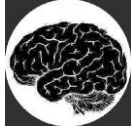
	<p style="text-align: center;">SKILLS CENTER</p> <p style="text-align: center;">STANDARD OPERATING PROCEDURE</p>	<p style="text-align: center;">A BIOFIZZ</p>  <p style="text-align: center;">PRODUCTON</p>
<p style="text-align: center;">Introduction to Jalview</p> <p style="text-align: center;">Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>


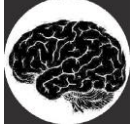
Background:

Sequence alignment in a nutshell is simply the process of comparing biological sequences. These “similarities” being detected refer to the number of identical residues/nucleotides across sequences.



Researchers have used sequence editing for decades to determine the evolutionary heritage of organisms. Over the course of evolution, much of the code for life is maintained. The DNA polymerase of a mouse has a near identical sequence to the DNA polymerase of a human. A DNA polymerase in an E. Coli will have a similar sequence to a human but not as similar as a mouse. This data shows us that a mouse is more closely related to a human than an E. Coli bacterium, but it also points to a common “code of life” inferring that organisms are all related. Sequence alignment helps us determine how closely related we are.

In addition to evolutionary study, sequence alignment helps us compare structures to determine functions. An **alignment score** measures the number of residues that are identical in the sequences being compared. If the alignment score is high, the sequence is closer to being identical. A more highly related sequence often points to a similar structure, and consequently, a similar function. If we were studying an abnormal protein found in stomach cells, but discovered its sequence alignment was very similar to a proteasome, we could infer that the protein probably has a function that is similar to a proteasome. While this strategy is often effective, comparing actual structure rather than just alignment score is necessary to form these inferences. Sometimes, 2 proteins could have a completely different sequence but a very similar structure. Additionally, 2 proteins could have a very similar sequence but have a very different structure, depending on the properties of the residues. However, alignment score is a piece of the puzzle that can be used to draw conclusions.

	<p align="center">SKILLS CENTER</p> <p align="center">STANDARD OPERATING PROCEDURE</p>	<p align="center">A BIOFIZZ</p>  <p align="center">PRODUCTON</p>
<p align="center">Introduction to Jalview</p> <p align="center">Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>

While incredibly applicable, protein sequence analysis is difficult to do without a suitable program to store and compare a large number of sequences. Manually comparing protein sequences is tedious and costly. However, with a computer program to line up the data for us, it becomes easier and more manageable to run multiple sequence alignments.

Jalview is a free, cross-platform program used for multiple sequence alignment editing, visualization and analysis. First developed in 1995 by the University of Oxford, Jalview has been run by the University of Dundee since 2001. In addition to its sequence alignment capabilities, Jalview has built in DNA, RNA and protein sequence and structure visualization and analysis capabilities. It provides a linked view of aligned DNA and Protein products.

1. Purpose

The purpose of this procedure is to recognize the multitude of applicable uses of Jalview sequence analysis and to familiarize oneself with its most common applications.

2. Scope


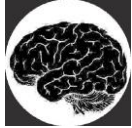
- This procedure applies to qualified skills center users.

3. Responsibility

- It is the responsibility of the user to understand and perform the procedure described in this document.
- It is the responsibility of the user performing the procedure to fully document any deviations from the written procedure (in a computer lab, document your findings using the database.)
- It is the responsibility of the user to become trained in the use of this application.

4. Definitions

- RCSB Protein Data Bank – A database that houses information about 3D macromolecule structures, such as proteins and nucleic acids.
- Residue – an amino acid within a polypeptide sequence (protein)
- Sequence – chain of amino acids

	<p style="text-align: center;">SKILLS CENTER</p> <p style="text-align: center;">STANDARD OPERATING PROCEDURE</p>	<p style="text-align: center;">A BIOFIZZ</p>  <p style="text-align: center;">PRODUCTON</p>
<p style="text-align: center;">Introduction to Jalview</p> <p style="text-align: center;">Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>

- JalView – an application that allows for detailed protein, DNA, and RNA sequencing, structure visualization, and tree creation

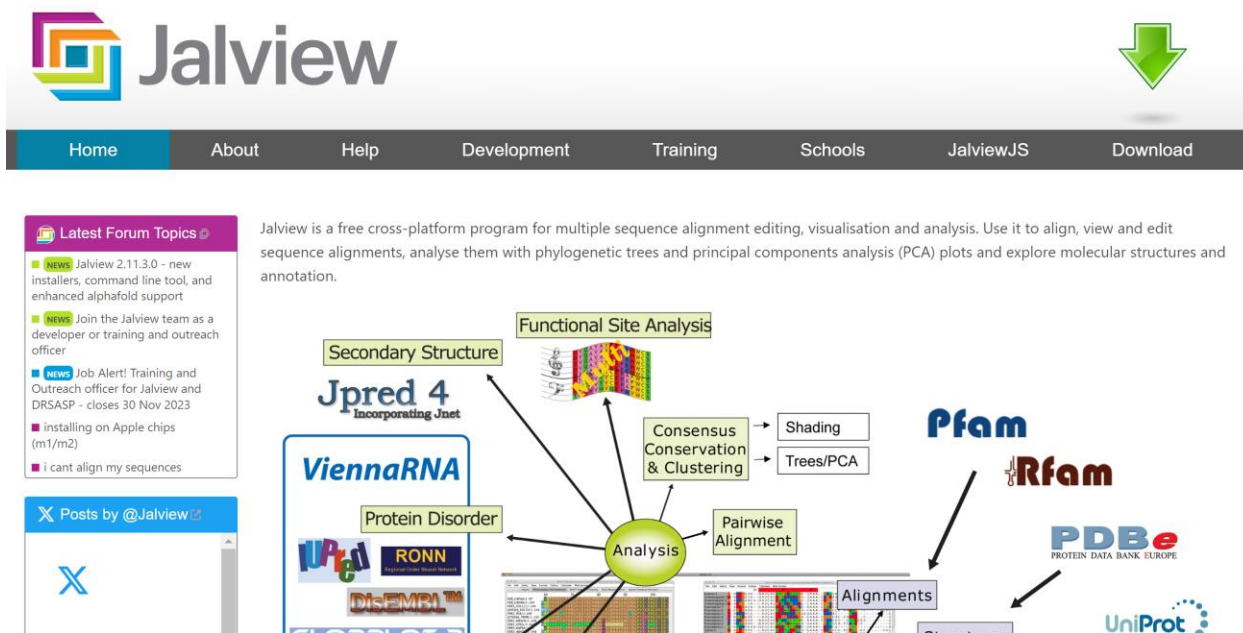
5. Materials/Equipment

- Jalview Application
- RCSB PDB

6. Procedures


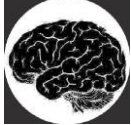
- 6.1 Downloading the Jalview Application:

- o Go to the following link and hover your mouse over the “download” tab on the top right bar of the home page: <https://www.jalview.org/>

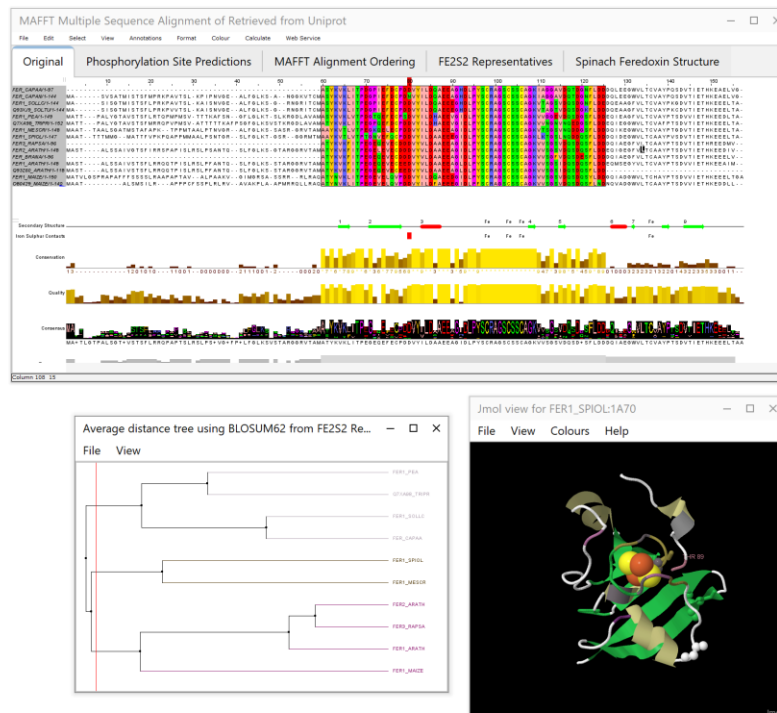


Jalview is a free cross-platform program for multiple sequence alignment editing, visualisation and analysis. Use it to align, view and edit sequence alignments, analyse them with phylogenetic trees and principal components analysis (PCA) plots and explore molecular structures and annotation.


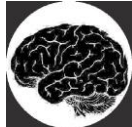
- o Click the “download” tab which corresponds to your computer (windows, mac, etc.).

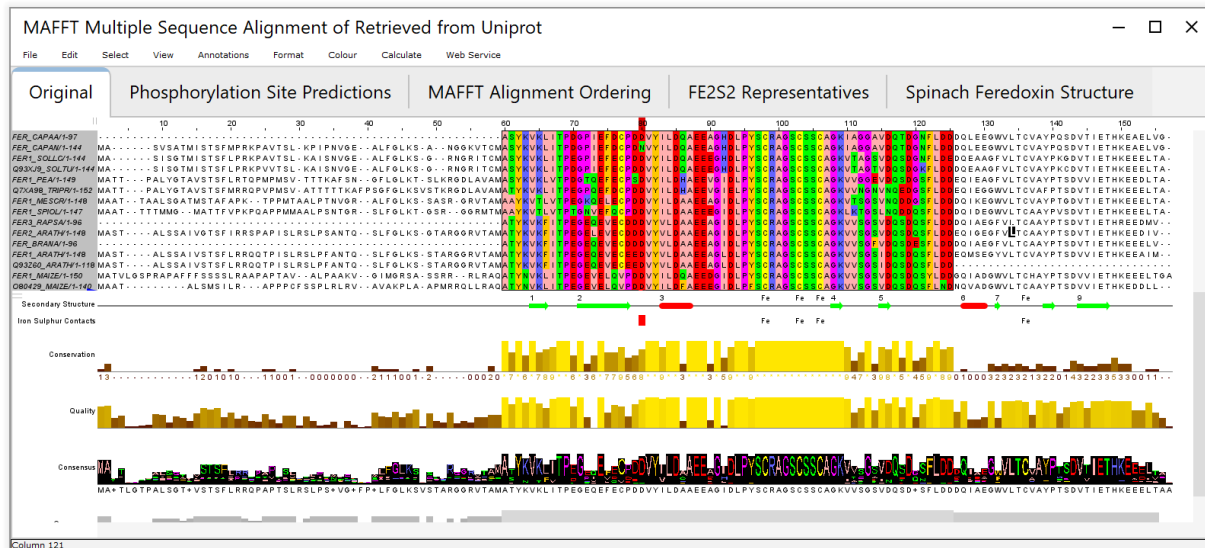
	<p style="text-align: center;">SKILLS CENTER</p> <p style="text-align: center;">STANDARD OPERATING PROCEDURE</p>	<p style="text-align: center;">A BIOFIZZ</p>  <p style="text-align: center;">PRODUCTON</p>
<p style="text-align: center;">Introduction to Jalview</p> <p style="text-align: center;">Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>

- Open the application from your files.
- **6.2 Navigating the Home Page:**
 - When you open the application, you will notice that three windows appear. One window is the alignment window, one is the structure window, and the final one is the tree window as shown below:





- In the alignment window, there are three panels: the sequence ID, the sequence alignment, and the annotation panels.

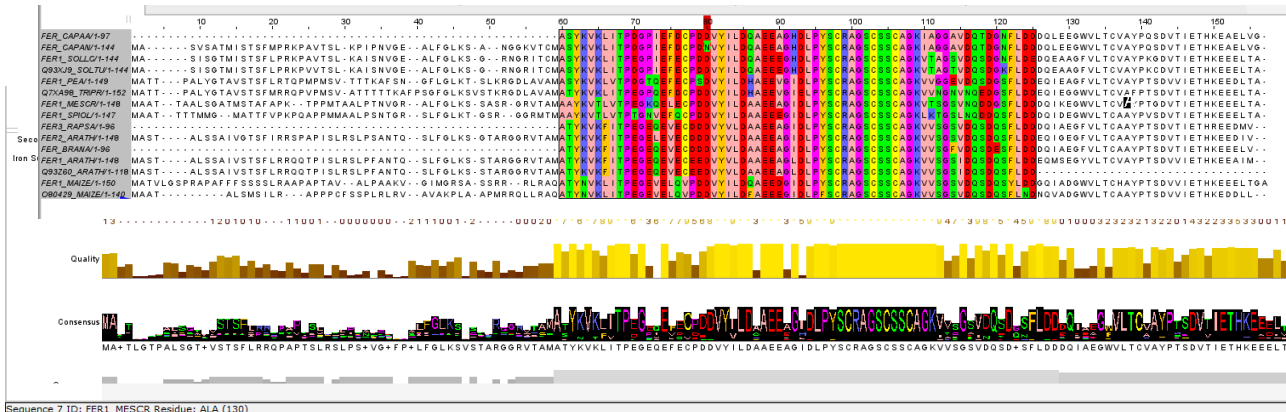
	<p align="center">SKILLS CENTER</p> <p align="center">STANDARD OPERATING PROCEDURE</p>	<p align="center">A BIOFIZZ</p>  <p align="center">PRODUCTON</p>
<p align="center">Introduction to Jalview</p> <p align="center">Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>



- In the alignment window you can locate residues very easily with the “curser” feature: hit the F2 key to go into curser mode and click on a residue (it should appear highlighted in black). If you are not in curser mode, it will merely have a black outline.
- You can locate residues by typing the column of interest first “,” followed by the row of interest, then type “enter.” For example, if I wanted to identify the 140th column in the 2nd row’s residue, I would type, 140,2 “enter.”


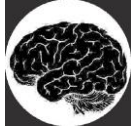
	<p style="text-align: center;">SKILLS CENTER</p> <p style="text-align: center;">STANDARD OPERATING PROCEDURE</p>	<p style="text-align: center;">A BIOFIZZ</p>  <p style="text-align: center;">PRODUCTON</p>
<p style="text-align: center;">Introduction to Jalview</p> <p style="text-align: center;">Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>

- In the bottom left corner of the window, there is information about that specific residues ID and information (in this example I am looking at residue 137,7).

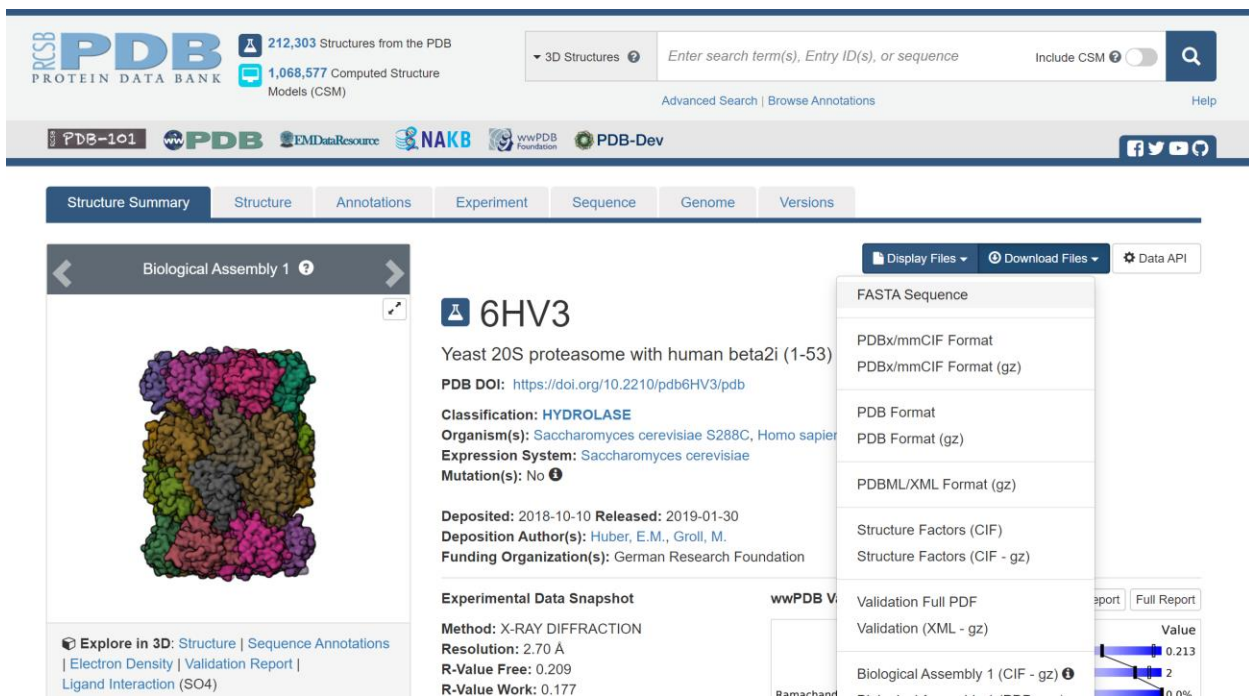


- You can also search the alignment panel for a common sequence of residues. Click “Select” followed by “find” (or ctrl F) and then type in your desired sequence. Make sure the entire sequence is selected before using the “find” feature. Do this by clicking and dragging your mouse across the whole sequence (or ctrl A).
- You can hide columns or rows by clicking on the desired sequence, “right click” and select hide sequence, or type ctrl H. This can be undone by right clicking the blue arrow that appears when a sequence is hidden, and selecting the “reveal all” tab.
- You can reorder sequences by selecting the sequence and typing the up or down arrow.
- You can change the highlighted color of the residues by selecting “color.” There are multiple ways you can sort your residues (like hydrophobicity, percent identity, etc.).

- 6.3 Importing a protein of interest:


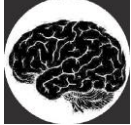
	<p align="center">SKILLS CENTER</p> <p align="center">STANDARD OPERATING PROCEDURE</p>	<p align="center">A BIOFIZZ</p>  <p align="center">PRODUCTON</p>
<p align="center">Introduction to Jalview</p> <p align="center">Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>

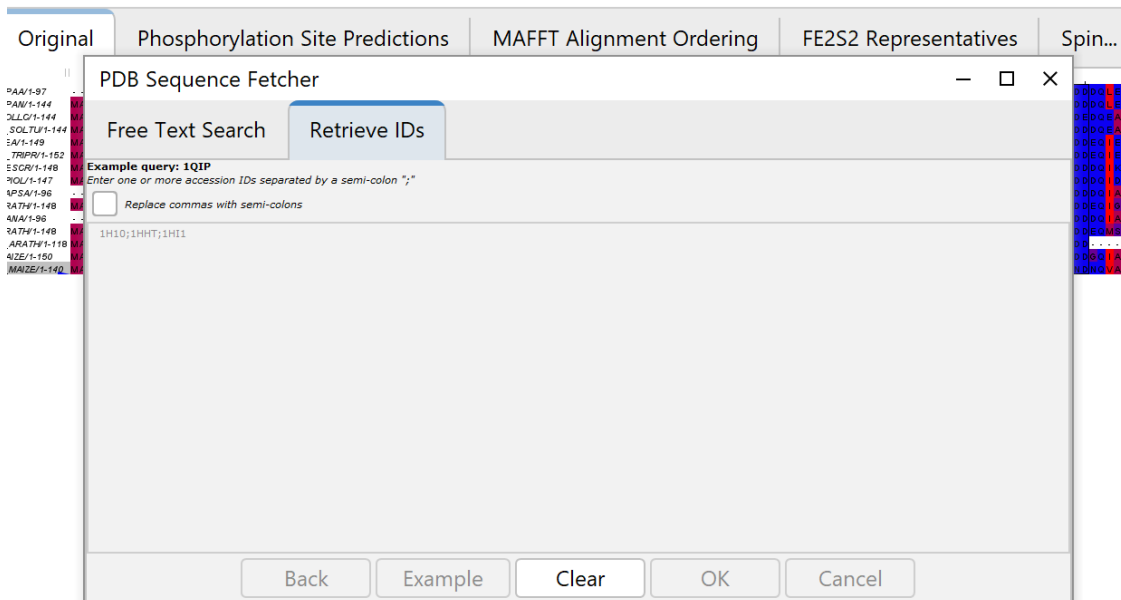
- First, select a protein of interest from the PDB database:
- Visit the RCSB databank website: <https://www.rcsb.org/>
- Search for a protein of interest, and note the ID. The ID in this example is 6HV3. For the module mastery tasks, you will be selecting 3 related proteins of interest.



The screenshot shows the RCSB PDB website interface. At the top, there are navigation tabs for Structure Summary, Structure, Annotations, Experiment, Sequence, Genome, and Versions. The main content area displays the entry for 6HV3, including a 3D structure visualization and detailed metadata. A dropdown menu is open, showing options for downloading files in various formats like FASTA, PDB, and CIF.

- Return to the Jalview app and upload your sequences: select “file” “fetch sequences” “retrieve IDs” and input the IDs of your chosen related proteins separated by semicolons. Be careful not to import the sequences into the list originally in the app. Make sure to click the top left “file” rather than the alignment window “file.”

	<p align="center">SKILLS CENTER</p> <p align="center">STANDARD OPERATING PROCEDURE</p>	<p align="center">A BIOFIZZ</p>  <p align="center">PRODUCTON</p>
<p>Introduction to Jalview</p> <p>Module Hours: 2.5</p>	<p>Effective Date: 12/1/2023</p> <p>PRQs:</p>	<p>Revision 1.0</p> <p>Author: D. Roberts</p> <p>Checked by Editor: M. Stowell</p>


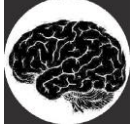


7. Troubleshooting

- If the application will not run on your PC because it is labelled as “not safe,” click “more information” and then at the bottom of the pop up, click “run anyway”
- You might need to update the app to the existing installation: just follow the prompts.
- The Jalview website has some good training videos on the home page that might be helpful if you are confused with any of the tasks.

8. References

“Jalview Home Page.” *Jalview*, www.jalview.org/. Accessed 13 Oct. 2023.

	SKILLS CENTER STANDARD OPERATING PROCEDURE	A BIOFIZZ  PRODUCTON
Introduction to Jalview Module Hours: 2.5	Effective Date: 12/1/2023 PRQs:	Revision 1.0 Author: D. Roberts Checked by Editor: M. Stowell

9. Module Mastery Task (MMT)

- 9.1 Open the app: on the home screen, what residue is at row 13, column 143?
- 9.2 What is the sequence ID of this residue?
- 9.3 Use the “find” function to determine the number of “EEE” sequences are there.
- 9.4 Rearrange your sequences and submit a resulting image. Why might reordering sequences be beneficial to researchers?
- 9.5 Hide the FER3_RAPSA sequence and submit the resulting image.
- 9.6 Select 5 related proteins of interest from the RCSB PDB database and load them into the application. Submit the resulting image. HINT: it might be helpful to find the same protein of 5 different organisms.
- 9.7 Which proteins did you choose and why did you choose them? What were their PDB ID’s? What function do these proteins serve?
- 9.8 Apply a percentage identity color filter on your sequences. Submit a resulting image. What does a darker shade tell you about your sequence comparison? Were you surprised by the results? Why or why not?
- 9.9 Save this file and be prepared to use it for the next JalView computer module. Make sure you save the whole file (click the very top left file rather than the alignment file).
- 9.10 Write a concise conclusion that encompasses the following key elements:
 - o Restate the purpose of the Module Mastery Task (ie. the objectives you were expected to achieve).
 - o Describe the most significant lessons or insights you gained from completing the tasks.
 - o Explain how you intend to apply the knowledge and skills acquired during this module in your academic or professional pursuits.

10. Feedback

- This is your space to help make this module better! Which aspects of the module did you like? Which aspects can be improved upon? Do you have any remaining areas of confusion? Did the number of hours assigned to this SOP accurately reflect the time spent on this skill?