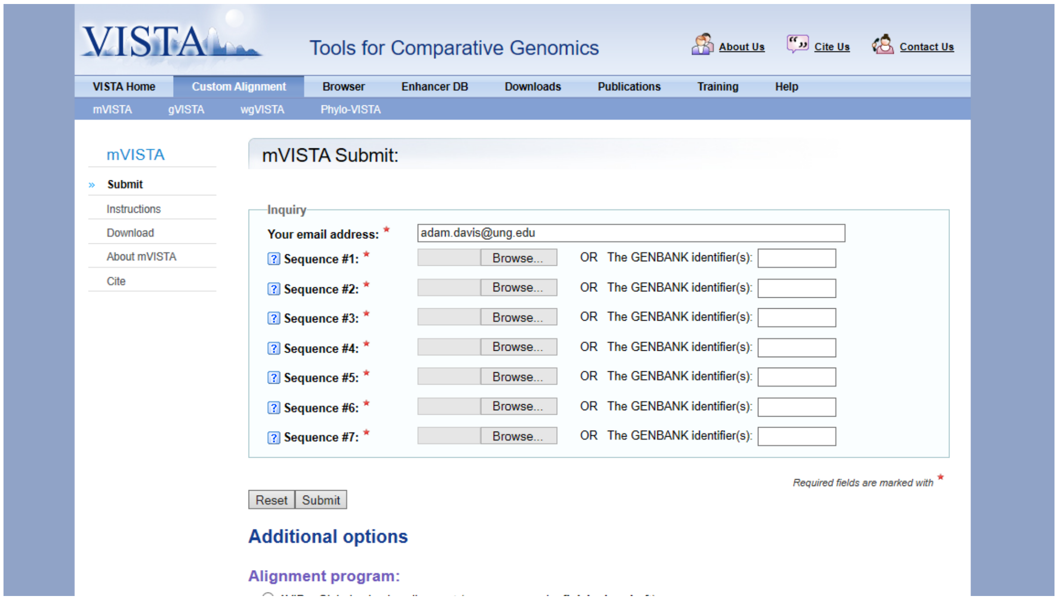
Please enter the number of species you would like to compare and click the Submit button.
This will take you to a form where you may input your species/sequences.

Total number of sequences: * (2 - 100) 7

Required fields are marked with *

3. A new screen will appear asking for specific information for the sequences.

C. Enter your email address in the box to the right of "Your email address:" (see image below).



mVISTA Submit:

Inquiry

Your email address: * adam.davis@ung.edu

Sequence #1: * Browse... OR The GENBANK identifier(s):

Sequence #2: * Browse... OR The GENBANK identifier(s):

Sequence #3: * Browse... OR The GENBANK identifier(s):

Sequence #4: * Browse... OR The GENBANK identifier(s):

Sequence #5: * Browse... OR The GENBANK identifier(s):

Sequence #6: * Browse... OR The GENBANK identifier(s):

Sequence #7: * Browse... OR The GENBANK identifier(s):

Required fields are marked with *

Additional options

Alignment program:

<u>NOTE 16</u>: All results obtained from mVISTA are through email alone.

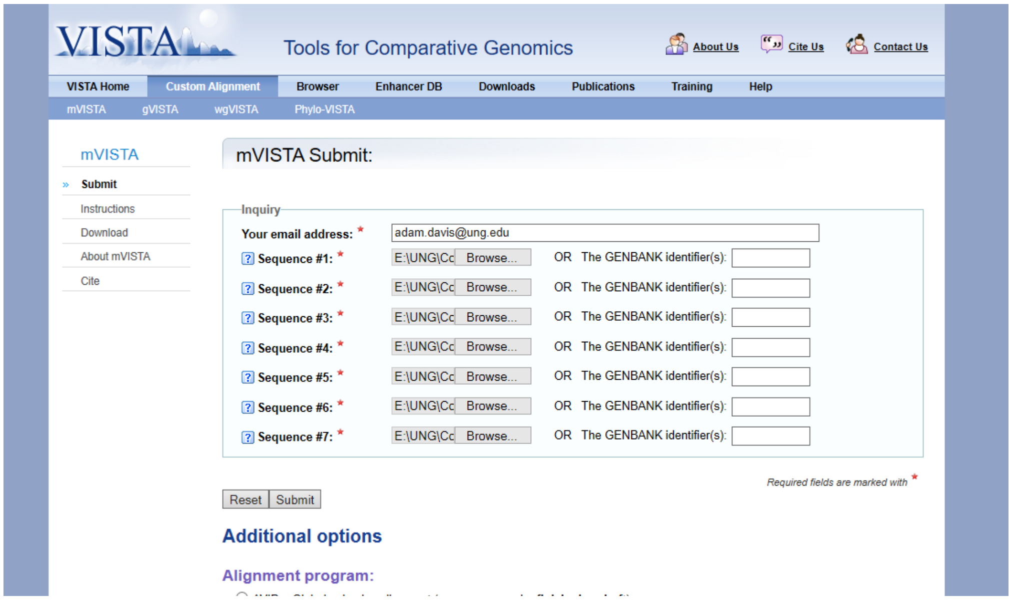
D. Enter the Plain Text files of each *Hoxa2* genomic DNA sequence from steps I.-III. of this SOP.

1. Click on the topmost button that is labeled “Choose File”. This is to the right of the text labeled “Sequence #1:” (see image above).

2. Choose the “Human_Hoxa2_mVista_Sequence” Plain Text file from the “Hoxa2_mVista_Sequences” folder.

3. Continue adding in the rest of the Hoxa2 sequence files in the order using the table below:

Sequence #1	Human_Hoxa2_mVista_Sequence
Sequence #2	Chimpanzee_Hoxa2_mVista_Sequence
Sequence #3	Gorilla_Hoxa2_mVista_Sequence
Sequence #4	Orantutan_Hoxa2_mVista_Sequence
Sequence #5	Mouse_Hoxa2_mVista_Sequence
Sequence #6	Chicken_Hoxa2_mVista_Sequence
Sequence #7	Zebrafish_Hoxa2_mVista_Sequence



Your email address: * adam.davis@ung.edu

Sequence #1: * E:\UNG\Cc OR The GENBANK identifier(s):

Sequence #2: * E:\UNG\Cc OR The GENBANK identifier(s):

Sequence #3: * E:\UNG\Cc OR The GENBANK identifier(s):

Sequence #4: * E:\UNG\Cc OR The GENBANK identifier(s):

Sequence #5: * E:\UNG\Cc OR The GENBANK identifier(s):

Sequence #6: * E:\UNG\Cc OR The GENBANK identifier(s):

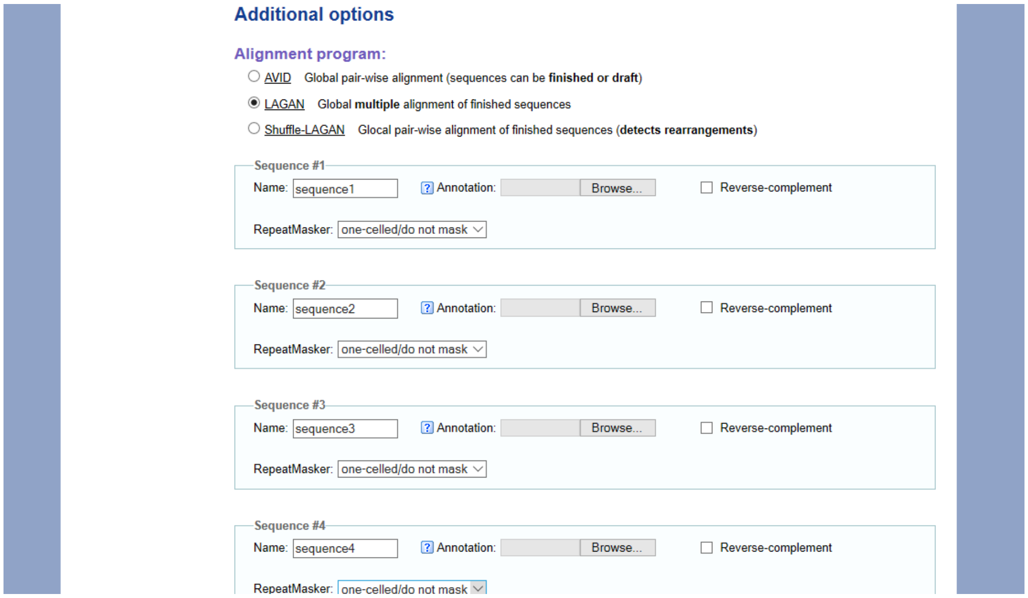
Sequence #7: * E:\UNG\Cc OR The GENBANK identifier(s):

Required fields are marked with *

E. Enter the Plain Text files of each *Hoxa2* genomic DNA annotation generated from steps IV.A.-IV.E. of this SOP.

1. Scroll down the mVista screen to the section titled “Additional Options”.

2. If it is not already selected, choose the “LAGAN” option (see image below).



○ AVID Global pair-wise alignment (sequences can be **finished or draft**)

● LAGAN Global **multiple** alignment of finished sequences

○ Shuffle-LAGAN Glocal pair-wise alignment of finished sequences (**detects rearrangements**)

Sequence #1
Name: sequence1 Annotation: ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

Sequence #2
Name: sequence2 Annotation: ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

Sequence #3
Name: sequence3 Annotation: ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

Sequence #4
Name: sequence4 Annotation: ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

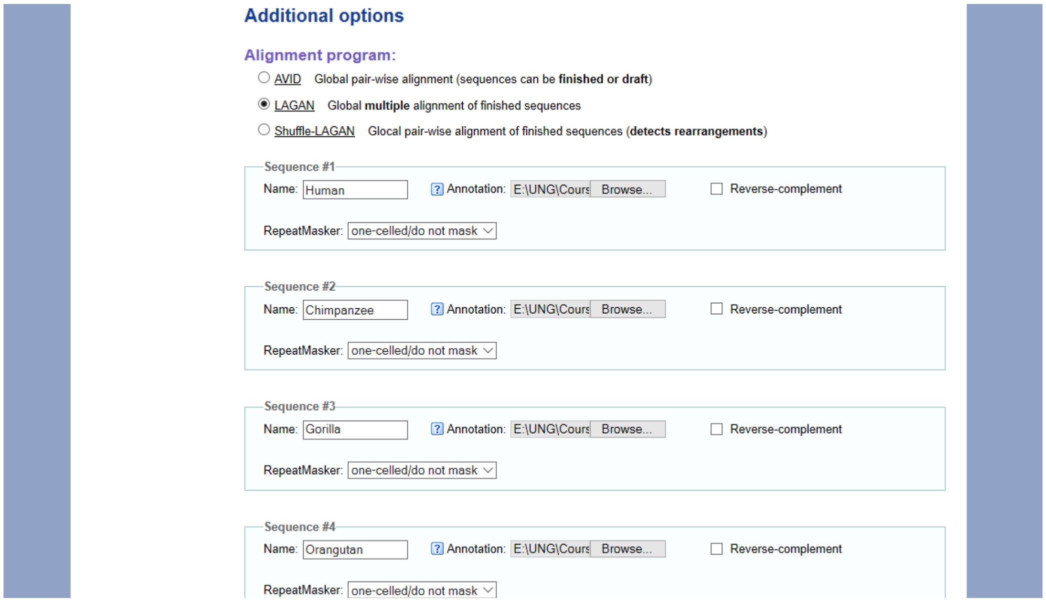
<u>NOTE 17</u>: LAGAN is a program for global pairwise and multiple sequence alignment of finished sequences.

3. Replace the “Sequence1” text in the topmost text box with “Human”.

4. Click on the button that is labeled "Choose File". This is to the right of the text box that is now labeled "Human".

5. Choose the "Human_Hoxa2_mVista_Annotation" Plain Text file from the "Hoxa2_mVista_Annotations" folder.

6. Follow steps V.E.1.-V.E.5. for the remaining six sequence annotation files. **Add the sequence names and files in the same order as those for the sequence files (see the Table on page 13)**. Be sure to label the sequences with the correct organism name (see image below).



Additional options

Alignment program:

○ AVID Global pair-wise alignment (sequences can be **finished or draft**)

● LAGAN Global **multiple** alignment of finished sequences

○ Shuffle-LAGAN Glocal pair-wise alignment of finished sequences (**detects rearrangements**)

Sequence #1
Name: Human ? Annotation: E:\UNG\Cours Browse... ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

Sequence #2
Name: Chimpanzee ? Annotation: E:\UNG\Cours Browse... ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

Sequence #3
Name: Gorilla ? Annotation: E:\UNG\Cours Browse... ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

Sequence #4
Name: Orangutan ? Annotation: E:\UNG\Cours Browse... ☐ Reverse-complement
RepeatMasker: one-celled/do not mask

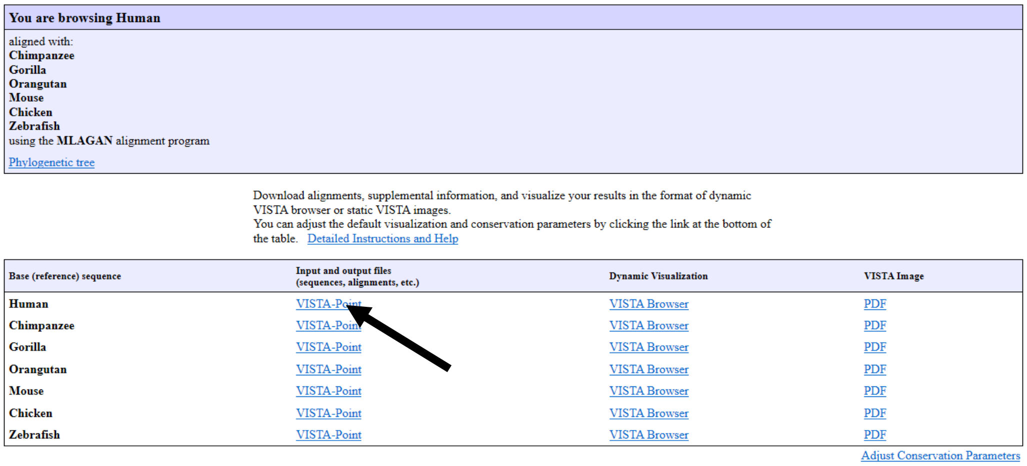
F. Scroll to the bottom of the mVista screen and click the "Submit" button.

<u>NOTE 18</u>: Depending on the internet traffic, it may take several days to receive the mVista results by email.

VI. Convert the mVista results to a figure in Powerpoint.

A. In your email from Vista, click on the link that provides the mVista results. A screen that reads "You are browsing Human" at the top will open.

B. Under "Input and output files (sequences, alignments, etc.)", click on the top-most "VISTA-Point" selection to the right of "Human" (see image below).



You are browsing Human

aligned with:
Chimpanzee
Gorilla
Orangutan
Mouse
Chicken
Zebrafish
using the **MLAGAN** alignment program

Phylogenetic tree

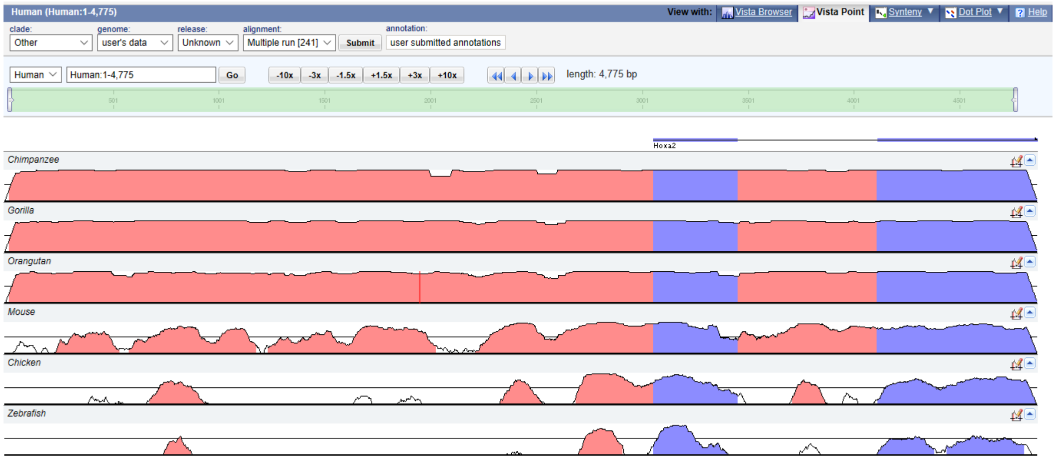
Download alignments, supplemental information, and visualize your results in the format of dynamic VISTA browser or static VISTA images.
You can adjust the default visualization and conservation parameters by clicking the link at the bottom of the table. Detailed Instructions and Help

Base (reference) sequence	Input and output files (sequences, alignments, etc.)	Dynamic Visualization	VISTA Image
Human	VISTA-Point	VISTA Browser	PDF
Chimpanzee	VISTA-Point	VISTA Browser	PDF
Gorilla	VISTA-Point	VISTA Browser	PDF
Orangutan	VISTA-Point	VISTA Browser	PDF
Mouse	VISTA-Point	VISTA Browser	PDF
Chicken	VISTA-Point	VISTA Browser	PDF
Zebrafish	VISTA-Point	VISTA Browser	PDF

Adjust Conservation Parameters

NOTE 19: Choosing this option will make the Human sequence as the reference sequence and all other sequences will be compared to it.

NOTE 20: A new screen will emerge showing the graphical output of the DNA sequences. Red corresponds to noncoding DNA (upstream regions and introns). Blue corresponds to coding DNA (exons). The Human sequence is not shown because it is providing a reference for all other sequences to be compared. Taller colored regions correspond to regions of higher conservation between orthologous sequences.



Human (Human:1-4,775)

clade: Other
genome: user's data
release: Unknown
alignment: Multiple run [241]
annotation: user submitted annotations

Human Human:1-4,775

length: 4,775 bp

C. If necessary, shrink the screen to be able to see all sequences on the same screen.

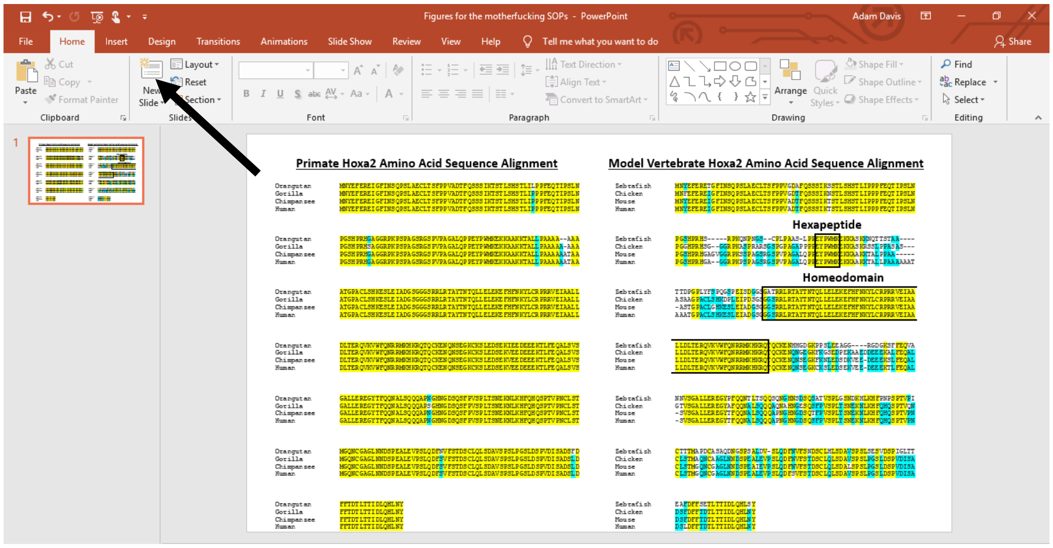
1. Press and hold down the “Ctrl” key on the keyboard of the computer.

2. While this is held down, press the “-“ key on the keyboard once. Each time this key is typed while the “Ctrl” key is pressed and held, the screen will shrink. Now all sequences are displayed on the screen.

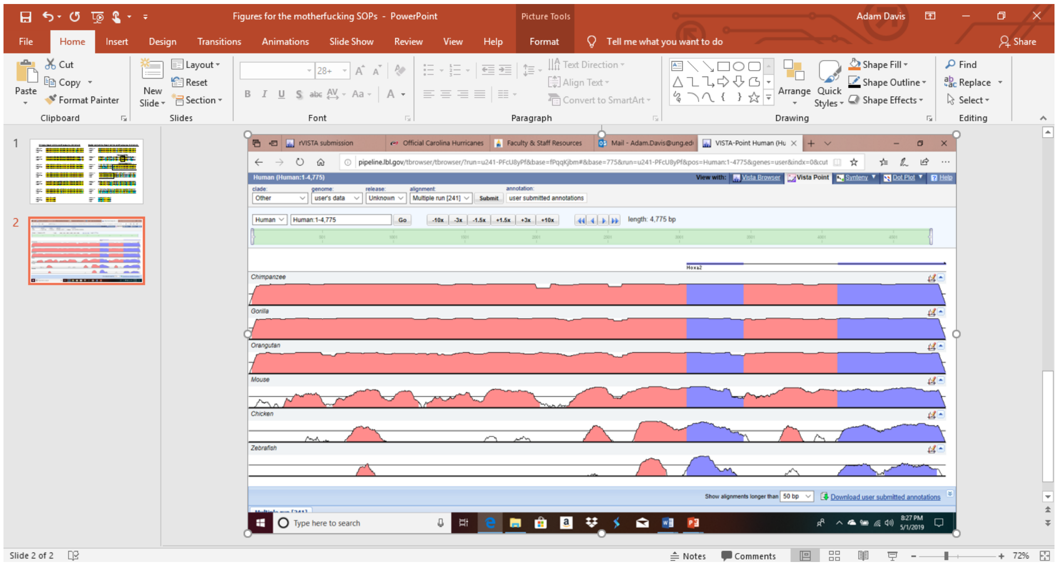
D. Use the “Print Screen” option on the keyboard to copy the entire screen containing the mVista results.

E. Open the “Hoxa2_Figures_Powerpoint” file made from BIO-001.

F. Under the “File” tab, click on the “New Slide” button to place a new slide after the slide containing the amino acid sequence alignment image generated from BIO-003 (see image below).

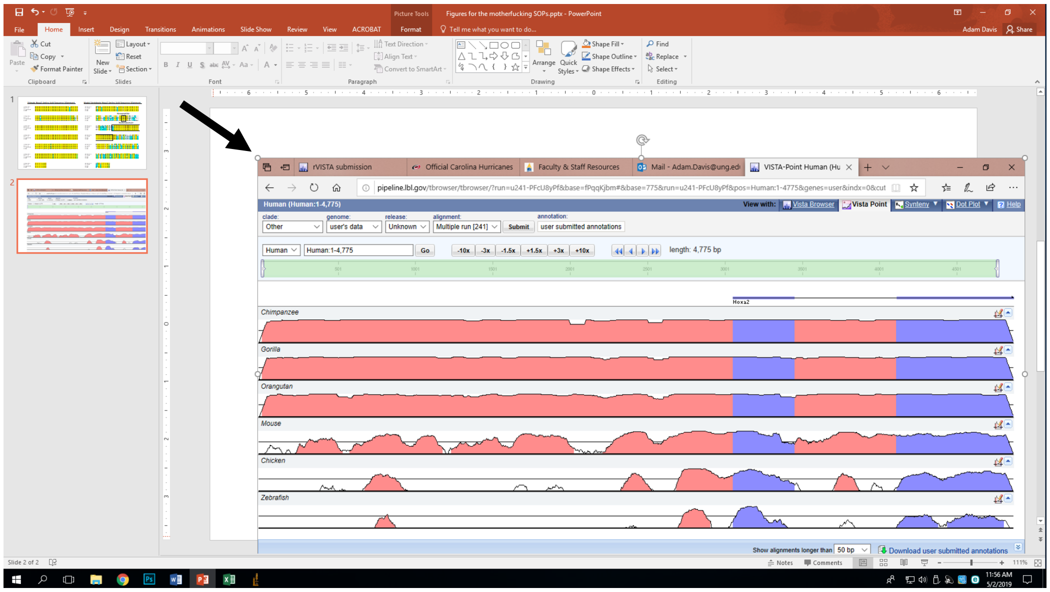


G. Paste the image of the mVista results retrieved from the Print Screen option from step VI.D. (see image below).

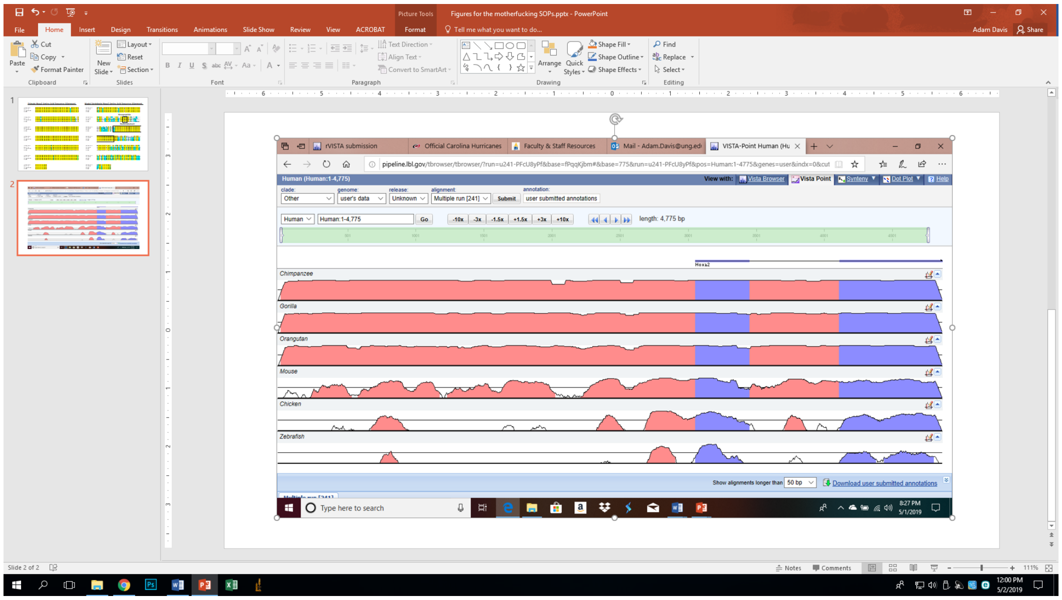


H. If necessary, resize the image of the mVisa results so that the entire image can be viewed in the Powerpoint slide.

1. Click on the image and drag it so that one of its corners is visible (see image below).



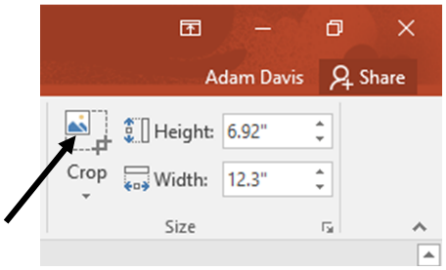
2. Click on the corner of the image and drag it inward to reduce the size of the image (see image below).



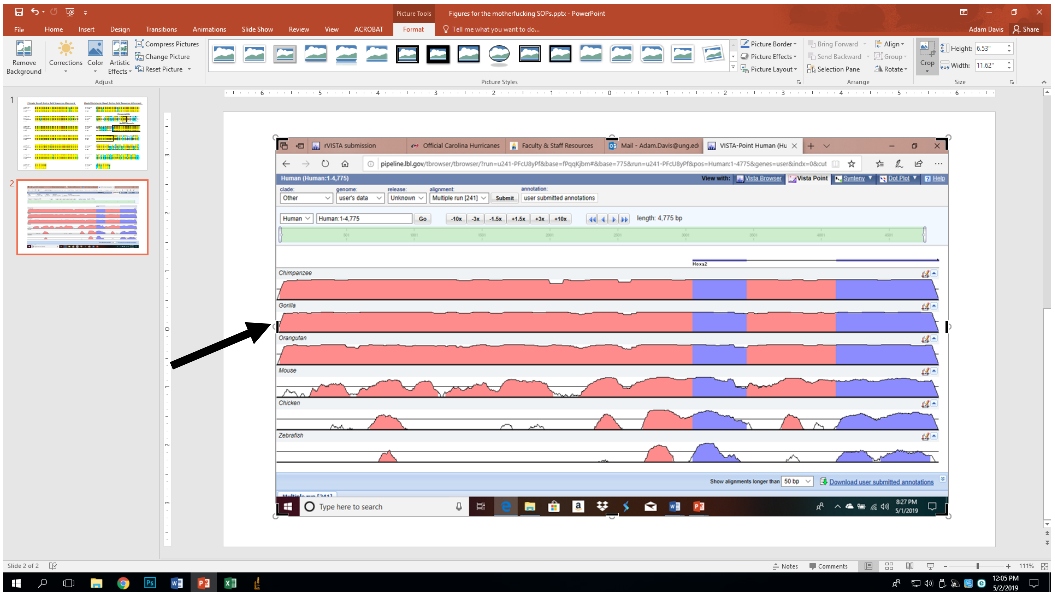
I. Remove all extraneous information by cropping the image.

1. Click anywhere on the mVista figure. A Format tab will appear in the menu.

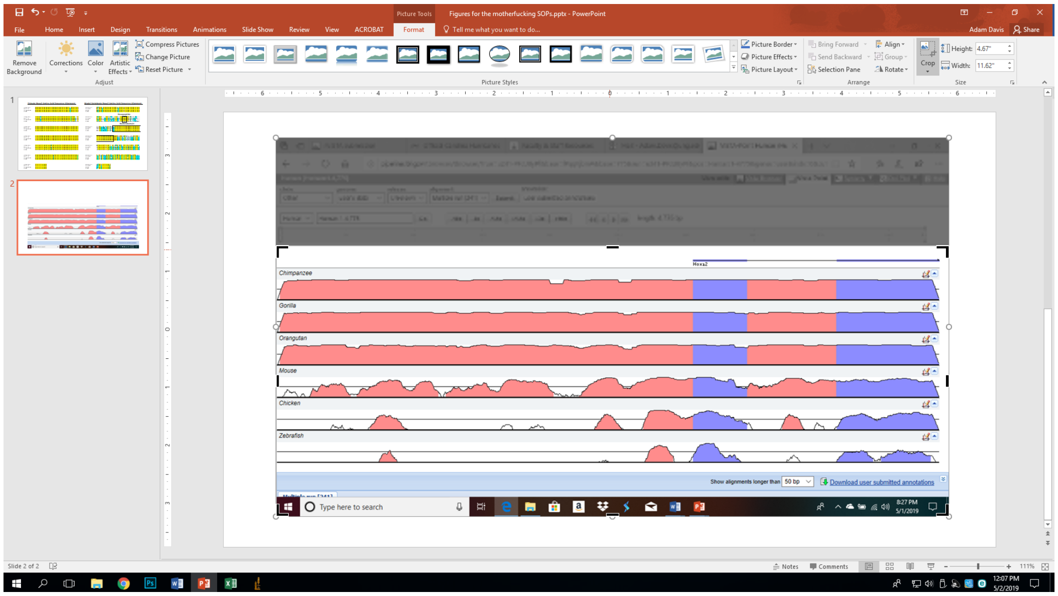
2. Under the Format tab, click on the “Crop” button at the top right of the powerpoint screen (see image below).



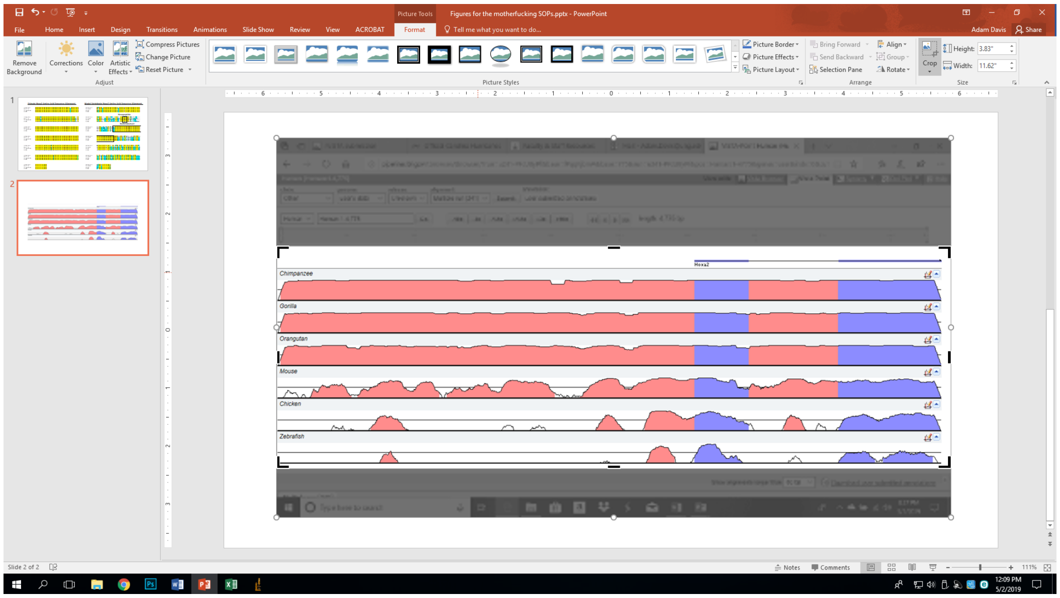
   a. A border will appear at the edges of the figure with several thicker lines that serve as anchors for cropping the image (see image below).



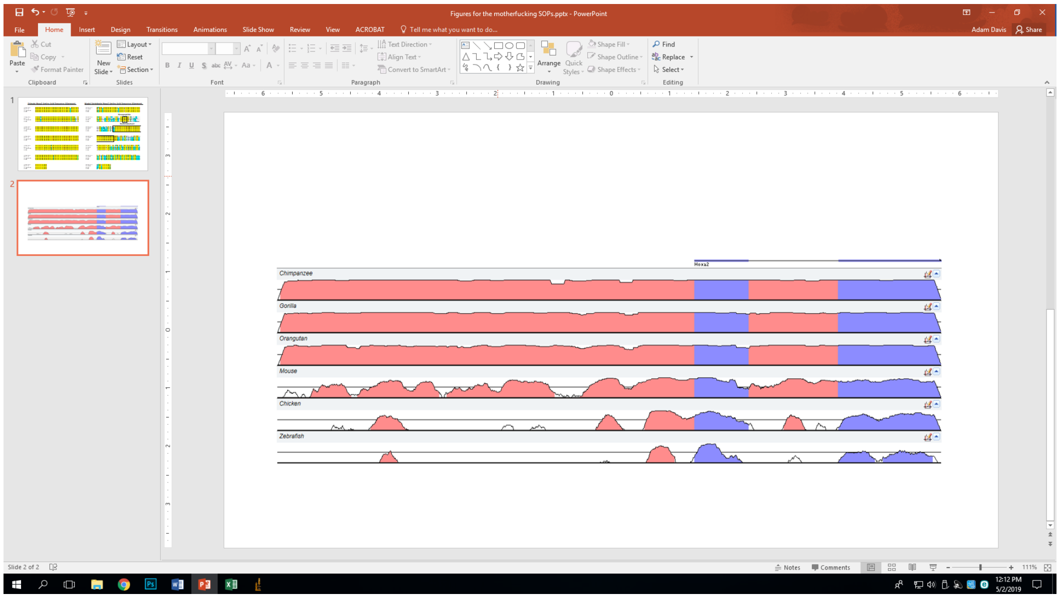
3. Click on and drag the top-middle anchor down so as to remove all extraneous information above the sequences including the green ruler (see image below).



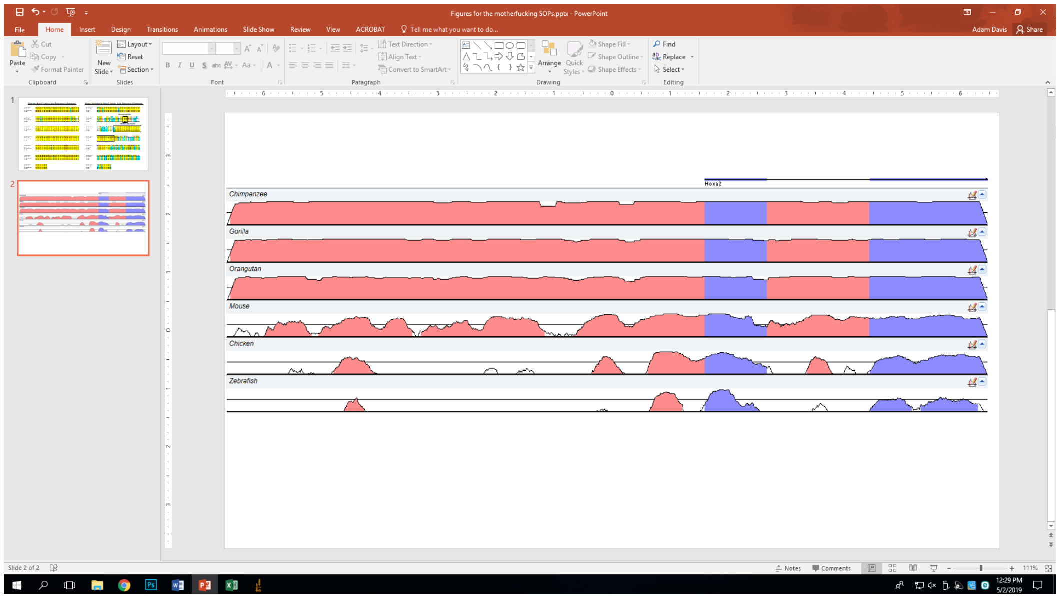
4. Click on and drag the bottom-middle anchor up so as to remove all extraneous information below the sequences up to the black line below the zebrafish sequence (see image below).



5. Click anywhere on the slide outside of the cropped image to exit the cropping procedure. There should no longer be any shaded portions of the figure that represent regions to be cropped (see image below).



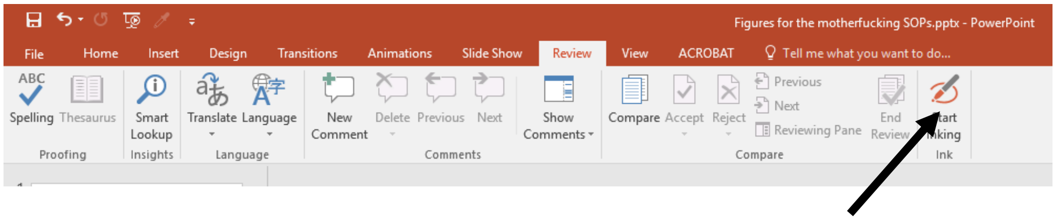
J. Enlarge the image by clicking on the corners and dragging outward so that the image fully spans the left-right axis of the PowerPoint slide (see image below).



NOTE 21: The figure will gently lock into place when you drag it to the left and right ends of the slide.

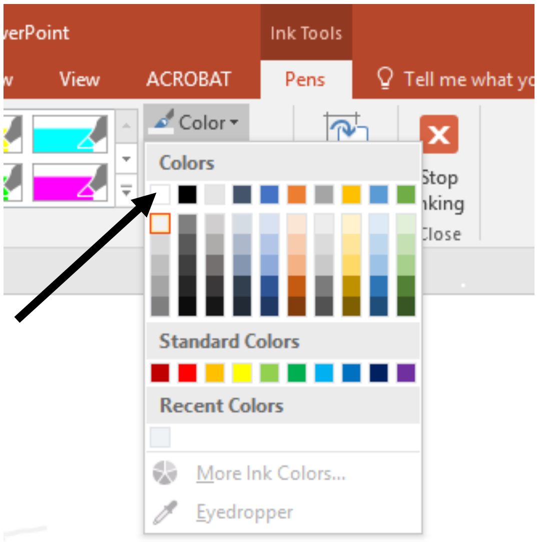
K. Cover the text for each sequence, as well as the “Hoxa2” text, and replace it with larger bolder text so it can be easily visualized.

1. Under the “Review” tab, click on the “Start Inking” button (see image below).

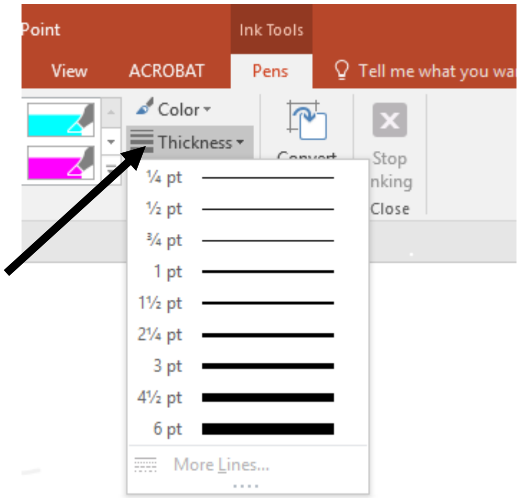


2. A new set of commands will appear in the menu. The “Pen” button should already be highlighted. If it is not highlighted, click on the “Pen” button (see image below).

3. Click on the “Color” button to access the assortment of colors to use. Click on the first and lightest gray color in the left-most column of choices. This most closely resembles the gray background on which the species names are written (see image below).

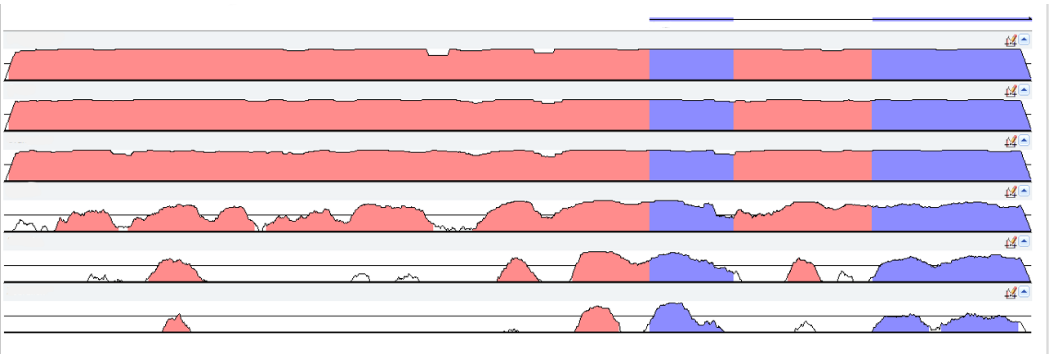


4. If necessary, change the thickness of the pen point by clicking on the “Thickness” button and choosing a size that you are most comfortable with (see image below).



5. Using the pen, color over the species names to erase them. Also use the white color to remove the “Hoxa2” text underneath the line representing Exon 1 of Hoxa2 (see image below).

   NOTE 22: You may need to increase the Zoom on the bottom right of the Microsoft PowerPoint screen to over 200% to be able to color the “Hoxa2” text.



6. Label the sequences with their specific organism names as well as *Hoxa2* and its associated exons.

   a. Follow steps VIII.D.3.a.-VIII.D.3.b. from the BIO-001 SOP to create a text box.

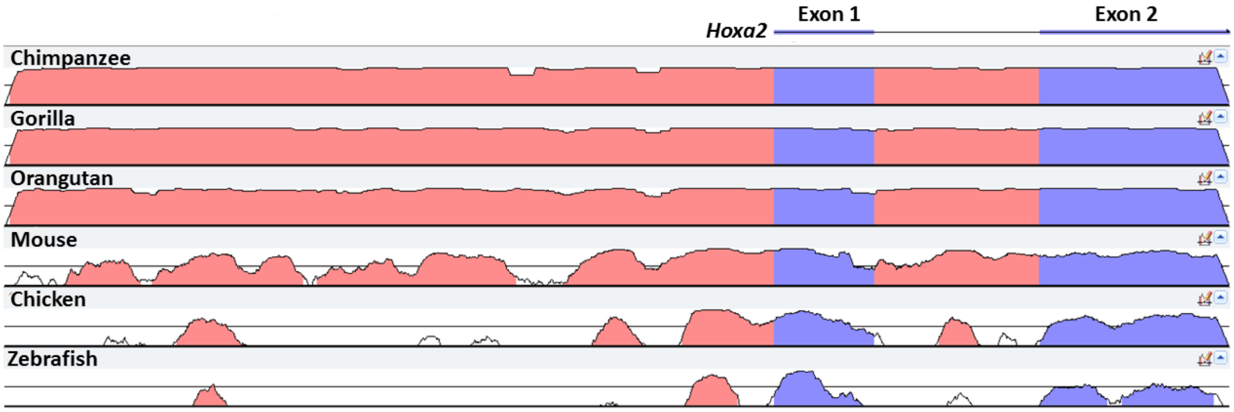
   b. Keep the Font type and size as is or change it to your liking.

   c. Bold all Text. Italicize the “Hoxa2” text.

   d. Move the bolded organism name text box over the region of the image that contained the original organism name text.

e. Move the bolded and italicized "*Hoxa2*" text just before the line corresponding to Exon 1.

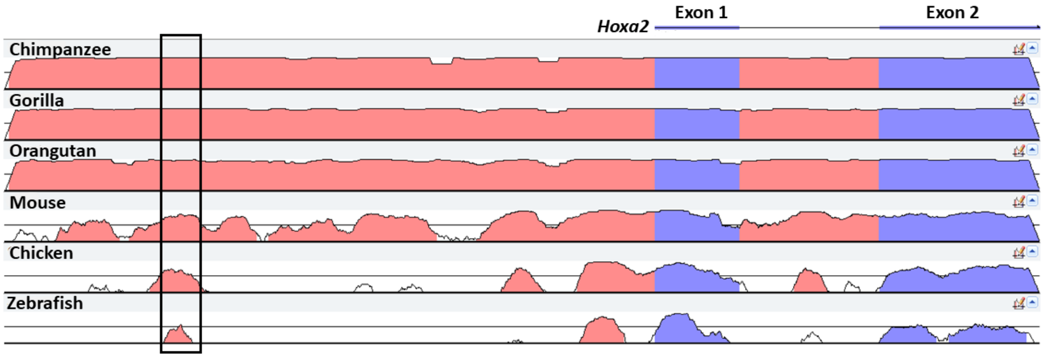
f. Write "Exon 1" and "Exon 2" above each of the lines that correspond to these sequences (see image below).



L. Label the enhancer regions that drive *Hoxa2* gene expression in rhombomreres (r) 3 and 5 and the pharyngeal arches (PAs) and r4 as well as the proximal promoter (PP).

NOTE 23: The r3/r5/PA enhancer region lies ~1000-2000 bp upstream of the ATG start site of *Hoxa2.* The r4 enhancer region lies within the intron of *Hoxa2*. The PP lies just upstream of exon 1 of *Hoxa2.* The only way these sequences can be visualized from this figure is by looking at the regions that are conserved among all species. These region are best visualized by observing the zebrafish sequence.

1. Follow steps II.B.1.b.-II.B.1.f.iv. of the BIO-002 SOP to make a transparent window. Have the window be just widen enough to encase the zebrafish r3/r5/PA enhancer but tall enough to cover the conserved regions of the orthologous sequences from all other organisms (see image below).



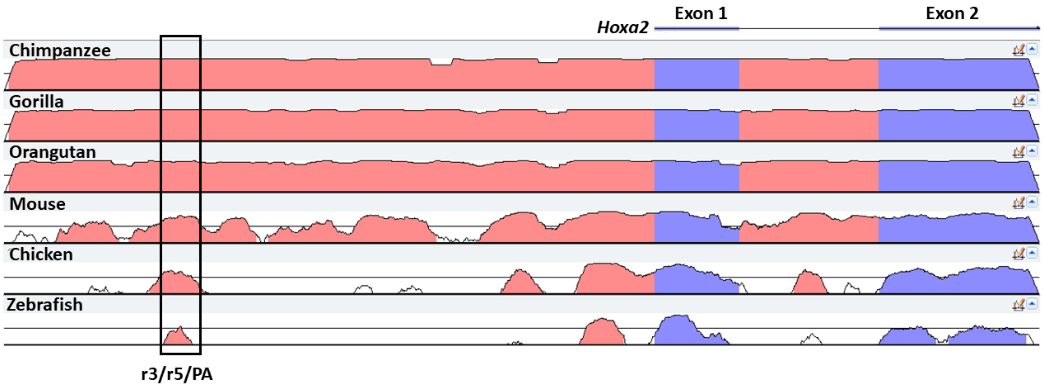
2. Label the r3/r5/PA region.

a. Follow steps VIII.D.3.a.-VIII.D.3.b. of the BIO-001 SOP to create a text box.

b. Type “r3/r5/PA”.

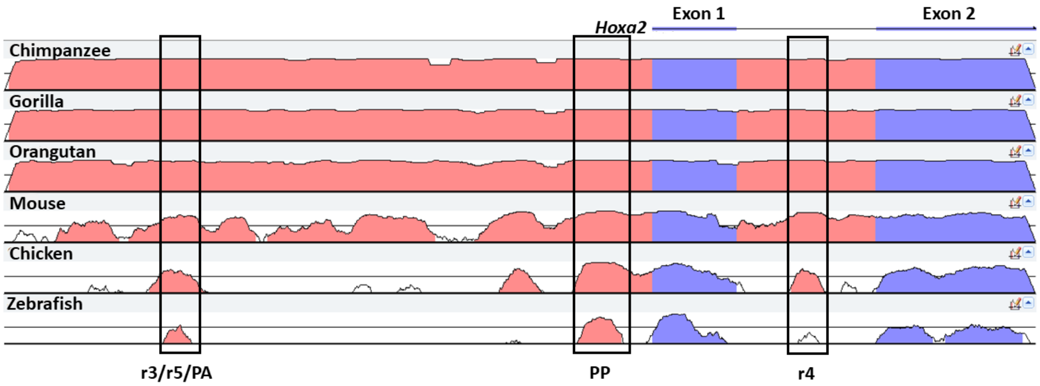
c. Bold the text.

d. Move the bolded “r3/r5/PA” text just below the middle of the rectangle encompassing the r3/r5/PA region (see image below).



3. Label the PP and r4 enhancer regions.

a. Follow steps VI.L.1.-VI.L.2.c. but place the transparent rectangles over the conserved region for the PP and r4 enhancer for all species. Label the proximal promoter as “PP” and the r4 enhancer region as “r4” (see image below).



VII. Save the “Hoxa2_Figures_Powerpoint” file.

VIII. Convert the sequence alignment data into a TIFF image file.

A. Click the “File” tab.

B. Under the "File" tab, click on "Save As".

C. Once a destination folder is chosen, click on the dropdown menu to the right of the text, "Save as type:".

D. Choose "TIFF Tag Image File Format".

E. Label the file "Hoxa2_Figure_2".

F. A new window will pop up asking, "Which slides do you want to export?". Click on the button labeled, "Just This One".

IX. Add the new figure to the "Hoxa2_Figures_Word" document generated from BIO-002.

A. Open the "Hoxa2_Figures_Word" document.

B. Add a new page to the document to which you will input the figure made from this SOP.

C. Click on the "Hoxa2_Figure_2" TIFF file and drag it into the second page of the Word document.

D. Underneath the figure, write a legend that describes:

1. The layout of the figure (what type of DNA was used, what species are represented, etc.)
2. The methods used to develop this figure (i.e.: what software program was used, which organism is used as the reference sequence, etc.)
3. The red and blue peaks and the significance on the size of these peaks.

E. Save the "Hoxa2_Figures_Word" file.

X. Turn in the "Hoxa2_Figures_Word" file according to the deadline set by the instructor.

References:

Brudno, M., C.B. Do, G.M. Cooper, M.F. Kim, E. Davydov, E.D. Green, A. Sidow, and S. Batzoglou, 2003. LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Res 13(4):721-731.

Davis, A., J.L. Scemama, and E.J. Stellwag. 2008. Japanese medaka *Hox* paralog group 2: insights into the evolution of *Hox* PG2 gene composition and expression in the Osteichthyes. J Exp Zool (Mol Dev Evol) 310(8):623-641.

Davis, A., M.C. Reubens, and E.J. Stellwag. 2016. Function and comparative genomics of *Hoxa2* gene *cis*-regulatory elements: evidence for evolutionary modification of ancestral core element activity. J Dev Biol 4(2):15.

Frazer, K.A., L. Pachter, A. Poliakov, E.M. Rubin, and I. Dubchak. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32(Web Server issue):W273-9

Gendron-Maguire, M., M. Mallo, M. Zhang, T. Gridley. 1993. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75(7):1317-1331.

Maconochie, M.K., R. Krishnamurthy, S. Nonchev., P. Meier, M. Manzanares, P.J. Mitchell, and R. Krumlauf. 1999. Regulation of *Hoxa2* in cranial neural crest cells involves members of the *AP-2* family. Development 126(7):1483-1494.

Maconochie, M.K., S. Nonchev, M. Manzanares, H. Marshall, and R. Krumlauf. 2001. Differences in Krox20-dependent regulation of Hoxa2 and Hoxb2 during hindbrain development. Dev Biol 233(2):468-481.

Mayor, C.., M. Brudno, J.R. Schwartz, A. Poliakov, E.M. Rubin, K.A. Frazer, L.S. Pachter, and I. Dubchak. 2000. VISTA: Visualizing Global DNA Sequence Alignments of Arbitrary Length. Bioinformatics 16(11):1046-1047.

McEllin, J.A., T.B. Alexander, S. Tümpel, L.M. Wiedemann, and R. Krumlauf. 2016. Analyses of fugu *hoxa2* genes provide evidence for subfunctionalization of neural crest cell and rhombomere *cis*-regulatory modules during vertebrate evolution. Dev Biol 409(2):530-542.

Nonchev, S., C. Vesque, M. Maconochie, T. Seitanidou, L. Ariza-McNaughton, M. Frain, H. Marshall, M.H. Sham, R. Krumlauf, and P. Charnay. 1996. Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. Development 122(2):543-554.

Rijli, F.M., M. Mark, S. Lakkaraju, A. Dierich, P. Dolle, and P. Chambon. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75(7):1333-1349.

Tümpel, S., F. Cambronero, L.M. Wiedemann, and R. Krumlauf. 2006. Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes). Proc Natl Acad Sci 103(14):5419-5424.

Tümpel, S., F. Cambronero, E. Ferretti, F. Blasi, L.M. Wiedemann, and R. Krumlauf, 2007. Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. Dev Biol 302(2):646-660.

Tümpel, S., L.M. Wiedemann, and R. Krumlauf. 2009. Hox genes and segmentation of the vertebrate hindbrain. Curr Top Dev Biol 88:103-137.

SOP #	BIO-003
Title	Analysis of *Hoxa2* Genomic DNA Sequences I: Use of mVISTA, Microsoft Word, and Microsoft PowerPoint Software to Analyze Conserved Coding and Intergenic DNA.
Site	Adam Davis, Ph.D. Department of Biology University of North Georgia

SOP Assessment

1. True or False: Intergenic DNA sequences are part of the coding domains of a gene (they are translated into amino acids).

2. The orange/red regions of the graphical view of the genomic DNA correspond to:

    a. Coding DNA
    b. Amino acids
    c. Noncoding DNA
    d. Exons

3. Exons 1 and 2 of the figure give rise to the:

    a. Promoter and *cis*-regulatory elements of *Hoxa2*
    b. Noncoding DNA of *Hoxa2*
    c. Amino acid sequence of *Hoxa2*
    d. Intron of *Hoxa2*

Date: ____

Name (Print): ____

Signature: ____

SOP #	BIO-004
Title	Analysis of *Hoxa2* Genomic DNA Sequences II: Use of Clustal, Microsoft Word, and Microsoft PowerPoint Software to Analyze *Cis*-Regulatory Elements Responsible for Directing *Hoxa2* Gene Expression in Rhombomeres 3 and 5 and the Pharyngeal Arches.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

Objective:

To understand the function and evolution of *Hoxa2*, especially with directing the development of the rhombomere-derived cranial nerves and the pharyngeal arch-derived skeletal elements, the upstream genomic DNA sequences, which contain regulatory elements that direct the expression of *Hoxa2*, must be retrieved from the appropriate database for analysis. Furthermore, genomic DNA sequences from evolutionarily divergent species must be analyzed to fully understand how this protein functions. Genomic DNA sequences will be retrieved from several primate species, including Human (*Homo sapiens*), Chimpanzee (*Pan troglodytes*), Gorilla (*Gorilla gorilla*), and Orangutan (*Pongo abelii*) and several vertebrate biomedical models, including Zebrafish (*Danio rerio*), Chicken (*Gallus gallus*), and Mouse (*Mus musculus*). The software program, Clustal, will be employed for this analysis.

Relevant Terms and their Definitions

Hoxa2 – Developmental regulatory gene that is evolutionarily conserved and functions to pattern the development of the facial nerve.

National Center for Biotechnology Information (NCBI) – Web-based database containing genetic information submitted by scientists. Used for genetic, developmental, medical, ecological and evolutionary research analyses.

Genbank Accession number – Identifying number for genomic DNA sequences

FASTA – Sequence format that must be obtained for nucleotide sequences for downstream analyses.

Intergenic regions – Regions of noncoding genomic DNA that span between genes and that can harbor regulatory elements.

Clustal – Multiple sequence alignment program for DNA or proteins.

Indel mutations – Point mutation of a nucleotide that results in an insertion of an amino acid or deletion of a nucleotide.

Cis-regulatory elements (CREs) – short sequences of genomic DNA that bind transcription factor (TF) proteins and regulate the spatial and temporal expression patterns of their respective genes.

r3/r5/PA enhancer region – region of genomic DNA upstream of *Hoxa2* that contains CREs that direct *Hoxa2* expression in rhombomeres 3 and 5 (r3 and r5) and the pharyngeal arches (PAs).

Krox20 - Transcription factors that function to bind to *cis*-regulatory elements and direct *Hoxa2* gene expression in r3 and r5.

Sox - Transcription factors works in conjunction with Krox20 to bind to *cis*-regulatory elements and direct *Hoxa2* gene expression in r3 and r5.

Hox/Pbx - Transcription factors that function to bind to *cis*-regulatory elements and direct *Hoxa2* gene expression in the PAs.

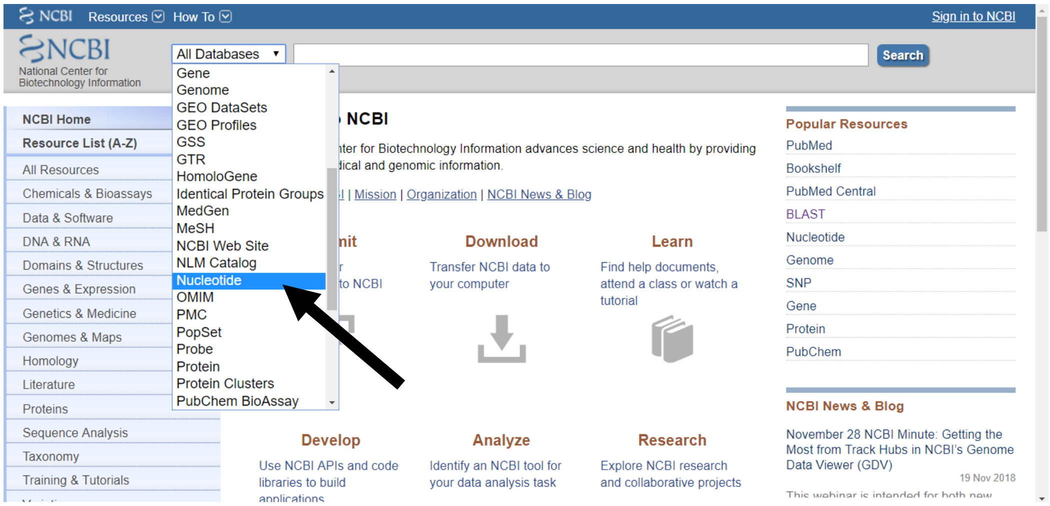
Prep/Meis - Transcription factors that function with Hox/Pbx factors to bind to *cis*-regulatory elements and direct *Hoxa2* gene expression in the PAs.

Procedure

I. Retrieve the CREs responsible for directing *Hoxa2* gene expression in r3, r5, and the PAs from the Genbank database.

A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), type in https://www.ncbi.nlm.nih.gov

B. Click on the drop down menu titled "All Databases" and select "Nucleotide" (see image below).



C. Type in the appropriate Genbank accession number in the text box next to the Drop-down menu to retrieve the appropriate coding DNA sequences and click on the "Search" button. The Genbank accession numbers for the species-specific *Hoxa2* noncoding DNA sequences are listed in the table below.

Organism	***Hoxa2* Genbank Accession**	**Sequence Begin Coordinate**	**Sequence End Coordinate**	**Reverse Complement Necessary**
Human	AC004079	87067	87415	Yes
Chimpanzee	NC_036886	27221268	27221616	Yes
Gorilla	NC_018431	27159477	27159825	Yes
Orangutan	NC_036910	46400521	46400872	No
Mouse	CH466597	4376330	4376657	Yes
Chicken	NC_006089	32586583	32586983	Yes
Zebrafish	AL645795	59409	59802	Yes

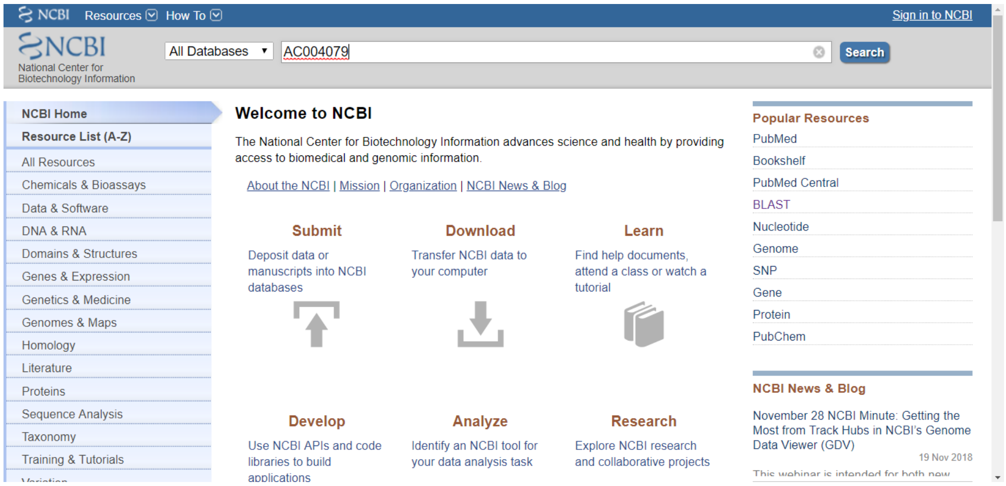
D. Once the Genbank sequence information is displayed, retrieve the designated region of the *Hoxa2* upstream intergenic sequence in FASTA format and in the 5'-3' orientation.

NOTE 1: FASTA format is necessary for future downstream analyses, including DNA sequence alignment.

NOTE 2: The Human *Hoxa2* sequence (Accession #: AC004079) will be used as an example for the next several steps of this SOP. This genomic sequence is from the entire chromosome 7, which is 102,717 bp in length. Due to the extreme length, it is computationally too intensive to show the entire sequence. For this reason, sequence start and end positions are listed in the table on page 2 of this SOP.

1. Type "AC004079" in the text box to the right of the dropdown menu.

2. Click on the "Search" button to the right of the textbox (see image below).



AC004079

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

Submit: Deposit data or manuscripts into NCBI databases

Download: Transfer NCBI data to your computer

Learn: Find help documents, attend a class or watch a tutorial

Develop: Use NCBI APIs and code libraries to build applications

Analyze: Identify an NCBI tool for your data analysis task

Research: Explore NCBI research and collaborative projects

November 28 NCBI Minute: Getting the Most from Track Hubs in NCBI's Genome Data Viewer (GDV)

19 Nov 2018

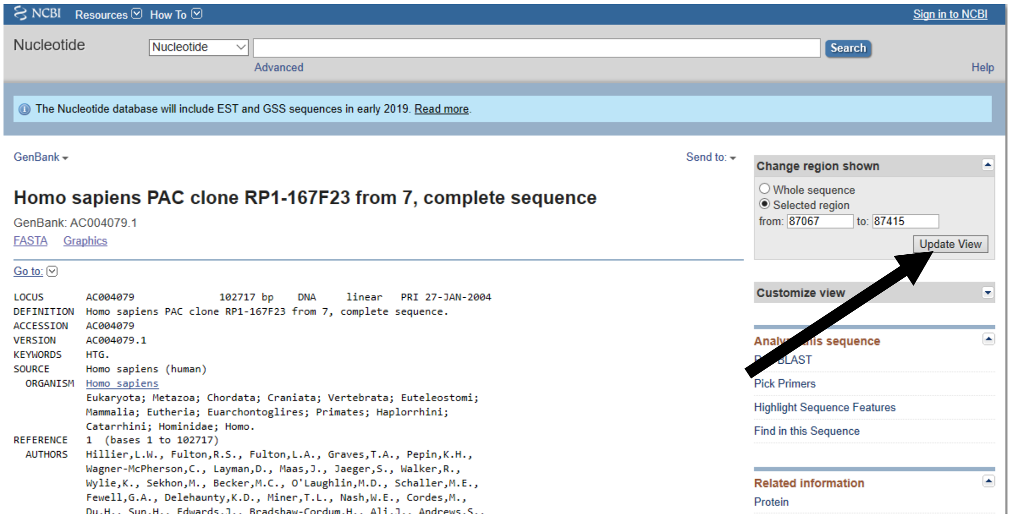
3. The Human *Hoxa2* Genomic DNA sequence Genbank information will be displayed. Click on the Down arrow next to the words "Change Region Shown" on the right side of the screen.

4. Click on "Selected region".

5. Type in the sequence begin position (for Human – 87067) from the table above in the box labeled "begin".

6. Type in the sequence end position (for Human – 87415) from the table above in the box labeled "end".

7. Click on the "Update View" button (see image below).



The Nucleotide database will include EST and GSS sequences in early 2019. Read more.

Homo sapiens PAC clone RP1-167F23 from 7, complete sequence

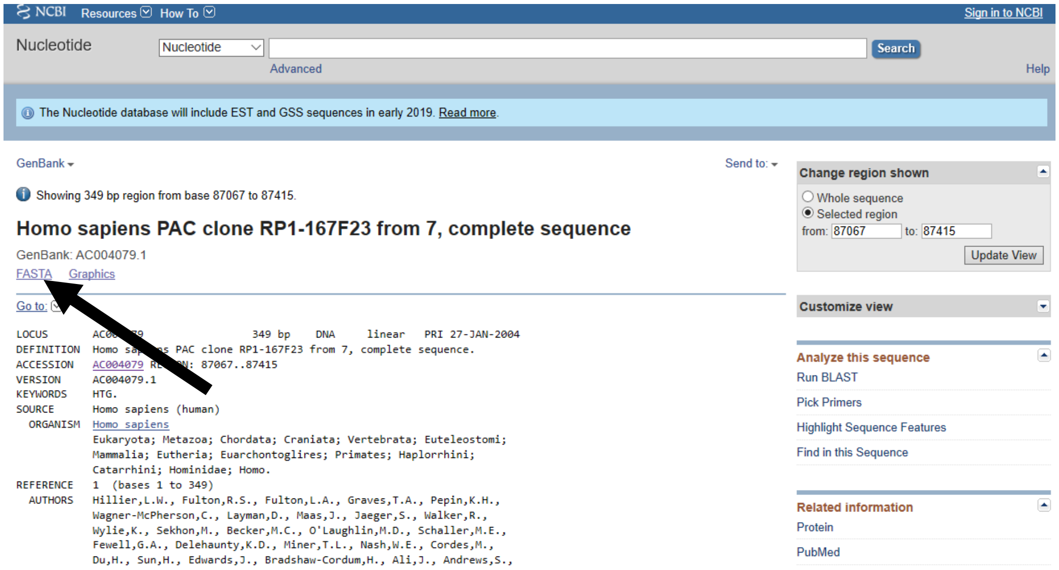
GenBank: AC004079.1

○ Whole sequence
● Selected region
from: 87067 to: 87415

```
LOCUS       AC004079              102717 bp    DNA     linear   PRI 27-JAN-2004
DEFINITION  Homo sapiens PAC clone RP1-167F23 from 7, complete sequence.
ACCESSION   AC004079
VERSION     AC004079.1
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102717)
  AUTHORS   Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
            Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
            Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
            Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
            Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
```

NOTE 3: A new screen will be displayed showing just the length of sequence specified (for Human it is 349 bp). For this SOP, only a portion of the specified genomic DNA corresponding to *Hoxa2* will be displayed.

E. Click on "FASTA" on the upper left corner of the screen. The FASTA sequence format removes all identifying information from the sequence file (see image below).



The Nucleotide database will include EST and GSS sequences in early 2019. Read more.

Showing 349 bp region from base 87067 to 87415.

Homo sapiens PAC clone RP1-167F23 from 7, complete sequence

GenBank: AC004079.1

from: 87067 to: 87415

```
LOCUS       AC0    79               349 bp    DNA     linear   PRI 27-JAN-2004
DEFINITION  Homo sa    s PAC clone RP1-167F23 from 7, complete sequence.
ACCESSION   AC004079 RE    N: 87067..87415
VERSION     AC004079.1
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
            Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
            Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
            Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
            Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
```

F. If necessary, change the orientation of the sequence so that it is displayed in the 5'-3' orientation.

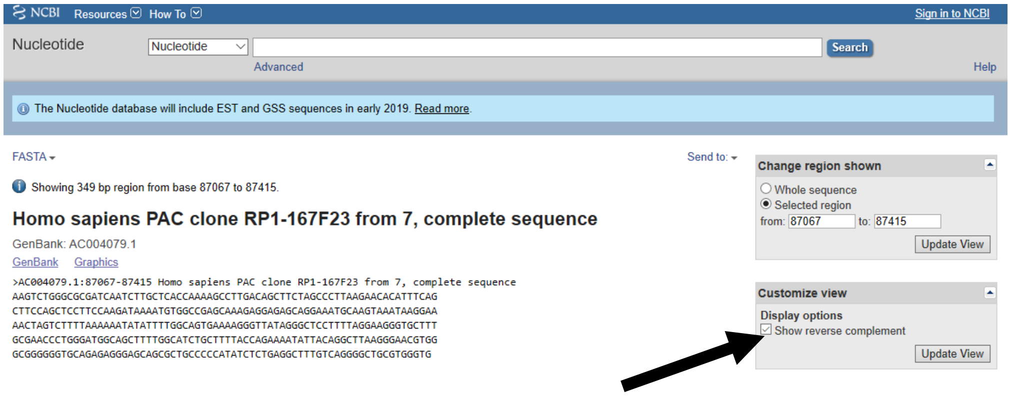
NOTE 4: Since DNA is double-stranded and both strands are complementary to each other, separate genes can be located on either strand. Depending on how the genomic DNA sequence (normally an entire chromosomal DNA sequence) was loaded into Genbank, the sequence of interest could be in either orientation. In order to perform sequence alignments, all sequences retrieved for this analysis must be in the 5'-3' orientation. **Reverse complements of the genomic DNA corresponding to *Hoxa2* must be obtained for ALL organisms except for Orangutan and Frog (see Table above).**

NOTE 5: The human sequence should be in the orientation: 5'-CACCC…GACTT-3'

1. Click on the Down arrow next to the words "Customize view" on the right side of the screen.

2. Under "Display options", click on "Show reverse complement".

3. Click on the "Update View" button (see images below).



Nucleotide: Nucleotide

The Nucleotide database will include EST and GSS sequences in early 2019. Read more.

Showing 349 bp region from base 87067 to 87415.

Homo sapiens PAC clone RP1-167F23 from 7, complete sequence

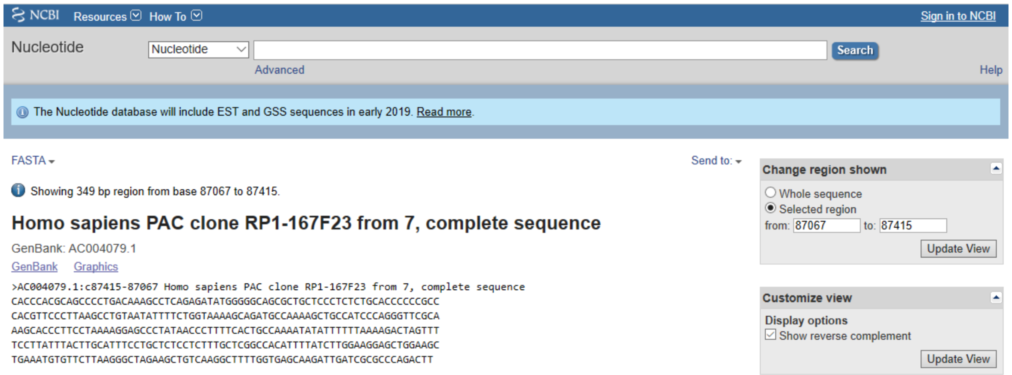
GenBank: AC004079.1

GenBank Graphics

```
>AC004079.1:87067-87415 Homo sapiens PAC clone RP1-167F23 from 7, complete sequence
AAGTCTGGGCGCGATCAATCTTGCTCACCAAAAGCCTTGACAGCTTCTAGCCCTTAAGAACACATTTCAG
CTTCCAGCTCCTTCCAAGATAAAATGTGGCCGAGCAAAGAGGAGAGCAGGAAATGCAAGTAAATAAGGAA
AACTAGTCTTTTAAAAAATATATTTTGGCAGTGAAAAGGGTTATAGGGCTCCTTTTAGGAAGGGTGCTTT
GCGAACCCTGGGATGGCAGCTTTTGGCATCTGCTTTTACCAGAAAATATTACAGGCTTAAGGGAACGTGG
GCGGGGGGTGCAGAGAGGGAGCAGCGCTGCCCCCATATCTCTGAGGCTTTGTCAGGGGCTGCGTGGGTG
```

Change region shown
○ Whole sequence
● Selected region
from: 87067 to: 87415

Customize view
Display options
☑ Show reverse complement



Nucleotide: Nucleotide

The Nucleotide database will include EST and GSS sequences in early 2019. Read more.

Showing 349 bp region from base 87067 to 87415.

Homo sapiens PAC clone RP1-167F23 from 7, complete sequence

GenBank: AC004079.1

GenBank Graphics

```
>AC004079.1:c87415-87067 Homo sapiens PAC clone RP1-167F23 from 7, complete sequence
CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGCACCCCCCGCC
CACGTTCCCTTAAGCCTGTAATATTTTCTGGTAAAAGCAGATGCCAAAAGCTGCCATCCCAGGGTTCGCA
AAGCACCCTTCCTAAAAGGAGCCCTATAACCCTTTTCACTGCCAAAATATATTTTTTAAAAGACTAGTTT
TCCTTATTTACTTGCATTTCCTGCTCTCCTCTTTGCTCGGCCACATTTTATCTTGGAAGGAGCTGGAAGC
TGAAATGTGTTCTTAAGGGCTAGAAGCTGTCAAGGCTTTTGGTGAGCAAGATTGATCGCGCCCAGACTT
```

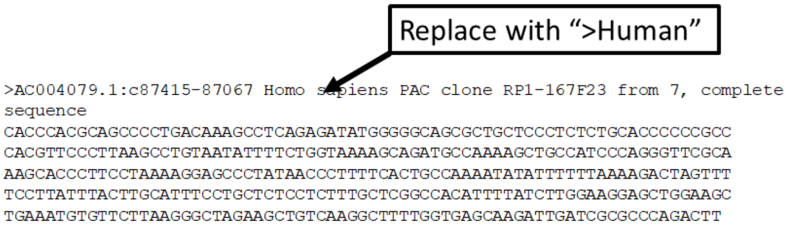
Change region shown
○ Whole sequence
● Selected region
from: 87067 to: 87415

Customize view
Display options
☑ Show reverse complement

G. Highlight the entire sequence, copy it, and paste it into a new word document.

II. Format the sequence in Microsoft Word so that it can be read and processed by the Clustal software program.

A. Replace all information after the carrot symbol (>) in the sequence identification line with the word “Human” (see image below).



Replace with “>Human”

```
>AC004079.1:c87415-87067 Homo sapiens PAC clone RP1-167F23 from 7, complete
sequence
CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGCACCCCCCGCC
CACGTTCCCTTAAGCCTGTAATATTTTCTGGTAAAAGCAGATGCCAAAAGCTGCCATCCCAGGGTTCGCA
AAGCACCCTTCCTAAAAGGAGCCCTATAACCCTTTTCACTGCCAAAATATATTTTTTAAAAGACTAGTTT
TCCTTATTTACTTGCATTTCCTGCTCTCCTCTTTGCTCGGCCACATTTTATCTTGGAAGGAGCTGGAAGC
TGAAATGTGTTCTTAAGGGCTAGAAGCTGTCAAGGCTTTTGGTGAGCAAGATTGATCGCGCCCAGACTT
```

B. Repeat steps I.C. – II.A. for adding all other *Hoxa2* upstream intergenic DNA sequences from all other species listed in the table on page 2 of this SOP.

C. Once all sequences have been added to the Word file, save this file as ”Hoxa2_r3r5PA_Unaligned”.

NOTE 6: Several FASTA sequences are shown below as an example.

```
>Human
CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGCACCCCCCGCC
CACGTTCCCTTAAGCCTGTAATATTTTCTGGTAAAAGCAGATGCCAAAAGCTGCCATCCCAGGGTTCGCA
AAGCACCCTTCCTAAAAGGAGCCCTATAACCCTTTTCACTGCCAAAATATATTTTTTAAAAGACTAGTTT
TCCTTATTTACTTGCATTTCCTGCTCTCCTCTTTGCTCGGCCACATTTTATCTTGGAAGGAGCTGGAAGC
TGAAATGTGTTCTTAAGGGCTAGAAGCTGTCAAGGCTTTTGGTGAGCAAGATTGATCGCGCCCAGACTT
>Chimpanzee
CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGCACCCCCCGCC
CACGTTCCCTTAAGCCTGTAATATTTTCTGGTAAAAGCAGATGCCAAAAGCTGCCATCCCAGGGTTCGCA
AAGCACCCTTCCTAAAAGGAGCCCTATAACCCTTTTCACTGCCAAAATATATTTTTTAAAAGACTAGTTT
TCCTTATTTACTTGCATTTCCTGCTCTCCTCTTTGCTCGGCCACATTTTATCTTGGAAGGAGCTGGAAGC
TGAAATGTGTTCTTAAGGGCTAGAAGCTGTCAAGGCTTTTGGTGAGCAAGATTGATCGCGCCCAGACTT
```

III. Perform two separate genomic DNA sequence alignments using the Clustal software program. One will compare the sequences of the four primates (Human, Chimpanzee, Gorilla, and Orangutan) and the other will compare the Human sequence to the three model vertebrate sequences (Mouse, Chicken, and Zebrafish).

A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), pull up the Clustal Omega website.

1. Type in https://www.ebi.ac.uk/Tools/msa/clustalo/

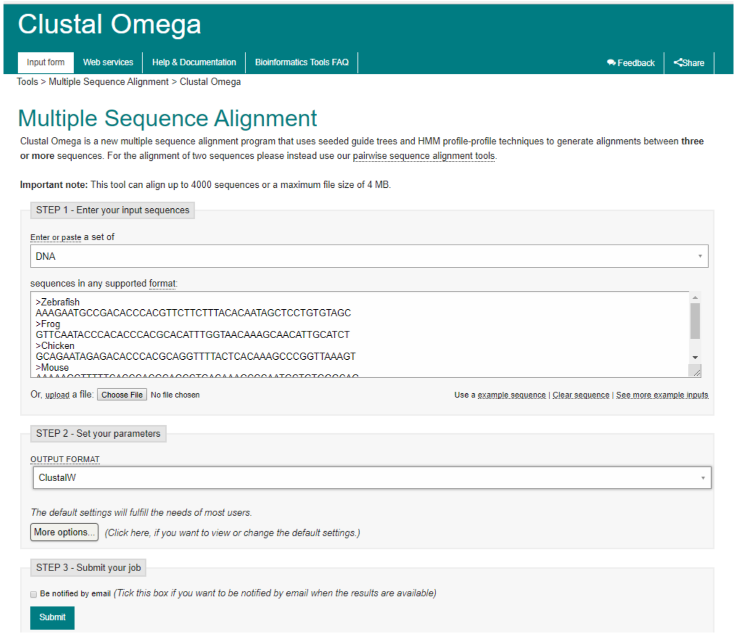
2. Click on the drop-down menu to select the appropriate sequence data to align.

3. Select "DNA" from the drop-down menu.

4. Copy the four model vertebrate sequences (Human, Chimpanzee, Gorilla, and Orangutan) in FASTA format and paste them into the open data box.

5. Under "Step 2 – Set your parameters", select "ClustalW".

6. Click on the "Submit" button on the bottom of the screen (see image below).



Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of: DNA

sequences in any supported format:

```
>Zebrafish
AAAGAATGCCGACACCCACGTTCTTCTTTACACAATAGCTCCTGTGTAGC
>Frog
GTTCAATACCCACACCCACGCACATTTGGTAACAAAGCAACATTGCATCT
>Chicken
GCAGAATAGAGACACCCACGCAGGTTTTACTCACAAAGCCCGGTTAAAGT
>Mouse
```

Or, upload a file: Choose File No file chosen

Use a example sequence | Clear sequence | See more example inputs

STEP 2 - Set your parameters

OUTPUT FORMAT: ClustalW

The default settings will fulfill the needs of most users.

More options... *(Click here, if you want to view or change the default settings.)*

STEP 3 - Submit your job

☐ Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Submit

7. Once the results of the sequence alignment are complete, copy all aligned sequences and paste them into a new Word file and save the file as “Hoxa2_r3r5PA_Alignment_Primates” (see sample data below).

```
Orangutan       CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGC
Human           CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGC
Chimpanzee      CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGC
Gorilla         CACCCACGCAGCCCCTGACAAAGCCTCAGAGATATGGGGGCAGCGCTGCTCCCTCTCTGC
                ************************************************************
```

8. Repeat steps III.A.1.-III.A.7. for the four model vertebrate sequences (Human, Mouse, Chicken, and Zebrafish). Save the Word file as “Hoxa2_r3r5PA_Alignment_ModelVerts” (see sample data below).

```
Zebrafish      CACCCACGTTCTTCTTTACACAATAGCTCCTGTGTAGCTGACAAAAACACGCTTTTC-CT
Chicken        CACCCACGCAGGTTTTACTCACAAAGCCCGGTTAA-AGTCCAAACCTCTCTCTTTATTAT
Mouse          CACCCACGC--AGCC---TGACAAAGC---------------------CCAATGCTGTGG
Human          CACCCACGCAGCCCC---TGACAAAGCCTCAGAGA-TATGGGGGCAGCGCTGCTCCCTCT
               ********              * ***                      *
```

IV. Follow steps VI.A.-VI.E. of the BIO-001 SOP for formatting the DNA sequence alignment documents for nucleotide color coding and eventual figure development.

V. Follow steps VII.A.-VII.D. of the BIO-001 SOP for color code the nucleotide data to highlight conserved DNA sequence regions.

VI. Develop a figure of the genomic DNA sequence alignments.

A. Follow steps VI.E.-VI.F. of the BIO-003 SOP for making a new slide in the "Hoxa2_Figures_Powerpoint" file.

B. Follow steps VIII.-IX. of the BIO-001 SOP for importing the primate and model vertebrate genomic sequence alignments into the new slide of the "Hoxa2_Figures_Powerpoint" file.

   1. Label the Primate alignment data as "Primate *Hoxa2* Upstream Genomic DNA Sequence Alignment".

   2. Label the Model Vertebrate alignment data as "Model Vertebrate *Hoxa2* Upstream Genomic DNA Sequence Alignment".

C. Label the Krox20, Sox, Prep/Meis, and Hox/Pbx *cis*-regulatory elements in the Model Vertebrates alignment.

   NOTE 7: *Krox20* and *Sox* are transcription factors that work in conjunction to direct *Hoxa2* gene expression in r3 and r5. *Hox/Pbx* and *Prep/Meis* are transcription factors that work in conjunction to direct *Hoxa2* gene expression in the PAs.

   1. The *cis*-regulatory sequence elements are described in McEllin et al. (2015) and Davis et al. (2016).

   2. Follow steps II.B.1.b-II.B.1.f.iv. of the BIO-002 SOP to make transparent rectangles.

   3. Follow steps II.B.3.a.-II.B.3.e. of the BIO-002 SOP to make the Krox20, Sox, Hox/Pbx, and Prep/Meis labels.

D. Save the PowerPoint file.

E. Follow steps IV.A.-IV.F. of the BIO-002 SOP to convert the genomic DNA sequence alignment data into a TIFF image file. Save the file as "Hoxa2_Figure_3".

F. Add a new page to the "Hoxa2_Figures_Word" document and drag the "Hoxa2_Figure_3 TIFF file to this page.

G. Underneath the third figure, write a legend that describes:

   1. The layout of the figure (describe the alignments on each side of the figure)

2. The methods used to develop this figure (i.e.: what software programs were used to obtain the genomic DNA sequences, align the sequences, identify the functional *cis*-regulatory elements, etc.)
3. The yellow and blue color-coding process of the genomic DNA sequence alignments.

VII. Turn in the "Hoxa2_Figures_Word" file according to the deadline set by the instructor.

References:

Davis, A., M.C. Reubens, and E.J. Stellwag. 2016. Function and comparative genomics of *Hoxa2* gene *cis*-regulatory elements: evidence for evolutionary modification of ancestral core element activity. J Dev Biol 4(2):15.

Gendron-Maguire, M., M. Mallo, M. Zhang, T. Gridley. 1993. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75(7):1317-1331.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947-2948.

Maconochie, M.K., R. Krishnamurthy, S. Nonchev., P. Meier, M. Manzanares, P.J. Mitchell, and R. Krumlauf. 1999. Regulation of *Hoxa2* in cranial neural crest cells involves members of the *AP-2* family. Development 126(7):1483-1494.

Maconochie, M.K., S. Nonchev, M. Manzanares, H. Marshall, and R. Krumlauf. 2001. Differences in Krox20-dependent regulation of Hoxa2 and Hoxb2 during hindbrain development. Dev Biol 233(2):468-481.

McEllin, J.A., T.B. Alexander, S. Tümpel, L.M. Wiedemann, and R. Krumlauf. 2016. Analyses of fugu *hoxa2* genes provide evidence for subfunctionalization of neural crest cell and rhombomere *cis*-regulatory modules during vertebrate evolution. Dev Biol 409(2):530-542.

Nonchev, S., C. Vesque, M. Maconochie, T. Seitanidou, L. Ariza-McNaughton, M. Frain, H. Marshall, M.H. Sham, R. Krumlauf, and P. Charnay. 1996. Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. Development 122(2):543-554.

Rijli, F.M., M. Mark, S. Lakkaraju, A. Dierich, P. Dolle, and P. Chambon. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75(7):1333-1349.

Tümpel, S., F. Cambronero, L.M. Wiedemann, and R. Krumlauf. 2006. Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes). Proc Natl Acad Sci 103(14):5419-5424.

Tümpel, S., L.M. Wiedemann, and R. Krumlauf. 2009. Hox genes and segmentation of the vertebrate hindbrain. Curr Top Dev Biol 88:103-137.

SOP #	BIO-004
Title	Analysis of *Hoxa2* Genomic DNA Sequences II: Use of Clustal, Microsoft Word, and Microsoft PowerPoint Software to Analyze *Cis*-Regulatory Elements Responsible for Directing *Hoxa2* Gene Expression in Rhombomeres 3 and 5 and the Pharyngeal Arches.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

SOP Assessment

1. True or False: Intergenic DNA sequences are part of the coding domains of a gene (they are translated into amino acids).

2. Krox20 is a:

   a. Translation factor
   b. Transcription factor
   c. Paracrine factor
   d. Splicing factor

3. Krox20 functions to:

   a. Inhibit *Hoxa2* gene expression in r3 and r5
   b. Enhance *Hoxa2* gene expression in r3 and r5
   c. Enhance *Hoxa2* gene expression in r4
   d. Enhance *Hoxa2* gene expression in the PAs

Date: ____

Name (Print): ____

Signature: ____

SOP #	BIO-005
Title	Analysis of *Hoxa2* Genomic DNA Sequences III: Use of Clustal, Microsoft Word, and Microsoft PowerPoint Software to Analyze *Cis*-Regulatory Elements Responsible for Directing *Hoxa2* Gene Expression in Rhombomere 4.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

Objective:

To understand the function and evolution of *Hoxa2*, especially with directing the development of the rhombomere-derived cranial nerves and the pharyngeal arch-derived skeletal elements, the upstream genomic DNA sequences, which contain regulatory elements that direct the expression of *Hoxa2*, must be retrieved from the appropriate database for analysis. Furthermore, genomic DNA sequences from evolutionarily divergent species must be analyzed to fully understand how this protein functions. Genomic DNA sequences will be retrieved from several primate species, including Human (*Homo sapiens*), Chimpanzee (*Pan troglodytes*), Gorilla (*Gorilla gorilla*), and Orangutan (*Pongo abelii*) and several vertebrate biomedical models, including Zebrafish (*Danio rerio*), Chicken (*Gallus gallus*), and Mouse (*Mus musculus*). The software program, Clustal, will be employed for this analysis.

Relevant Terms and their Definitions

Hoxa2 – Developmental regulatory gene that is evolutionarily conserved and functions to pattern the development of the facial nerve.

National Center for Biotechnology Information (NCBI) – Web-based database containing genetic information submitted by scientists. Used for genetic, developmental, medical, ecological and evolutionary research analyses.

Genbank Accession number – Identifying number for genomic DNA sequences

FASTA – Sequence format that must be obtained for nucleotide sequences for downstream analyses.

Intronic regions – Regions of noncoding genomic DNA that span between exons of genes and that can harbor regulatory elements.

Clustal – Multiple sequence alignment program for DNA or proteins.

Indel mutations – Point mutation of a nucleotide that results in an insertion of an amino acid or deletion of a nucleotide.

Cis-regulatory elements (CREs) – short sequences of genomic DNA that bind transcription factor (TF) proteins and regulate the spatial and temporal expression patterns of their respective genes.

r4 enhancer region – region of genomic DNA within the intron of *Hoxa2* that contains CREs that direct *Hoxa2* expression in r4.

Hox/Pbx – Transcription factors that function to bind to *cis*-regulatory elements and direct *Hoxa2* gene expression in r4.

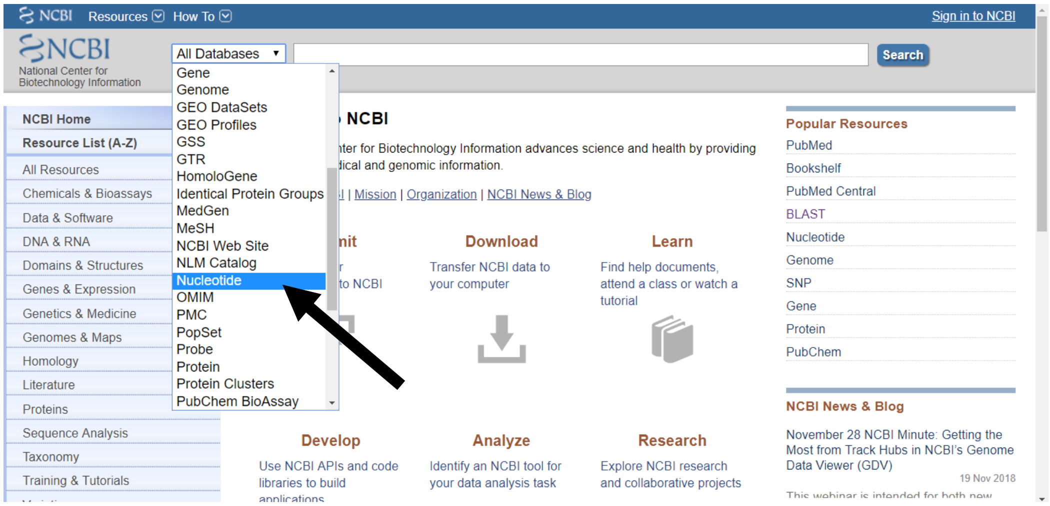
Prep/Meis – Transcription factors that function with Hox/Pbx factors to bind to *cis*-regulatory elements and direct *Hoxa2* gene expression in r4.

Procedure

I. Retrieve the CREs responsible for directing *Hoxa2* gene expression in r4 from the Genbank database.

A. Using an internet-based software program (e.g.: Google Chrome, Mozilla Firefox, Internet Explorer, etc.), type in https://www.ncbi.nlm.nih.gov

B. Click on the drop down menu titled "All Databases" and select "Nucleotide" (see image below).



C. Type in the appropriate Genbank accession number in the text box next to the Drop-down menu to retrieve the appropriate coding DNA sequences and click on the "Search" button. The Genbank accession numbers for the species-specific *Hoxa2* noncoding DNA sequences are listed in the table below.

Organism	***Hoxa2* Genbank Accession**	**Sequence Begin Coordinate**	**Sequence End Coordinate**	**Reverse Complement Necessary**
Human	AC004079	84164	84248	Yes
Chimpanzee	NC_036886	27218368	27218452	Yes
Gorilla	NC_018431	27156581	27156665	Yes
Orangutan	NC_036910	46403675	46403759	No
Mouse	CH466597	4373429	4373513	Yes
Chicken	NC_006089	32583964	32584047	Yes
Zebrafish	AL645795	57649	57741	Yes

D. Follow steps I.D.-I.G. of the BIO-004 SOP to retrieve the designated region of the *Hoxa2* intronic sequence in FASTA format and in the 5'-3' orientation for all of the organisms listed in the table above.

II. Follow steps II.A.-II.C. of the BIO-004 SOP to format the sequences in Microsoft Word so that they can be read and processed by the Clustal software program. Save the Word file as "Hoxa2_r4_Unaligned".

III. Follow steps III.A.1.-III.A.8. of the BIO-004 SOP to perform the primate and model vertebrate genomic DNA sequence alignments. Save the Word files as "Hoxa2_r4_Alignment_Primates" and "Hoxa2_r4_Alignment_ModelVerts".

IV. Follow steps VI.A.-VI.E. of the BIO-001 SOP for formatting the DNA sequence alignment documents for nucleotide color coding and eventual figure development.

V. Follow steps VII.A.-VII.D. of the BIO-001 SOP for color code the nucleotide data to highlight conserved DNA sequence regions.

VI. Develop a figure of the genomic DNA sequence alignments.

A. Follow steps VI.E.-VI.F. of the BIO-003 SOP for making a new slide in the "Hoxa2_Figures_Powerpoint" file.

B. Follow steps VIII.-IX. of the BIO-001 SOP for importing the primate and model vertebrate genomic sequence alignments into the new slide of the "Hoxa2_Figures_Powerpoint" file.

1. Label the Primate alignment data as "Primate *Hoxa2* Intronic Genomic DNA Sequence Alignment".

2. Label the Model Vertebrate alignment data as "Model Vertebrate *Hoxa2* Intronic Genomic DNA Sequence Alignment".

C. Label the Prep/Meis, and Hox/Pbx *cis*-regulatory elements in the Model Vertebrates alignment.

   NOTE 1: *Hox/Pbx* and *Prep/Meis* are transcription factors that work in conjunction to direct *Hoxa2* gene expression in r4.

   1. The *cis*-regulatory sequence elements are described in Tümpel et al. (2007).

   2. Follow steps II.B.1.b-II.B.1.f.iv. of the BIO-002 SOP to make transparent rectangles.

   3. Follow steps II.B.3.a.-II.B.3.e. of the BIO-002 SOP to make the Hox/Pbx and Prep/Meis labels.

D. Save the PowerPoint file.

E. Follow steps IV.A.-IV.F. of the BIO-002 SOP to convert the genomic DNA sequence alignment data into a TIFF image file. Save the file as "Hoxa2_Figure_4".

F. Add a new page to the "Hoxa2_Figures_Word" document and drag the "Hoxa2_Figure_4 TIFF file to this page.

G. Underneath the third figure, write a legend that describes:

   1. The layout of the figure (describe the alignments on each side of the figure)
   2. The methods used to develop this figure (i.e.: what software programs were used to obtain the genomic DNA sequences, align the sequences, identify the functional *cis*-regulatory elements, etc.)
   3. The yellow and blue color-coding process of the genomic DNA sequence alignments.

VII. Turn in the "Hoxa2_Figures_Word" file according to the deadline set by the instructor.

References:

Gendron-Maguire, M., M. Mallo, M. Zhang, T. Gridley. 1993. *Hoxa-2* mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. Cell 75(7):1317-1331.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947-2948.

Rijli, F.M., M. Mark, S. Lakkaraju, A. Dierich, P. Dolle, and P. Chambon. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. Cell 75(7):1333-1349.

Tümpel, S., F. Cambronero, L.M. Wiedemann, and R. Krumlauf. 2006. Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes). Proc Natl Acad Sci 103(14):5419-5424.

Tümpel, S., F. Cambronero, E. Ferretti, F. Blasi, L.M. Wiedemann, and R. Krumlauf, 2007. Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. Dev Biol 302(2):646-660.

Tümpel, S., L.M. Wiedemann, and R. Krumlauf. 2009. Hox genes and segmentation of the vertebrate hindbrain. Curr Top Dev Biol 88:103-137.

SOP #	BIO-005
Title	Analysis of *Hoxa2* Genomic DNA Sequences II: Use of Clustal, Microsoft Word, and Microsoft PowerPoint Software to Analyze *Cis*-Regulatory Elements Responsible for Directing *Hoxa2* Gene Expression in Rhombomere 4.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

SOP Assessment

1. True or False: Intronic DNA sequences are part of the coding domains of a gene (they are translated into amino acids).

2. Prep/Meis is a:

   a. Translation factor
   b. Transcription factor
   c. Paracrine factor
   d. Splicing factor

3. Prep/Meis functions to:

   a. Inhibit *Hoxa2* gene expression in r3 and r5
   b. Enhance *Hoxa2* gene expression in r3 and r5
   c. Enhance *Hoxa2* gene expression in r4
   d. Inhibit *Hoxa2* gene expression in the PAs

Date: ____

Name (Print): ____________________________________

Signature: _______________________________________

SOP #	BIO-006
Title	Performing whole-mount *in situ* hybridization (WISH) using zebrafish (*Danio rerio*) embryos.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

Objective:

To understand how genes function to pattern anatomical structures, their spatial and temporal expression patterns must be deduced. Whole-mount *in situ* hybridization will be performed on zebrafish to observe where *Hoxa2* is expressed during embryonic development.

Relevant Terms and their Definitions

Whole-mount *in situ* hybridization (WISH) – a technique used by biologists to understand when and where genes are transcribed into messenger RNA within tissues. A labeled antisense riboprobe is used to hybridize with expressed mRNAs

Antisense riboprobe – a RNA strand that is complementary to mRNA of interest and labeled with digoxigenin (DIG). This will hybridize with transcripts that are expressed in tissues. These are typically 300-1000 nucleotides in length to add for specificity in hybridizing to targeted gene transcripts. Roughly every 1 in every 4 uracil nucleotides of the riboprobe is covalently bonded to DIG. DIG provides an antigen target for the anti-DIG antibody.

Sense riboprobe – a RNA strand that is in the same orientation as the mRNA of interest and labeled with DIG. This is the same length as the antisense riboprobe and is used as a control. Since it is in the same orientation as the mRNA of interest, it will not hybridize with the mRNA.

Anti-DIG Antibody-Alkaline phosphatase (ADAAP) – An antibody that binds to DIG on the DIG-labeled antisense and sense riboprobes. It is coupled to an alkaline phosphatase enzyme that yields a purple fluorescence when in the presence of its substrate.

Procedure

I. Obtain appropriate developmentally staged zebrafish embryos for analysis.

A. Using a bulb pipette, carefully transfer 5-6 embryos per developmental stage into a 1.7 mL microfuge tube.

NOTE 1: To save on materials, one microcentrifuge tube per analysis will be used. Different developmentally staged embryos show differential morphological features.

NOTE 2: Embryos are stored long-term in 100% methanol at -20 ºC. 100% methanol permeabilizes tissues to allow for access to endogenous mRNAs. Embryonic tissue is originally "fixed" with 4% paraforlmaldehyde (PFA) dissolved in phosphate buffered saline with 0.1% Tween 20 (0.137 M NaCL, 0.0027 M KCl, 0.01 M Na_2HPO_4, 0.0018 M KH_2PO_4, pH 7.4) (PBST). PFA cross-links proteins to preserve tissues. Methanol dissolves lipids to allow access of RNA probes and antibodies to their intracellular targets for the WISH analysis.

B. Transfer excess 100% methanol back to original stock tube.

II. Re-hydrate embryos.

NOTE 3: In order to allow any enzymatic activity to occur within the embryo during the WISH assay, the embryonic tissue must be hydrated and equilibrated in PBST. Most enzymatic reactions that take place within living tissue occur in an environment that contains readily available phosphate (PO_4^-) ions and is at a pH of 7.4. Tween 20 within PBST serves as a surfactant and prevents embryos from adhering to each other and potentially destroying tissue during subsequent washes.

NOTE 4: When replacing liquids from embryos, always be sure to leave embryos submerged within liquids. Otherwise the embryos will become dessicated and the tissue will be destroyed. Be sure that all embryos have fallen to the bottom of the 1.7 mL microfuge tube before replacing any liquids.

A. Remove 100% methanol from embryos in the 1.7 mL microcentrifuge tube and discard.

B. Gently add 1 mL of PBST to embryos, cap tube and incubate embryos at room temperature (RT) for 5 minutes (min).

C. Remove PBST from embryos and discard.

D. Gently add 1 mL of fresh PBST to embryos and incubate at RT for 5 min.

E. Repeat steps II.C. through II.D. three more times for a total of five PBST washes.

III. Digest embryos.

NOTE 5: In order to observe gene expression in deeply located tissue, embryos must be partially digested with Proteinase K. Proteinase K digests protein, thus exposing deeper tissues for riboprobe penetration. The Proteinase K stock solutions are at 10 mg/mL (or 10,000 μg/mL) in dH_2O. Working solutions must be at 10 μg/mL.

A. Add 1 μl of 10,000 μg/mL Proteinase K to the 1.7 mL microfuge tube containing the embryos and 1 mL of PBST.

B. Incubate embryos at RT for the appropriate amount of time depending on the developmental stage (see Table below).

Zebrafish Developmental Stage	Proteinase K Digestion Time
75% Epiboly (Gastrula)	30 sec
> Gastrula to 18-20 somites	1 min
> 18-20 somites to 24 hours (hr)	10 min
> 24 hr to < 48 hr	20 min
48 hr and later	30 min

C. Once the Proteinase K incubation time has ended, immediately remove the PBST with 10 μg/mL Proteinase K and discard.

D. Gently add 1 mL of 4% PFA to the embryos and incubate tube at RT for 20 min.

NOTE 6: 4% PFA is added to deactivate the Proteinase K so no further digestion of embryos will occur.

IV. Hybridize riboprobes to their endogenous mRNA targets.

A. Remove 4% PFA from embryos and discard.

B. Gently add 1 mL of PBST to the embryos and incubate tube at RT for 5 min.

C. Remove PBST from embryos and discard.

D. Repeat steps IV.B.-IV.C. four more times for a total of five PBST washes.

E. Remove final wash of PBST from embryos and discard.

F. Gently add 500 μL of Hybridization buffer (HB) to the embryos and incubate at 65 ºC for at least 5 min.

NOTE 7: It may take more than 5 min for all embryos to completely equilibrate in the HB and sink to the bottom of the microfuge tube. The HB is an extremely viscous solution and it will take a longer duration for embryos to equilibrate in HB than in other solutions.

NOTE 8: The HB is composed of 50% formamide, 50 μg/mL Heparin, 5X sodium saline citrate (0.75M NaCl, 0.075M $Na_3C_6H_5O_7$) (SSC), 500 μg/ mL torula RNA, and 0.1% Tween 20 and is at a pH of 6.0. All components of the HB, including the pH of 6.0 and the temperature at 65 ˚C, increase stringency and specificity of the antisense riboprobe to its endogenous target and decrease non-specific binding of the riboprobe to other mRNAs. The high temperature of 65 ºC and formamide, which is an organic solvent, both function to denature (melt) hydrogen bonds between the riboprobe and non-specific mRNA targets. The pH of 6.0 increases the concentration of free H^+ ions, which compete with the riboprobe for mRNA targets. Heparin and torula (yeast) RNA function as "blocking" reagents that prevent the riboprobe from binding to non-specific mRNA targets. 5X SSC is a high salt concentration and aids in stabilizing the riboprobe to its endogenous mRNA temperature and low pH. Thus, the HB is added to embryos to aid in allowing the antisense riboprobes to hybridize only to their endogenous mRNA targets.

NOTE 9: If it is not feasible to continue with the WISH assay due to time issues, embryos can be stored in HB buffer at -20 ºC until the assay can be continued.

G. Remove the HB from the embryos and discard.

H. Gently add 500 mL fresh HB to the embryos and incubate at 65 ºC for 1 hr.

I. Add 1 μL of the appropriate digoxigenin (DIG)-labeled riboprobe to the embryos, gently tap the tube to mix, and incubate at 65 ºC for at least 16 hr.

NOTE 10: Be sure to add only antisense riboprobe or sense riboprobe to each microfuge tube of embryos.

V. Wash excess DIG-labeled riboprobes from embryos.

NOTE 11: The washing of embryos post-hybridization is extremely important as it will remove any non-hybridized DIG-labeled riboprobes that can potentially generate background noise.

A. Remove HB with DIG-labeled antisense or sense riboprobe and discard.

B. Gently add 1 mL Wash 1 solution (50% formamide, 2X SSC, 0.1% Tween 20 (SSCT) to the embryos and incubate at 65 ºC for 1 hr.

NOTE 12: A decreased SSC concentration aids in de-stabilizing hydrogen bonding of DIG-labeled antisense riboprobes to any non-specific mRNA targets.

C. Remove Wash 1 solution and discard.

D. Gently add 1 mL fresh Wash 1 solution to the embryos and incubate at 65 ºC for 1 hr.

E. Repeat steps V.C.-V.D. two more times for a total of four Wash 1 washes.

F. Remove Wash 1 solution and discard.

G. Gently add 1 mL Wash 2 solution (2X SSCT) to the embryos and incubate at 65 ºC for 1 hr.

H. Remove Wash 2 solution and discard.

I. Gently add 1 mL Wash 3 solution (0.2X SSCT) to the embryos and incubate at 65 ºC for 1 hr.

J. Remove Wash 3 solution and discard.

K. Gently add 1 mL fresh Wash 3 solution to the embryos and incubate at 65 ºC for 1 hr.

NOTE 13: At this point of the Wash steps, the only mRNA that the antisense DIG-labeled riboprobe should be hybridized to is its endogenous, complementary sense mRNA target.

L. Remove Wash 3 solution and discard.

M. Gently add 1 mL PBST to the embryos, cap tube, and incubate at RT for 5 min.

N. Remove PBST and discard.

O. Gently add 1 mL fresh PBST to the embryos, cap tube, and incubate at RT for 5 min.

NOTE 14: If it is not feasible to continue with the WISH assay due to time issues, embryos can be stored in PBST at 4 ºC until the assay can be continued.

VI. Add Anti-DIG Antibody-Alkaline Phosphatase (ADAAP) to the embryos.

A. Remove PBST and discard.

B. Gently add 1 mL Blocking solution (2 mg/mL bovine serum albumin (BSA) in PBST) to the embryos, cap tube and incubate embryos at RT for at least 90 min.

NOTE 15: The BSA is protein that functions as a blocking agent. Specifically, it helps to limit the antibody from binding to any targets other than the DIG target on the riboprobe.

C. Remove blocking solution and discard.

D. Gently add 1 mL 1:5000 Anti-DIG Antibody-AP in PBST to the embryos, cap tube and incubate for at least 12 hr at 4 ºC.

NOTE 16: The 4 ºC temperature is used to decrease the rate of reaction of the antibody-antigen binding. This decreased temperature also aids in allowing the antibody to specifically bind to its DIG target. A 1:5000 dilution is made by adding 1:100 diluted antibody to 980 µL PBST.

VII. Wash excess Anti-DIG Antibody-AP from embryos.

NOTE 17: The washing of embryos post antibody incubation is performed to remove any non-bound ADAAP that can create background noise.

A. Remove ADAAP in PBST and discard.

B. Gently add 1 mL PBST to the embryos and incubate at RT for 15 min.

C. Remove PBST and discard.

D. Gently add 1 mL fresh PBST to the embryos and incubate at RT for 15 min.

E. Repeat steps VII.C.-VII.D. six more times for a total of eight washes in PBST.

NOTE 18: If it is not feasible to continue with the WISH assay due to time issues, embryos can be stored in PBST at 4 ºC until the assay can be continued.

VIII. Perform labeling reaction.

NOTE 19: The labeling reaction involves transitioning embryos from PBST to Alkaline phosphatase (AP) buffer (0.1 M Tris, pH 9.5, 0.05 M $MgSO_4$, 0.1 M NaCl, 0.1% Tween 20). This buffer provides the optimal pH environment for AP enzyme to function. This reaction also calls for nitro-blue tetrazolium chloride (NBT) and 5-bromo-4-chloro-3'-indolylphosphate p-toluidine salt (BCIP), which function as cofactors for the AP enzyme.

NOTE 20: The AP buffer must be made fresh for each WISH assay.

A. Remove PBST and discard.

B. Gently add 1 mL AP buffer to the embryos, cap tube, and incubate at RT for 5 min.

C. Remove AP buffer and discard.

D. Gently add 1 mL fresh AP buffer

E. Remove AP buffer and discard.

F. Gently add AP buffer with 0.4 mg/mL NBT and 0.2 mg/mL BCIP.

NOTE 21: The NBT and BCIP must be measured and added to the AP buffer prior to adding to the embryos. Once the NBT and BCIP are added to the AP buffer, the solution must be kept in a light-free zone. Covering the tube with foil works well.

G. Immediately after adding the AP buffer with NBT and BCIP, transfer the embryos and buffer from the 1.7 mL microfuge tube to a labeled well of a 24-well plate.

NOTE 22: Embryos can be visualized more easily under a microscope once they are transferred to a 24-well plate.

H. Cover the well plate with foil to inhibit any light from disrupting the reaction and incubate embryos at RT.

I. Check the progress of the labeling reaction using a light microscope every 30 min to 1 hr.

NOTE 23: The embryos can be stored overnight at 4 ºC to slow down the labeling reaction if it is not feasible to continue the labeling reaction at this time. The labeling reaction can be continued at RT the next day.

J. Embryos that show desirable gene expression must be transferred into 4% PFA using tweezers to stop the labeling reaction.

NOTE 24: Embryos that are treated with 4% PFA can be stored at 4 ºC overnight to ensure that the alkaline phosphatase enzyme has been deactivated.

K. Once the AP enzyme is deactivated, remove the 4% PFA and replace with PBST.

NOTE 25: Embryos can be stored in PBST at 4 ºC until they are prepared for mounting and digital microscopic photography.

References:

Davis, A., J.L. Scemama, and E.J. Stellwag. 2008. Japanese medaka *Hox* paralog group 2: insights into the evolution of *Hox* PG2 gene composition and expression in the Osteichthyes. J Exp Zool (Mol Dev Evol) 310(8):623-641.

Prince, V. E., L. Joly, M. Ekker, and R.K. Ho. 1998. Zebrafish *Hox* genes: Genomic organization and modified colinear expression patterns in the trunk. Development 125(3)**:**407–420.

SOP #	BIO-006
Title	Performing whole-mount *in situ* hybridization (WISH) using zebrafish (*Danio rerio*) embryos.
Author	Adam Davis, Ph.D. Department of Biology University of North Georgia

SOP Assessment

1. An antisense riboprobe is complementary to:

    a. The DNA expressed in the embryo
    b. The mRNA expressed in the embryo
    c. The protein expressed in the embryo
    d. The promoter expressed in the embryo

2. 4% PFA functions to cross-link __________ to preserve embryonic ___________.

    a. DNAs; Proteins
    b. Proteins; Tissues
    c. RNAs; DNAs
    d. DNAs; RNAs

3. All components of the HB buffer function to ____________ specificity of the __________ riboprobe to its ___________ target.

    a. Decrease; Sense; DNA
    b. Increase; Sense; Protein
    c. Increase; Antisense; RNA
    d. Decrease; Antisense; RNA

Date: ____

Name (Print): ____

Signature: ____