Multiple Sequence Alignment (MSA)

Option 1: Search for sequences and then align sequences

1. Identify sequences to align: mouse-over the “Search Data” tab and click “Nucleotide Sequences” or “Protein Sequences”. For this example, we will use nucleotide sequences.
2. Select search criteria on the Nucleotide Sequence Search page and click the “Search” button to run your query.
3. Select sequences from the search result page by clicking the checkboxes. Mouse-over the yellow “Run Analysis” button and click “Align Sequences (MSA)”. If you want to include sequences that are not in this search result or to use the sequences to do further analysis, select the desired sequences and click “Add to Working Set”. Then add other sequences to the same working set later by repeating the process.
4. A “Select Sequence Type” lightbox will pop-up. Select the appropriate sequence type and click “Continue”.
5. On the next page, select output format and output order. Then click “Run”.
6. If you have a large amount of input to align, the analysis may take a few minutes to run. While the analysis is running, you can choose to save it (upon completion) to your Workbench by entering a name for the analysis and then clicking the “Save to Workbench” button. Then you can move to other parts of the IRD site, and retrieve the alignment later from your Workbench.

IRD is funded by the National Institute of Allergy and Infectious Diseases (NIH/DHHS) under Contract No. HHSN266200400041C and is a collaboration between Northrop Grumman Health IT, University of Texas Southwestern Medical Center, Vecna Technologies, SAGE Analytica and Los Alamos National Laboratory. Comments, questions, suggestions? Contact us at feedback@virusbrc.org
Option 2: Align sequences from a working set or your own sequences

1. Mouse-over the “Analyze & Visualize” tab and click “Align Sequences (MSA)”.
2. On the MSA landing page, use one of the three options to input sequences:
   2.1 Upload a file containing sequences in FASTA format.
   2.2 Paste sequences in FASTA format.
   2.3 Use a working set from your Workbench. Select output format and order. Then click ”Run”.
3. If you have a large amount of input, the analysis may take a few minutes to run. While the analysis is running, you can choose to save it (upon completion) to your Workbench by entering a name for the analysis and then clicking the “Save to Workbench” button. Then you can move to other parts of the IRD site, and retrieve the alignment later from your Workbench.

Visualize and Customize the Alignment

1. After the alignment analysis is finished, mouse-over “Run Analysis” and click “Visualize Aligned Sequences”.
2. On the next page, you will have the option to customize the sequence labels by selecting the “Custom” radio button in the “Label sequence by” section. Click “Run” to load the alignment visualization window.
3. In the alignment visualization window, many customization options exist:
   - Rename a sequence label: Right-click a strain name in the alignment, mouse-over the sequence name in pop-up menu, click “Edit Name/Description”, modify the name and click “Accept”.
   - Highlight IRD-defined Sequence Features on the alignment:
   - Add your own sequence feature to the alignment: Click and drag a desired region of sequence alignment, right-click the selected region, mouse-over “Selection” and click “Create Sequence Feature”.
   - Color alignment based on sequence identity cutoff: Click the “Colour” pulldown menu and then the “Above Identity Threshold” option. Using sliding bar to adjust color display.
   - Manually adjust the alignment: Click a sequence and then use → on the keyboard to adjust the alignment.
4. Export the alignment from the “File” menu.

IRD is funded by the National Institute of Allergy and Infectious Diseases (NIAID) under Contract No. HHSN266200400041C and is a collaboration between Northrop Grumman Health IT, University of Texas Southwestern Medical Center, Vパーワナ Technologies, SAGE Analytica and Los Alamos National Laboratory. Comments, questions, suggestions? Contact us at feedback@virushrc.org.