Introduction to Bioinformatics online course: IBT

Multiple Sequence Alignment

Lec6: Selecting and editing sequences

By

Ahmed Mansour Alzohairy

Professor (Full) at Department of Genetics,
Zagazig University, Zagazig, Egypt
Selecting and editing sequences
Selecting parts of an alignment
Selections can be of arbitrary regions in an alignment, one or more complete columns, or one or more complete sequences.
A selected region can be copied and pasted as a new alignment using the:

*Edit ⇒ Copy and Edit ⇒ Paste ⇒ To New Alignment* in the alignment window menu options.
To clear (Deselect) the selection press the [ESC] (Escape) key.
Selecting arbitrary regions
To select part of an alignment, place the mouse at the top left corner of the region you wish to select. Press and hold the mouse button and drag the mouse to the bottom right corner of the chosen region then release the mouse button. A dashed red box appears around the selected region.

Selecting a region in an alignment
Selecting columns

- To select the same residues in all sequences, click and drag along the alignment ruler. This selects the entire height of the alignment.
- Ranges of positions can also be selected by clicking on the first position then holding down the [SHIFT] key whilst clicking the other end of the selection.
- Discontinuous regions can be selected by holding down [CTRL] and clicking on positions to add to the column selection. Note that each [CTRL] Selected columns are indicated by red highlighting in the ruler bar.
Selecting sequences

To select **multiple complete sequences**, click and drag the mouse **down** the sequence ID panel. The same technique as used for columns above can be used with **[SHIFT]-Click** for **continuous** and **[CTRL]-Click** to select **discontinuous** ranges of sequences.
Making selections in Cursor mode
To define a selection in cursor mode (which is enabled by pressing [F2] when the alignment window is selected), navigate to the top left corner of the proposed selection (using the mouse, the arrow keys, or the keystroke commands described). Pressing the [Q] key marks this as the corner. A red outline appears around the cursor. Navigate to the bottom right corner of the proposed selection and press the [M] key. This marks the bottom right corner of the selection. The selection can then be treated in the same way as if it had been created in normal mode.
Inverting the current selection

The current sequence or column selection can be inverted, using **Select ⇒ Invert Sequence/Column Selection** in the alignment window.

Inverting the selection is useful when selecting large regions in an alignment, simply select the region that is to be kept unselected, and then invert the selection.
This may also be useful when hiding large regions in an alignment. Instead of selecting the columns and rows that are to be hidden, simply select the region that is to be kept visible, invert the selection, then select **View ⇒ Hide ⇒ Selected Region**.
Creating groups

Selections are lost as soon as a different region is selected. Groups can be created which are labeled regions of the alignment. To create a group,

- first select the region which is to comprise the group.
- Then click the right mouse button on the selection to bring up a context menu.

Select **Selection ⇒ Group ⇒ Edit**

- name and description of current group
- then enter a name for the group in the dialogue box which appears. By default the new group will have a box drawn around it. The appearance of the group can be changed.

This group will stay defined even when the selection is removed.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>FER_CAPA/1-97</td>
<td>ALFGLKS-A-NG4KVTCMA</td>
</tr>
<tr>
<td>FER_CAPA/1-144</td>
<td>AVTSL-KPIPNVGE-ALFGLKS-G-</td>
</tr>
<tr>
<td>FER1_SOLCO/1-144</td>
<td>AVTSL-KAISNVGE-ALFGLKS-G-</td>
</tr>
<tr>
<td>Q93007/1-144</td>
<td>VVTSL-KAISNVGE-ALFGLKS-G-</td>
</tr>
<tr>
<td>FER1_PEA/1-149</td>
<td>MPMSV-TTTKAFSN-9FLGGLKT-SIRG1LAMAVASLYKVLTVPDGQEHGVIDYI2DEEVLHFPYSCRAGS</td>
</tr>
<tr>
<td>Q7X995_1-152</td>
<td>VPMYSV-ATTTTKAIPPSGFLKSVSIRG1LAMAVASLYKVLTVPDGQEHGVIDYI2DEEVLHFPYSCRAGS</td>
</tr>
<tr>
<td>FER1_MESUR/1-148</td>
<td>TPPMTAALFTNVR-ALFGLKS-SASR-6RVTAMAYKVLTVPDGQEHGVIDYI2DEEVLHFPYSCRAGS</td>
</tr>
<tr>
<td>FER1_SPILO/1-147</td>
<td>APPMMAALFSNTOR-SLFQGLKT-GSR-6RMTMAAYKVLTVPDGQEHGVIDYI2DEEVLHFPYSCRAGS</td>
</tr>
<tr>
<td>FER1_RAPSA/1-96</td>
<td>ATYKVKF1TTPSCEEVDCIDWYVLDAEEVHLSCAGS</td>
</tr>
<tr>
<td>FER1_ARAT7/1-149</td>
<td>APLSLRSLPSANTQ-SLFGLKS-GTARG5RVTAMAYKVIKTTPSCEEVDCIDWYVLDAEEVHLSCAGS</td>
</tr>
<tr>
<td>FER1_BIANW/1-96</td>
<td>ATYKVKFI1TTPSCEEVDCIDWYVLDAEEVHLSCAGS</td>
</tr>
<tr>
<td>FER2_ARAT7/1-149</td>
<td>TPSLRSLPFAN-</td>
</tr>
<tr>
<td>Q93300_ARAT7/1-116</td>
<td>TPSLRSLPFAN-</td>
</tr>
<tr>
<td>FER1_MAIZE/1-150</td>
<td>APTAV-ALPAA</td>
</tr>
<tr>
<td>O08429_MAIZE/1-145</td>
<td>APPPCSSFLRLR-V-AVAKPLA</td>
</tr>
</tbody>
</table>

**Secondary Structure**

**Iron Sulphur Contacts**

**Conservation**

**Quality**
Exporting the current selection
- The **current selection** can be copied to the **clipboard** (in PFAM format).
- It can also be **output to a textbox** using the output functions in the pop-up menu obtained by right clicking the current selection.
- The **textbox enables** quick manual editing of the alignment prior to importing it into a new window (using the [New Window] button) or saving to a file with the **File ⇒ Save As** pulldown menu option from the text box.
AACon Protein Alignment Conservation Services
Practical

“To try on your own time”
Exercise: Making selections and groups

1- Close all windows in Jalview and load the ferredoxin alignment (PFAM ID PF03460).

2- Choose a residue and place the mouse cursor on it. Click and drag the mouse cursor to create a selection. As you drag, a red box will ‘rubber band’ out to show the extent of the selection. Release the mouse button and a red box should border the selected region.

3- Now press [ESC] to clear the selection.
   - Select one sequence by clicking on the ID panel. Note that the sequence ID takes on a highlighted background and a red box appears around the selected sequence.
   - Now hold down [SHIFT] and click another sequence ID a few positions above or below. Note how the selection expands to include all the sequences between the two positions on which you clicked.
   - Now hold down [CTRL] and click on several sequences ID’s both selected and unselected. Note how unselected IDs are individually added to the selection and previously selected IDs are individually deselected.

4- Repeat the step above but selecting columns by clicking on the ruler bar instead of selecting rows by clicking on the sequence ID.

5- Press [F2] to enter Cursor mode. Navigate to column 59, row 1 by pressing 5 9 , 1 [RETURN].
- Press Q to mark this position. Now navigate to column 65, row 8 by pressing 6 5 , 8 [RETURN]. Press M to complete the selection.
- Open the popup menu by right-clicking the selected region with the mouse. Open the *Selection ⇒ Group ⇒ Group Colour* menu and select ‘Percentage Identity’.
- This will turn the selected region into a group and color it accordingly.
- Hold down [CTRL] and use the mouse to select and deselect sequences by clicking on their Sequence ID label. Note how the group expands to include newly selected sequences, and the ‘Percentage Identity’ coloring changes.
- Use the mouse to click and drag the right-hand edge of the selected group. Note again how the group resizes.
Right click on the text area to open the selection popup-menu. Follow the menus and pick an output format from the *Selection ⇒ Output to Textbox* . . . submenu. Try manually editing the alignment and then press the [New Window] button to import the file into a new alignment window.
Introduction to Bioinformatics Online Course: IBT

Multiple Sequence Alignment | Prof. Ahmed M. Alzohairy
By
Ahmed Mansour Alzohairy
Department of Genetics, Zagazig University, Zagazig, Egypt