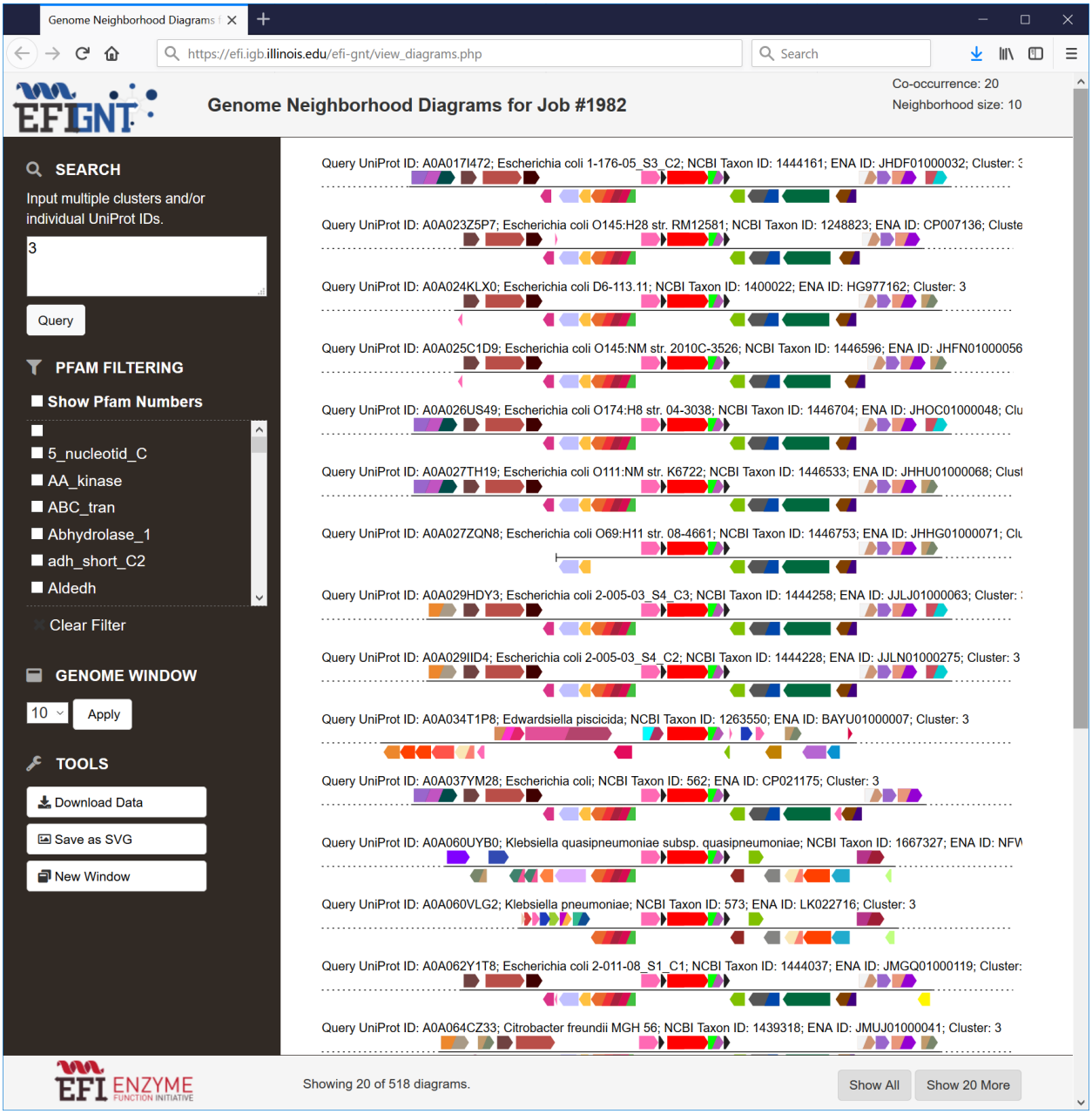
Genome Neighborhood Diagrams Tutorial



Genome Neighborhood Diagrams

The EFI-GNT tools provide the capability to view diagrams of genome neighborhoods for input SSNs or for a set of proteins obtained either through a list of IDs or a BLAST.

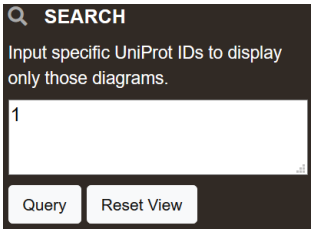
In a given diagram, the "query" ID—the protein that was input either through a list or corresponding to a node in an SSN—is located in the center of the diagram. It is always colored red, and is pointing in the "Forward" direction (5' to 3'). For query IDs that are on the reverse strand on the sequence information, the neighboring genes are flipped appropriately.

Each arrow is colored according to the Pfam that it belongs to. Some diagrams have multiple colors included on the same arrow  which represents a protein that is a fusion of multiple families. Each color in the arrow represents a specific Pfam in the fusion.

When the viewer is first loaded, the default behavior depends on if the diagrams were generated from a BLAST run or a list of protein IDs, or from an SSN through the GNN tool. In the latter case, the user must input a cluster number in the search box in order to view diagrams. The cluster numbers correspond do the clusters in the input SSN. If the diagrams were generated from BLAST or a list of IDs, then there is only one cluster in the entire dataset and that is displayed.

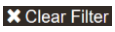
Loading and Filtering

Diagrams are displayed by putting cluster numbers, lists of independent query IDs, or a combination of these in the **Search** box. For example, if the user wishes to only view diagrams for 10 specific proteins, those protein IDs can be input in the search box. If multiple clusters numbers are input, the diagrams from the clusters are combined sequentially in the order that the cluster numbers are given.



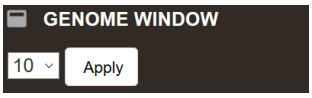
Diagrams can contain thousands of proteins, but only 20 diagrams are shown initially. The user can opt to add more to the view in increments of 20. By clicking the buttons at the bottom-right of the page

. The user can also request that the tool shows all diagrams; this can take several minutes depending on the cluster size, and the browser will present a warning indicating that the webpage or script has stopped responding. This error is safe to ignore.

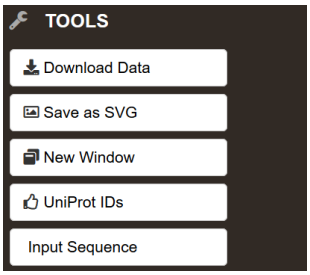
The viewer tool has the ability to highlight proteins containing specific Pfams, through the **Pfam Filtering** tool panel. All of the Pfams that are in the input proteins as well as the neighbors are listed in the Pfam Filtering box. Each of these can be clicked, and all proteins with the selected Pfams are highlighted by drawing a black border around the sequence, and by muting the color of all other proteins. All filters can be cleared by clicking the **Clear Filter** 

button.

When the diagrams are initially generated, the number of neighboring genes that are included in the diagrams is an input parameter and is called the genome window. This window cannot be retroactively increased, but it can be decreased through the **Genome Window** tool. This is useful for closely inspecting the immediate neighbors of the query protein. The number that is selected is the number that is applied on both the left and right of the query (e.g. 10 means 10 on either the left or right for a total of up to 20 proteins plus the query ID in the center). When the window is applied, the entire diagram is refreshed.



Tools



At the bottom of the left sidebar is the **Tools** section. The **Download Data** Download Data button downloads the original diagram database file so it can be viewed again in the future, or inspected through third-party tools. (This file is an SQLite database.) The **Save as SVG** Save as SVG button downloads the currently loaded diagram as an SVG diagram. The **New Window** New Window button creates a new viewer window with the current diagram. For jobs that originate from the *Retrieve Neighborhoods* → *Single Sequence BLAST*, *Retrieve Neighborhoods* → *Sequence ID Lookup*, and *Retrieve Neighborhoods* → *FASTA Sequence Lookup* GNT options, the **UniProt IDs** UniProt IDs button allows the user to view the UniProt IDs that were found in the BLAST or list of sequences. For the *Sequence ID Lookup*, and *FASTA Sequence Lookup* GNT options, there is a corresponding **Unmatched IDs** button that allows the user to view the IDs that were not in the database, or were not reverse matched if the input IDs were NCBI IDs. Both of these buttons display a dialog containing a list of IDs. The Input Sequence **Input Sequence** button is visible if the *Single Sequence BLAST* from the *Retrieve Neighborhoods* option is used to generate the diagrams, and contains the FASTA sequence used as input for the BLAST.

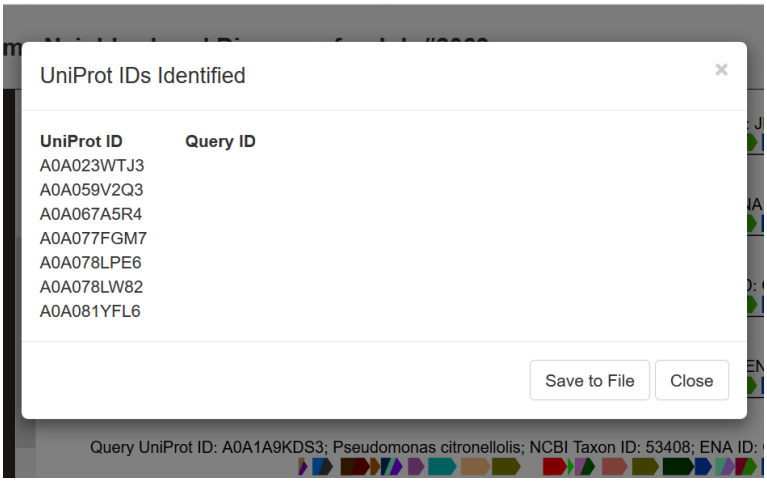


Figure 1: List of UniProt IDs

SVG Tutorial

The Neighborhood Diagram viewer tool ("Viewer") provides the user with the ability to download the currently-visible set of diagrams to an SVG format file. SVG—scalable vector graphics—is a type of data that describes the shapes in terms of relative coordinates rather than absolute positions (pixels) as is the case with a raster image. The SVG files downloaded from the Viewer contains arrows and diagrams that can be manipulated in an editor after the fact, without compromising the integrity of the image. This is useful for generating figure-quality images, or for generating an image with only a subset of diagrams in it through editing.

In the **Tools** section of the Viewer is the **Save as SVG** Save as SVG button. When this button is pressed, a file containing the currently received diagrams is made available for download. This is not just what's visible on the screen, but also whatever has been downloaded via the **Show All** or **Show 20 More** buttons at the bottom right. This tutorial will guide the user through opening and editing SVG files and exporting raster or PDF images for use in publications or presentations using the free Inkscape editor.

Inkscape

Inkscape is a free vector graphics editor that runs on Windows, Mac OS X, and Linux, and is available at https://inkscape.org/en/. In order to use it, it must be downloaded and installed, and instructions for doing this are available on the Inkscape website. The following steps assume that Inkscape is installed and is running.

Download a SVG file from the Viewer. In Inkscape, open the file through the *File → Open* menu option. You will see something similar to Figure 2. Each diagram is composed of a grouping of the various objects that constitute the diagram (the text, arrows, and lines). These groups can be deleted by clicking on a diagram and clicking the *Delete* key on the keyboard. Groups can also be shifted by using the mouse, or by selecting a diagram and using the keyboard arrow keys. In the case illustrated in Figure 2, we will remove the top five diagrams, and then retain the next 7 diagrams while removing all following diagrams. Each top diagram is selected and deleted using the mouse and the *Delete* key on the keyboard, and what is left is shown in Figure 3. Scrolling down the document, we select the 8th diagram and delete it. Next, we zoom out to see the whole document by pressing the *Ctrl* key and rolling the mouse wheel (or using the zoom box at the bottom right). Next, the rest of the diagrams are selected by clicking and dragging a box around the diagrams; the result is shown in Figure 4. These are deleted using the *Delete* key.

The next step is to reorder the diagrams and this is done by clicking and dragging the diagrams into the appropriate position. This is shown in Figure 5; notice that the diagrams aren't aligned correctly, which is remedied in the next step. Selecting the *Object → Align and Distribute* menu option shows a new panel on the right side of the Inkscape window. All of the diagrams are selected by pressing *Ctrl-A* on the keyboard, and then they are left-aligned by clicking the *Align left edges* button in the **Align and Distribute** panel. In order to make sure the diagrams are equally-spaced vertically, the diagrams are aligned using the **Distribute centers equidistantly vertically** button. These changes are illustrated in Figure 6. In more advanced usage that is outside the scope of this tutorial, lines can be drawn, text can be added or removed, and colors can be updated.

Inkscape allows export to a variety of file formats, but the most commonly used one is PNG. Other formats are PDF, EPS, EMF, and many others that Inkscape supports. A typical first step in the process of

exporting is to resize the document frame by selecting all of the diagrams (*Ctrl + A*) and then going to the *Edit → Resize Page to Selection* menu option. To export to PNG, go to *File → Export PNG Image*; this opens a window where export parameters can be edited. If the document was resized to fit the diagrams, the Export Area would typically be Page or Drawing. The Image Size Width and Height should remain the same typically, but the dpi settings could be changed depending on what the output image would be used for [presentations, web, email should use a lower dpi (72-100) whereas figures for publications should use 300 or 600 dpi]. These settings are illustrated in Figure 7.

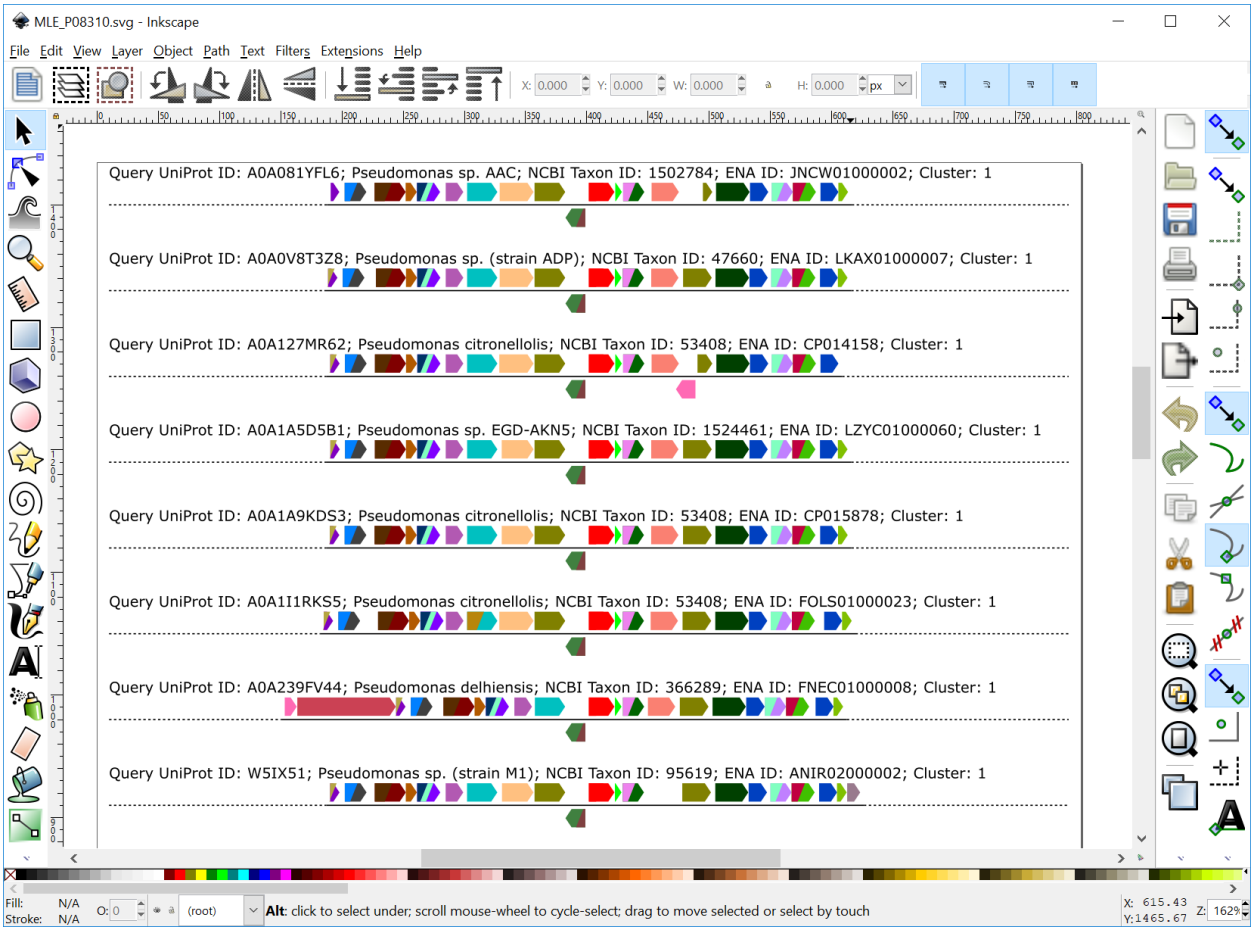


Figure 2: Initial Inkscape document view

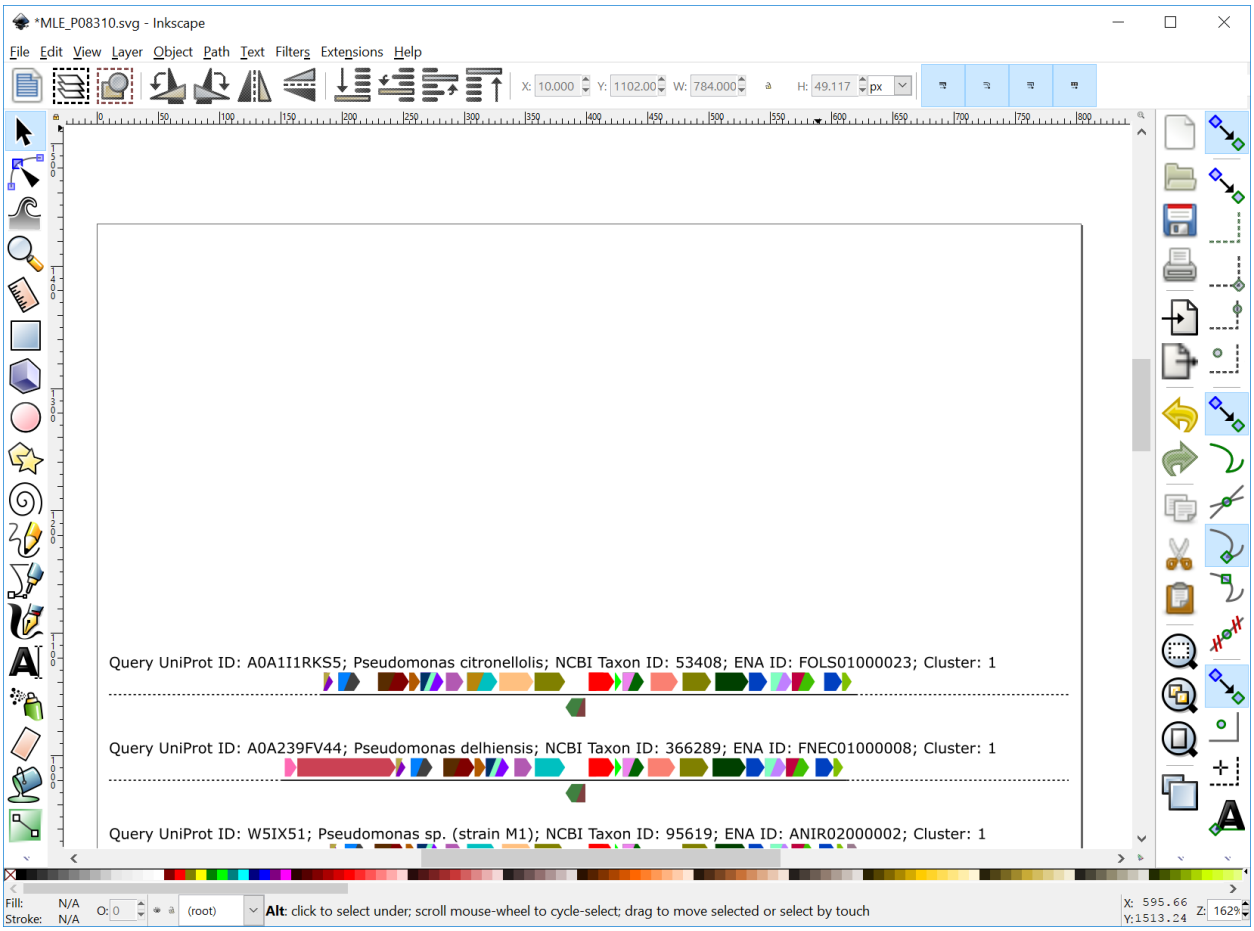


Figure 3: Document view after removal of some diagrams

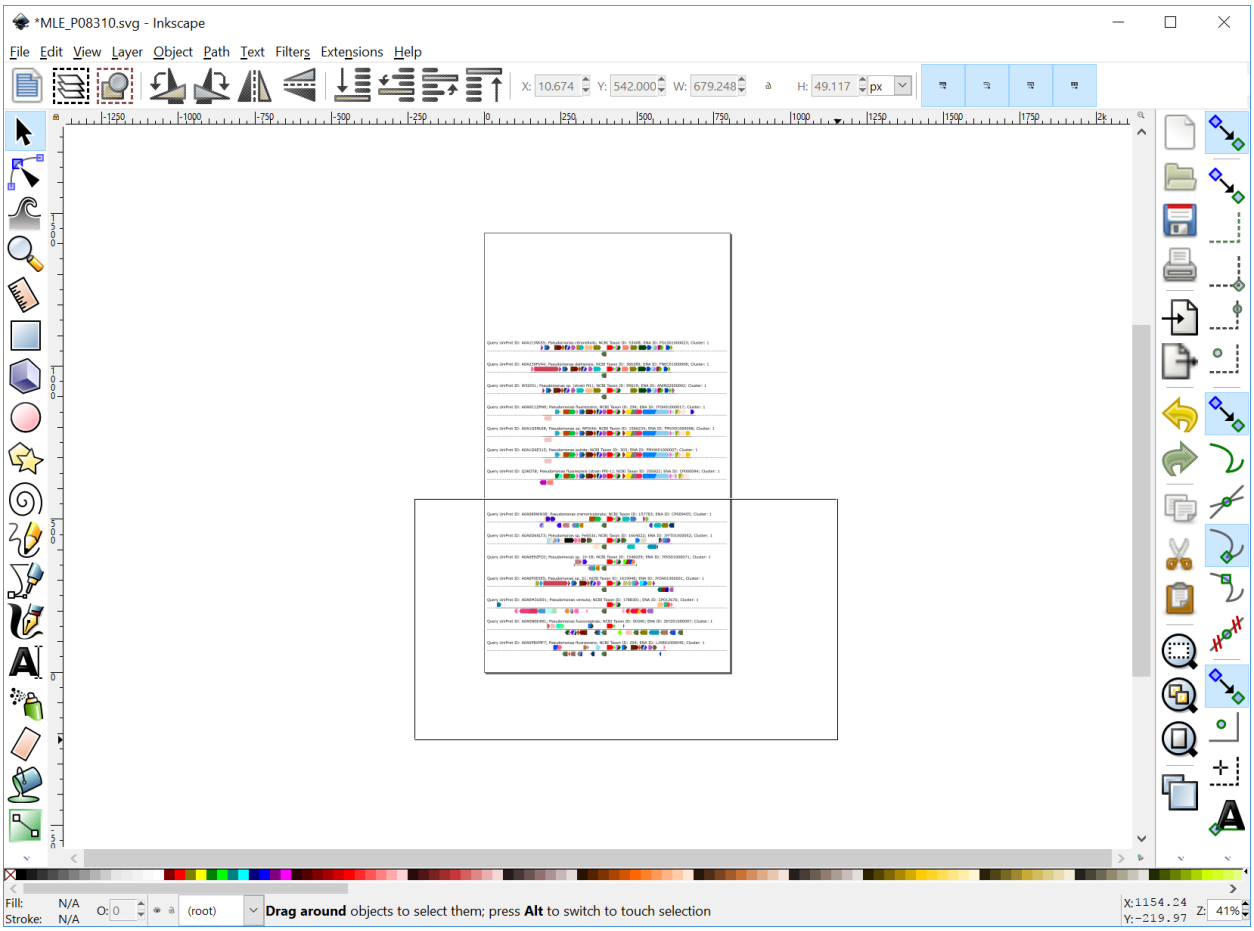


Figure 4: Document view after zooming out

Figure 5: Moving diagrams into desired position

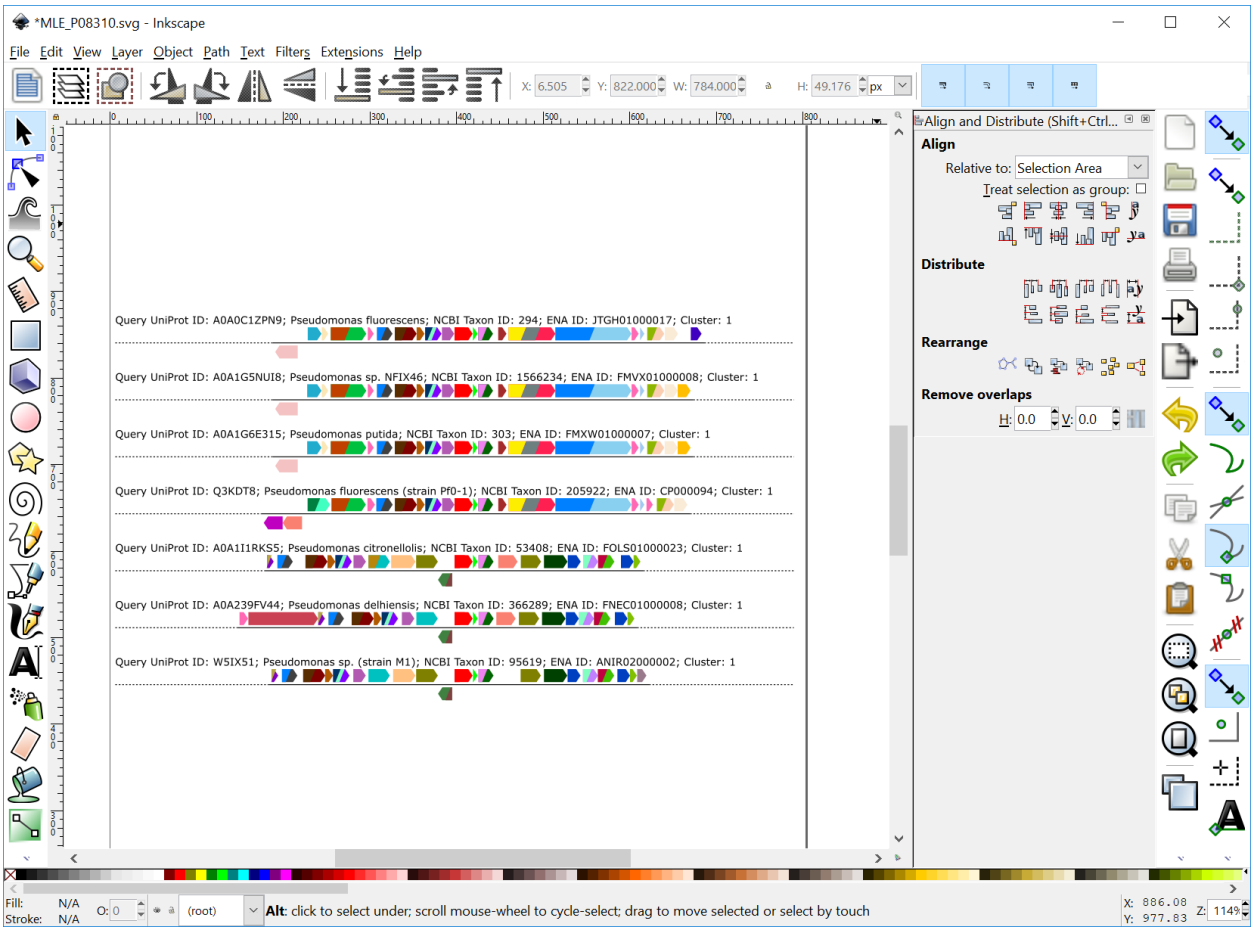


Figure 6: Document view after alignment

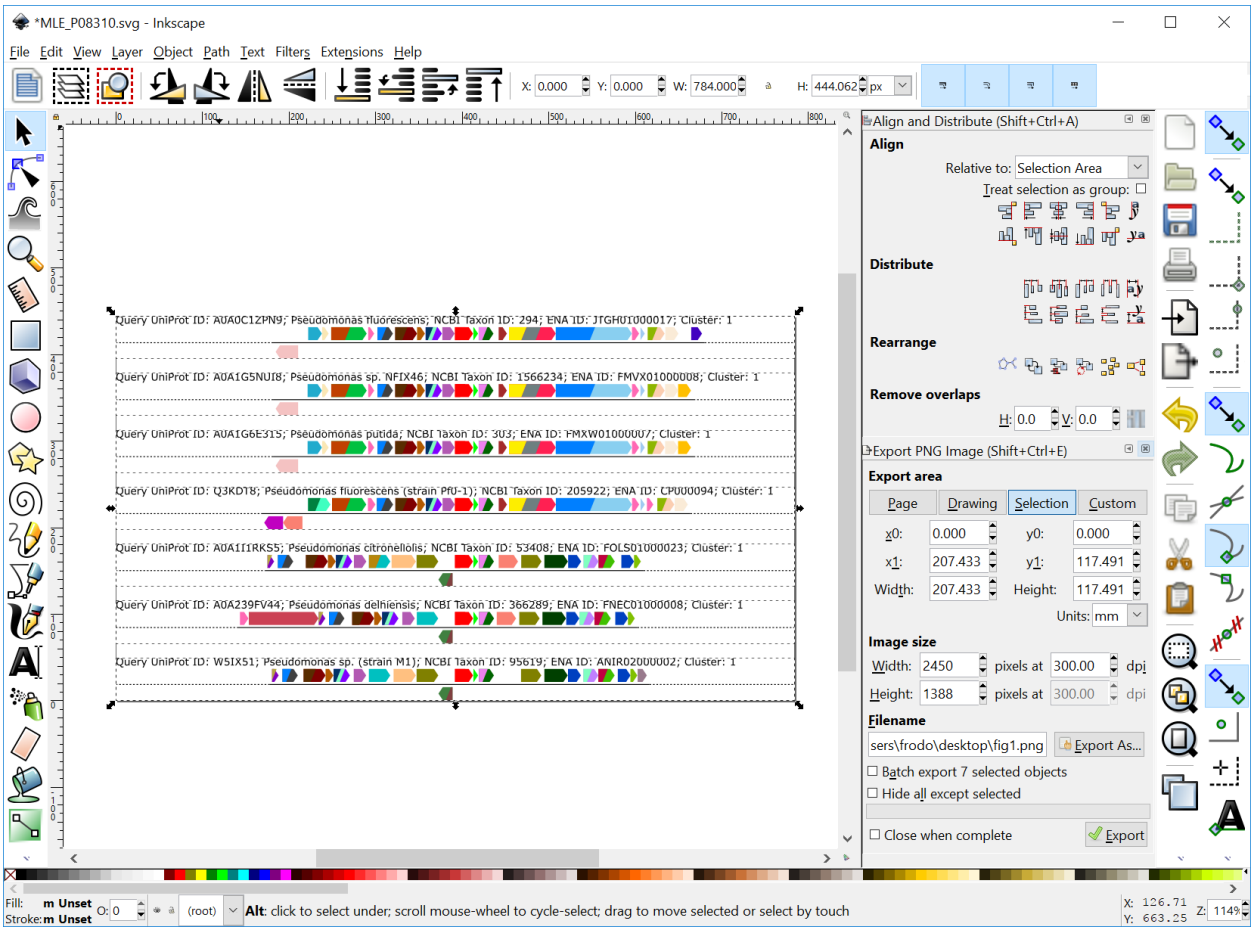


Figure 7: Exporting a PNG image