

Requirements



Python version 2.7 or newer -
<https://www.python.org/downloads/>



Python Libraries installed if not already

Bio Python – pip install biopython

Pillow – pip install Pillow

Tkinter – included in python installation

Numpy – pip install numpy



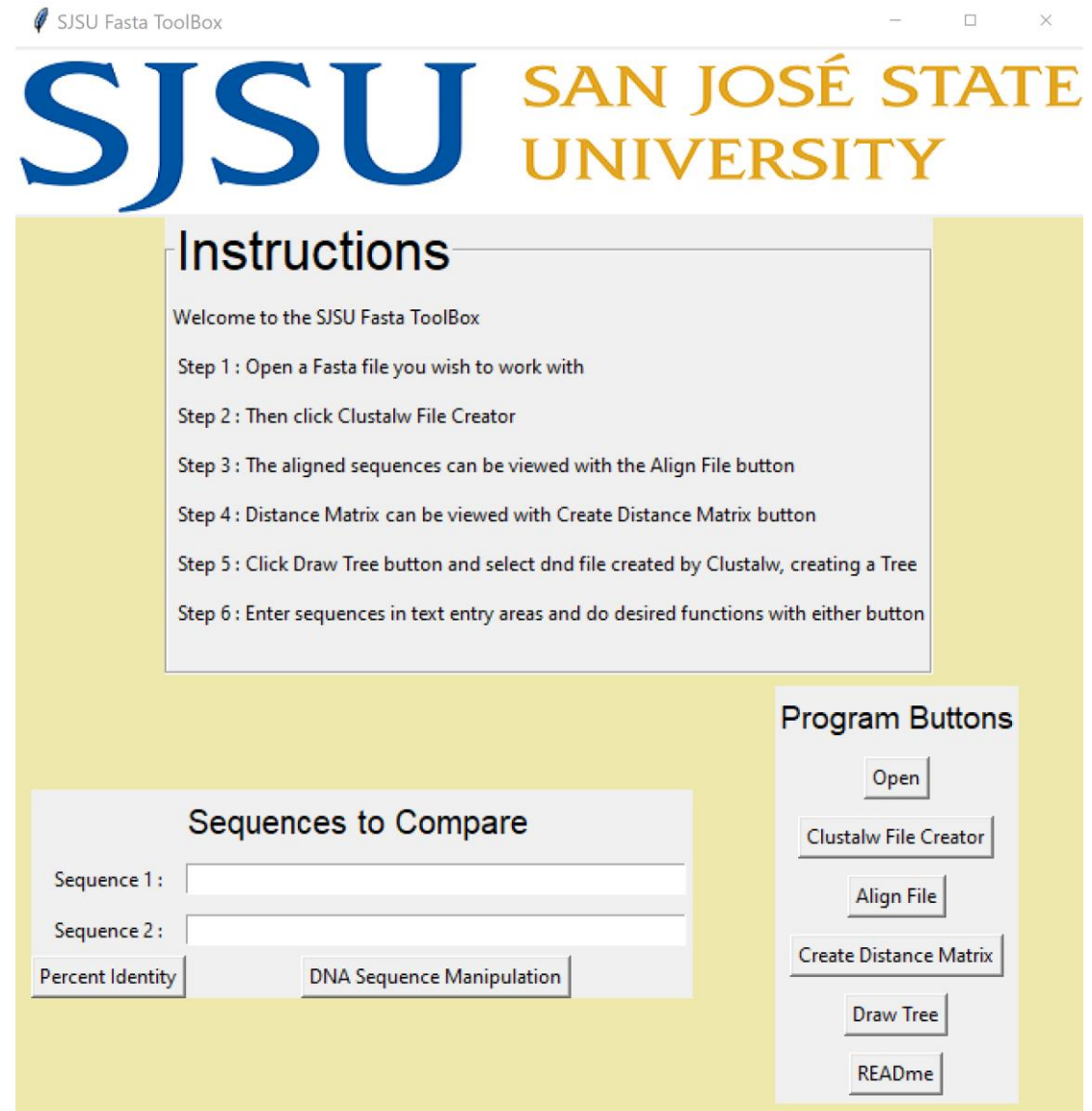
Install Clustalw -
<http://www.clustal.org/clustal2/#Download>



Fasta File

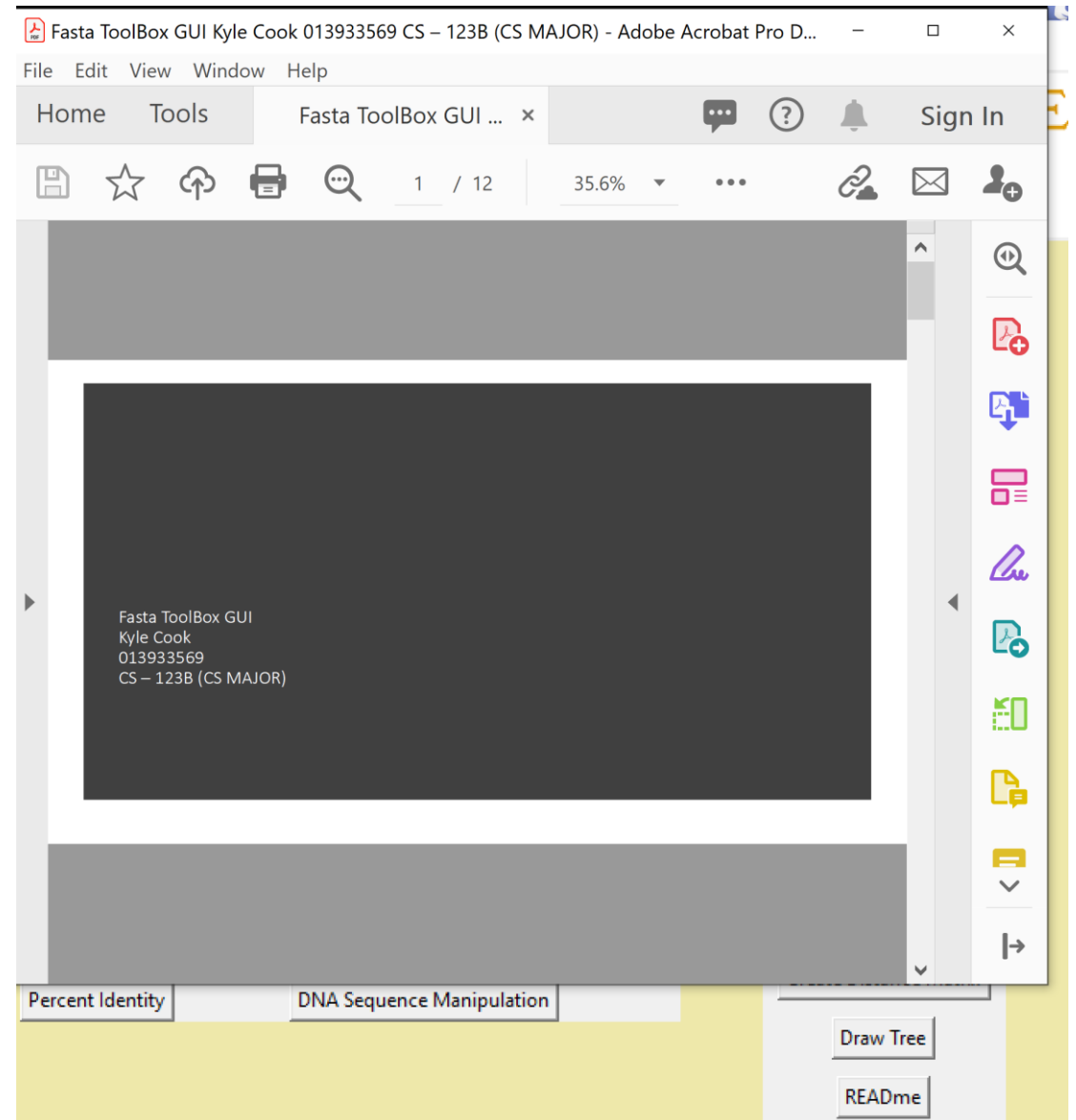
Program

- Once the program is running, this GUI will appear
- The GUI was created using Tkinter a library within python
- I have created 4 frames that hold all the GUI components
- Each button or entry box is linked to a function to perform desired output



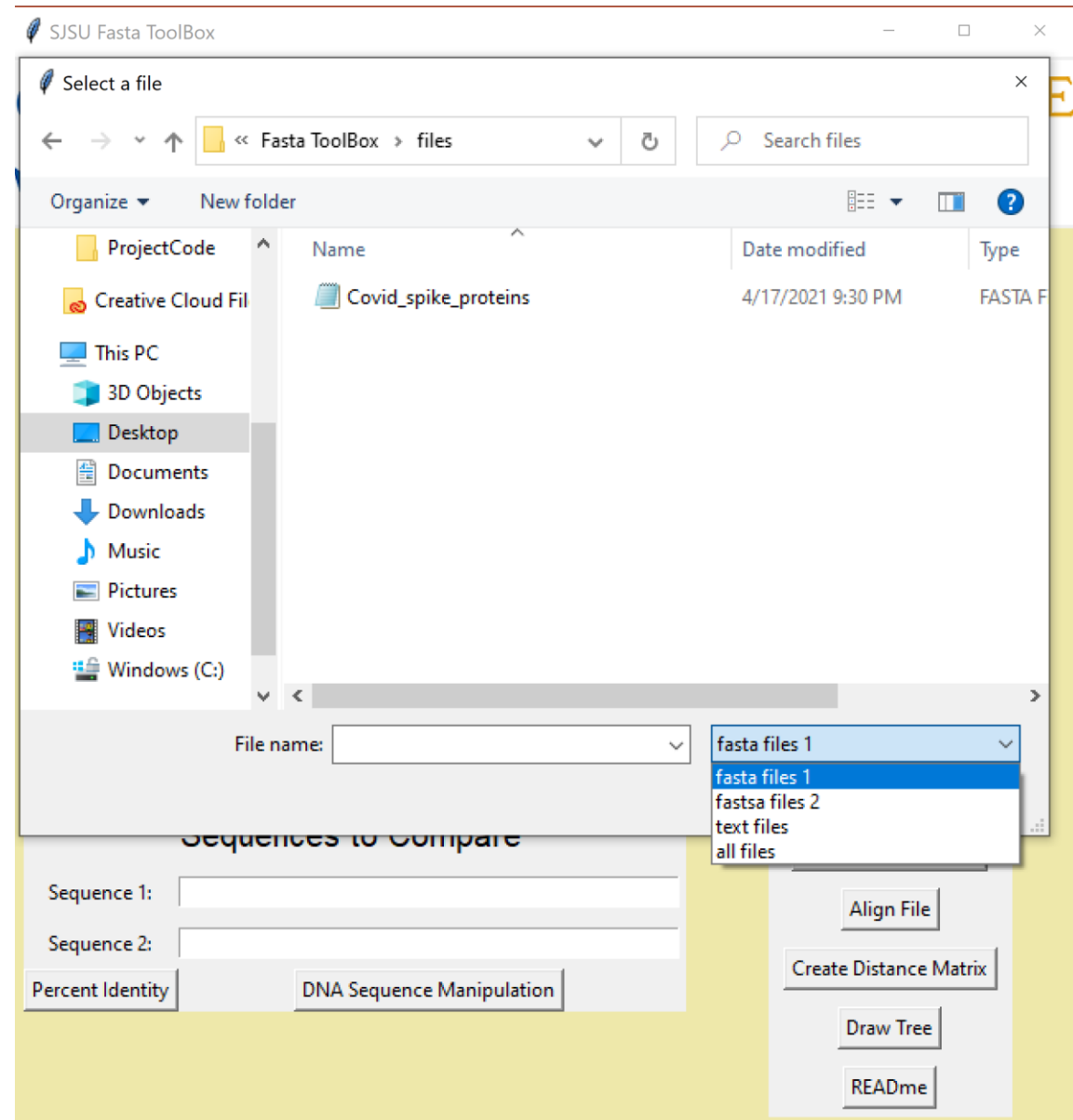
READme

- The READme button calls a function that will open a PDF version of these slides, so users have explanations and directions



Open File

- Once Open File is clicked it will prompt user to select a file
- It prompts desired file types, but user can click drop down menu and select which file type they want
- This file is then opened and stored within the program



Clustalw

- Once Clustalw button is clicked it will perform the Clustalw.exe which will create an alignment file (.aln) and a tree (.dnd), both files are used for other functions
- ** Bug
 - Clustalw takes a long time if fasta file is very long, in order to keep program from crashing, do not click anything while the Clustalw Button is sunken it, this means it is performing the file creations and if interrupted it will crash program.

SJSU Fasta ToolBox

SJSU SAN JOSÉ STATE UNIVERSITY

Instructions

Welcome to the SJSU Fasta ToolBox

Step 1 : Open a Fasta file you wish to work with

Step 2 : Then click Clustalw File Creator

Step 3 : The aligned sequences can be viewed with the Align File button

Step 4 : Distance Matrix can be viewed with Create Distance Matrix button

Step 5 : Click Draw Tree button and select dnd file created by Clustalw, creating a Tree

Step 6 : Enter sequences in text entry areas and do desired functions with either button

Sequences to Compare

Sequence 1 :

Sequence 2 :

Percent Identity DNA Sequence Manipulation

Program Buttons

Open

Clustalw File Creator

Align File

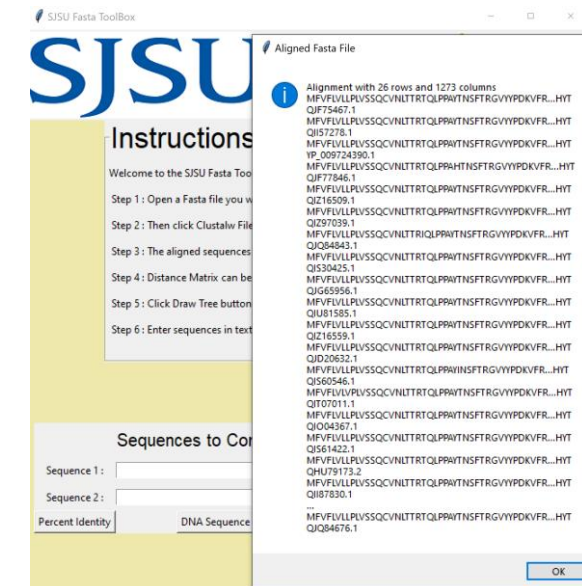
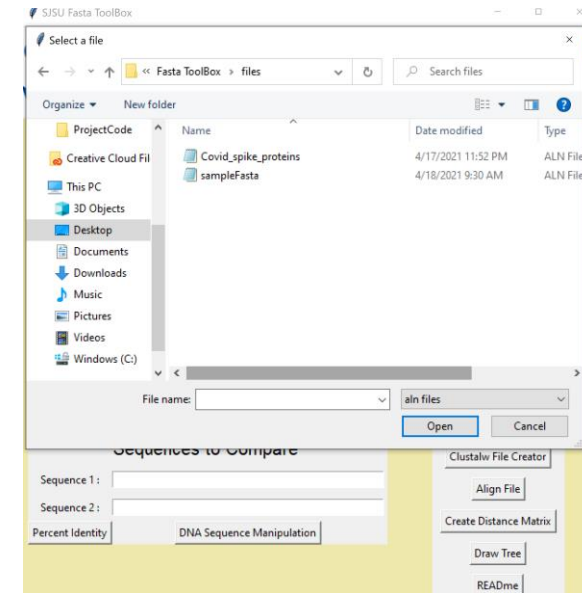
Create Distance Matrix

Draw Tree

READme

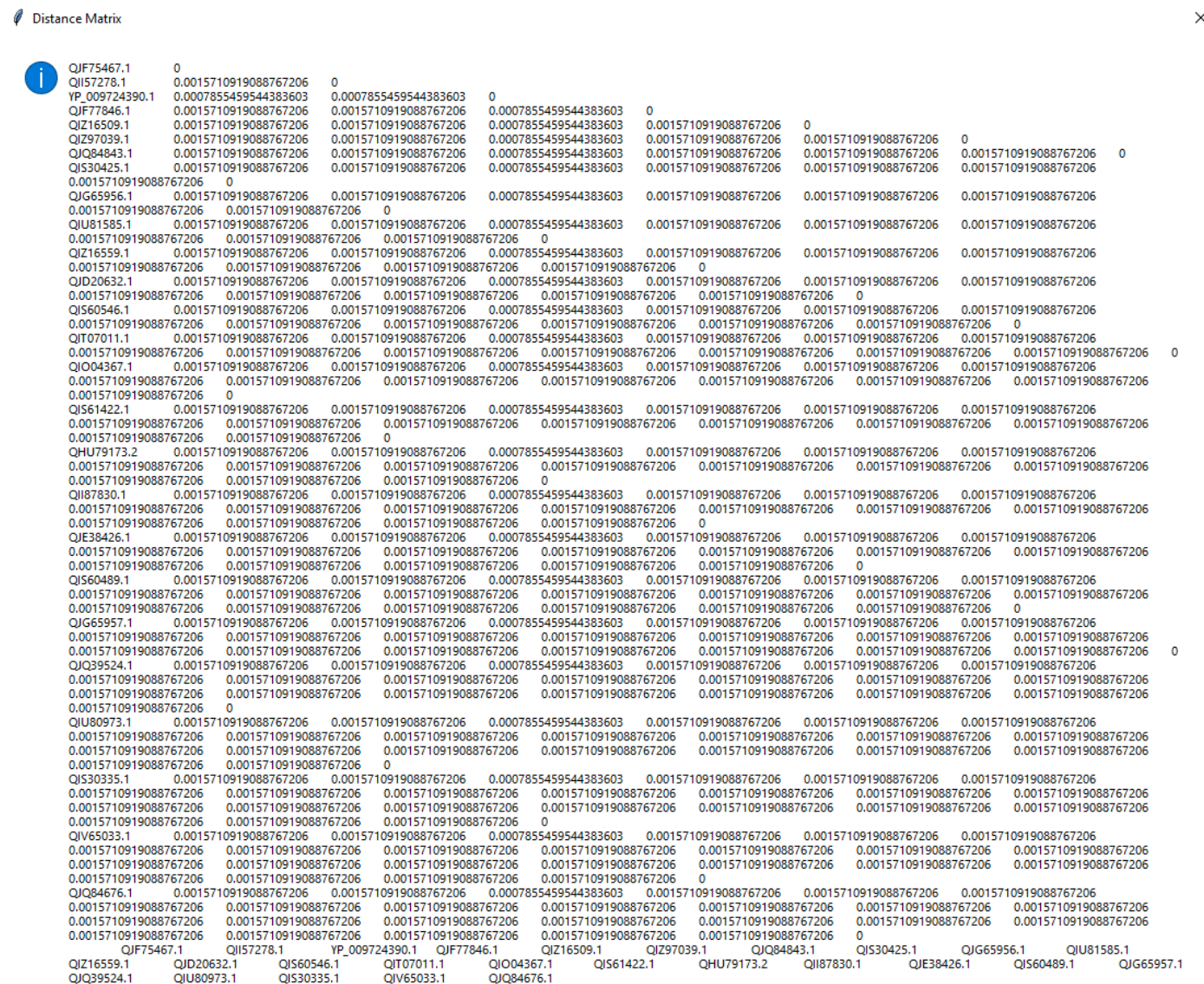
Align

- Align button will prompt user to select aln file created by Clustalw
- Once selected, the function will preview a list view of the alignments, if it is too large, whole alignment can be viewed with created aln file.



- The distance matrix button takes the aligned sequences and creates a distance matrix
- This distance matrix design comes from the biopython library

- The distance matrix button takes the aligned sequences and creates a distance matrix
- This distance matrix design comes from the biopython library



OK

Draw Tree

- The draw tree button prompts the user to open the dnd file created by Clustalw to view a tree representation of user's sequences from the file. Once file is selected a new window will appear
- User can view tree, zoom, and even save it to a file

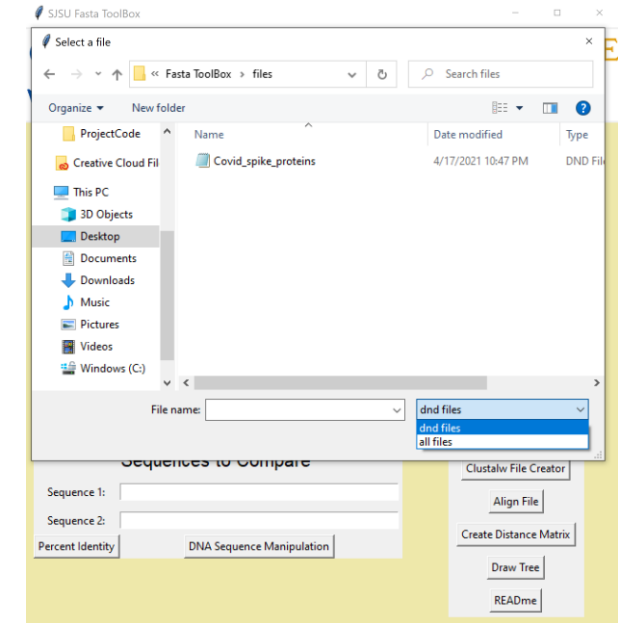
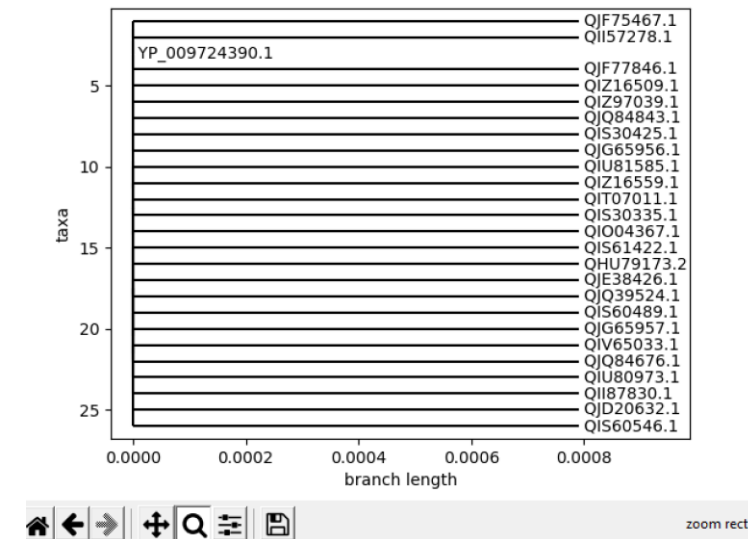
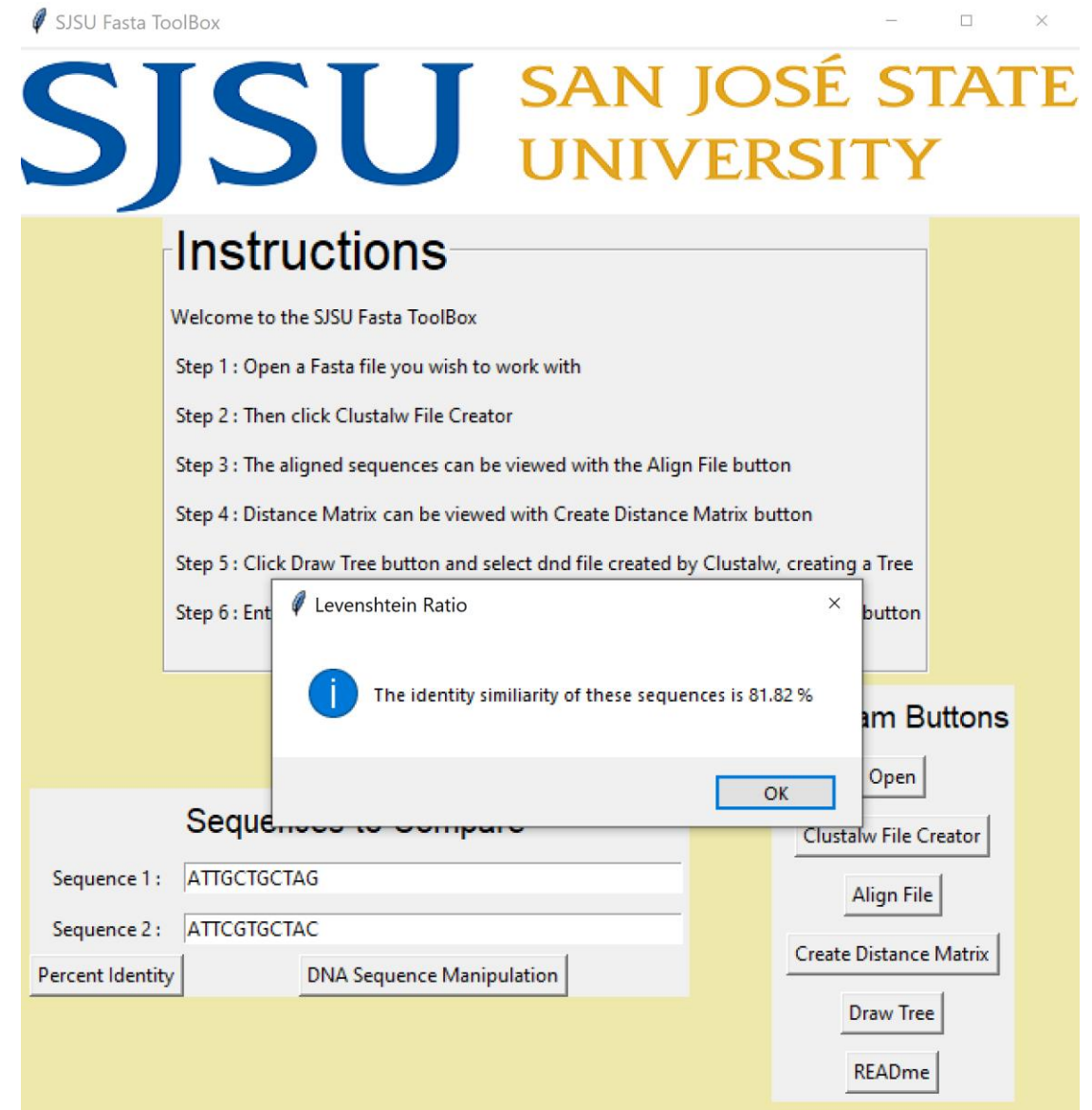


Figure 1



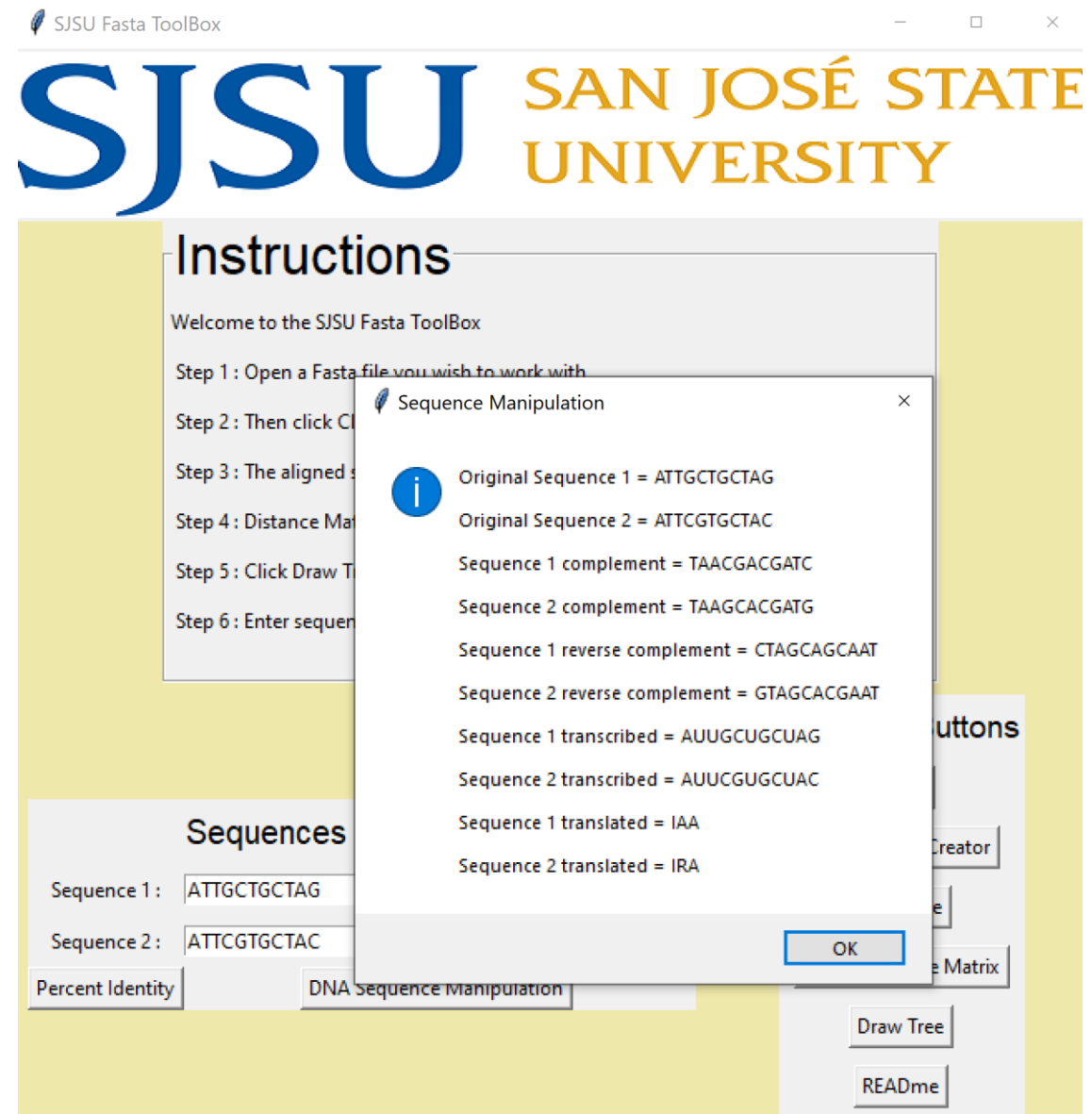
Percent Identity

- The 2 entry boxes are for users to enter 2 sequences
- Using these sequences, the percent identity can be calculated.
- This function uses my implementation of an algorithm designed by Vladimir Levenshtein, and it is used to measure difference between sequences. This is achieved by keeping track of the number of character edits needed (insertion, substitution, deletion) required to change one sequence into the other. Then a ratio is used to calculate a percent of how identical these sequences are.

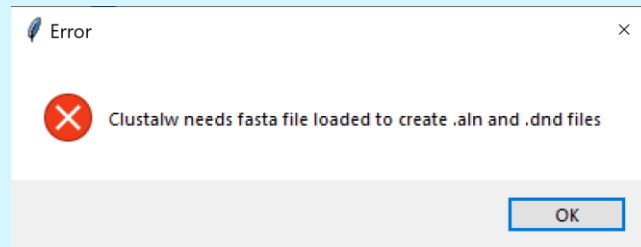
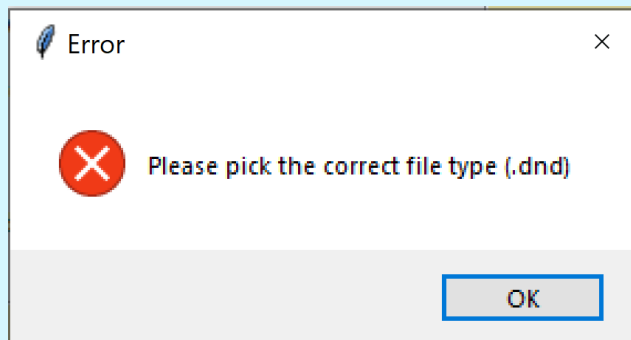
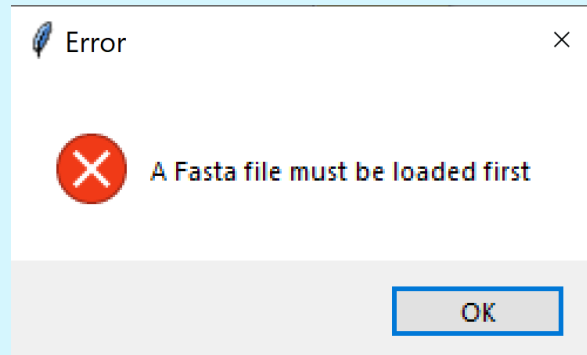
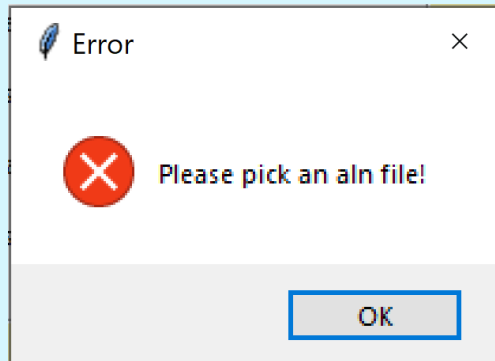


DNA Sequence Manipulation

- Using the same entry boxes, the DNA sequence manipulation button will take those sequences and manipulate them in 4 different ways
 - Complement
 - Reverse Complement
 - Transcription
 - Translation
- A new window will appear with the manipulated sequences displayed as a list view.



Error Catches



- Certain Error catches have been added to raise error to user so they can correct the issue.
- Most errors will be from not opening a file