BIONUMERICS Tutorial:

Sequence alignment and mutation analysis

1 Aim

The *Sequence alignment* window in BIONUMERICS has been designed for the calculation of multiple sequence alignments, subsequence searches and mutation analysis. Curves can be displayed for the trace files, allowing a quick and reliable evaluation of the correctness of positions of interest. In this tutorial you will learn how to create and calculate an alignment project in BIONUMERICS and how to perform a mutation search.

2 Preparing the database

1. Create a new database (see tutorial "Creating a new database") or open an existing database.

2. Import the example .SCF trace files as described in the tutorial: "Importing and assembling sequences in batch". The trace files originate from influenza A virus strains and represent partial sequences of the haemagglutinin (HA) and neuraminidase (NA) genes.

3 Creating a new alignment project

In the *Main* window, the *Alignments* panel is displayed in default configuration in the lower right corner.

1. Highlight the *Database entries* panel and select all entries using the shortcut **Ctrl+A**.

2. To create a new alignment project, highlight the *Alignments* panel in the *Main* window and select ***Edit*** > ***Create new object...*** (+).

A name for the new alignment project is prompted for (see Figure 1).

3. Specify a name and press <***OK***>.

The new alignment project is added to the *Alignments* panel in the *Main* window and the *Experiment types* dialog box opens. The *Experiment types* dialog box displays a list of available sequence types and the number of associated entries. From this list, the user can select the experiment type(s) that should be included in the alignment project.

4. Leave the experiment types **HA** and **NA** selected in the list and press <***OK***>.

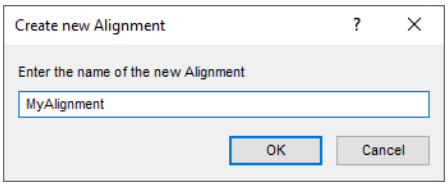

Figure 1: Specify an alignment project name.

4 Aligning sequences

In case multiple sequence types were imported for the selected entries, the active (i.e. the currently displayed) sequence alignment can be selected from the drop-down list in the main toolbar: HA.

1. Make sure **HA** is selected from the drop-down list.

2. Select ***Alignment*** > ***Calculate*** > ***Multiple alignment...*** ( ).

The settings for the successive pairwise and multiple alignment steps are grouped within this single dialog box.

3. Leave all parameters at their default value and press <***OK***> to calculate a multiple alignment using the BIONUMERICS algorithm.

The dendrogram and similarity matrix are still based on the pairwise similarity values. To calculate a global cluster analysis:

4. Choose ***Clustering*** > ***Calculate*** > ***Clustering (multiple alignment)...*** ( ) to call the *Clustering settings (multiple alignment)* dialog box.

5. Select **Neighbor Joining** and press <***OK***> (see Figure 2).

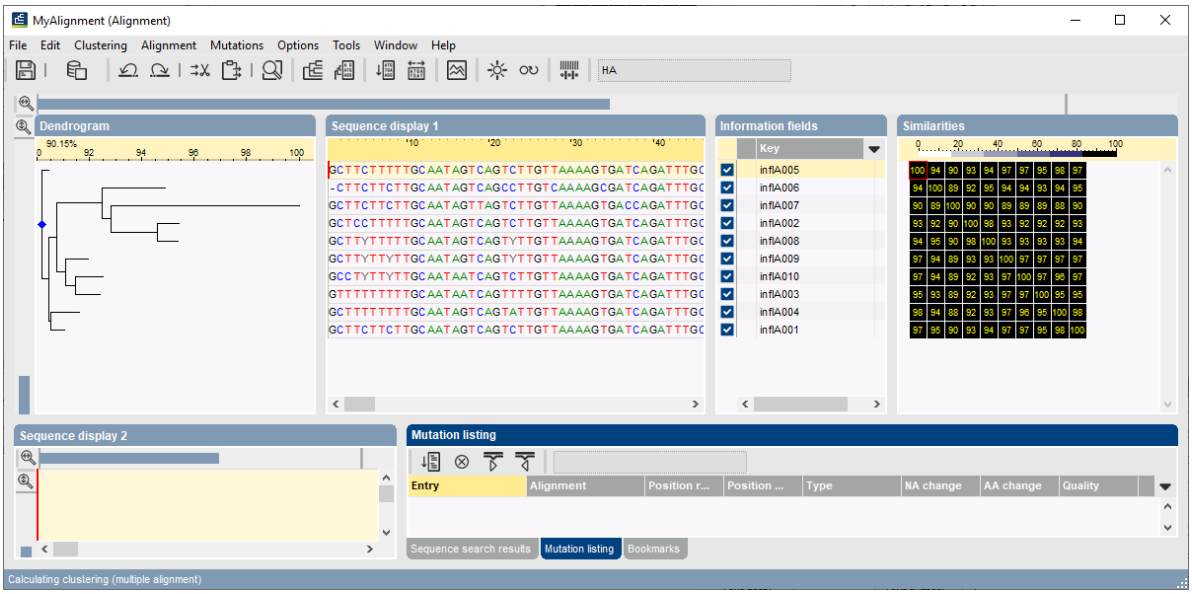

Figure 2: Neighbor joining tree.

5 Calculating a consensus sequence

A consensus sequence can be used to obtain a multiple alignment of all sequences against one single sequence. Depending on the type of analysis, the user may wish to assign a single sequence as consensus or may want to calculate the consensus sequence based on several sequences. A consensus sequence is also required for mutation searches and allows additional identity display settings for the alignment to be chosen.

To allow maximum flexibility, a consensus sequence is always calculated for the currently selected entries in an alignment.

1. Select all entries in the alignment (**Ctrl+A**).
2. Select ***Alignment*** > ***Consensus*** > ***Create from selected entries*** () (see Figure 3).

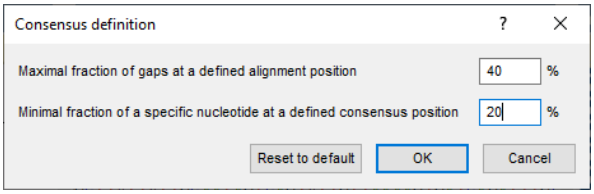

Figure 3: The *Consensus definition* dialog box.

3. Leave the setting for gaps at the default value but enter 20% for ***Minimal fraction of a specific nucleotide at a defined consensus position***.
4. Press <***OK***> to calculate a consensus based on the selected sequences.

The consensus is displayed in the header of the *Sequence display 1* panel (see Figure 4).

6 Display options

1. Make sure **HA** is selected from the drop-down list in the main toolbar.
2. To load the curves into the alignment project, choose ***Alignment*** > ***Load curves*** ().

The *Sequence display 2* panel now shows the curves for the experiment type **HA** (see Figure 4).

A number of options are designed to enhance the visualization of conserved parts in the alignment. They are specific to the *Sequence display 1* panel, and are grouped in the menu item ***Alignment*** > ***Identity display***.

3. Select ***Alignment*** > ***Identity display*** > ***Conserved blocks*** to display the sequence positions that are conserved throughout the alignment in gray.
4. Select ***Alignment*** > ***Identity display*** > ***Identity with consensus*** to display the sequence positions that are the same as in the consensus sequence in gray (see Figure 5).

7 Sequence translation

BIONUMERICS can automatically translate an alignment of nucleotide sequences into amino acids according to a selected translation table and within a certain translation frame. The trans-

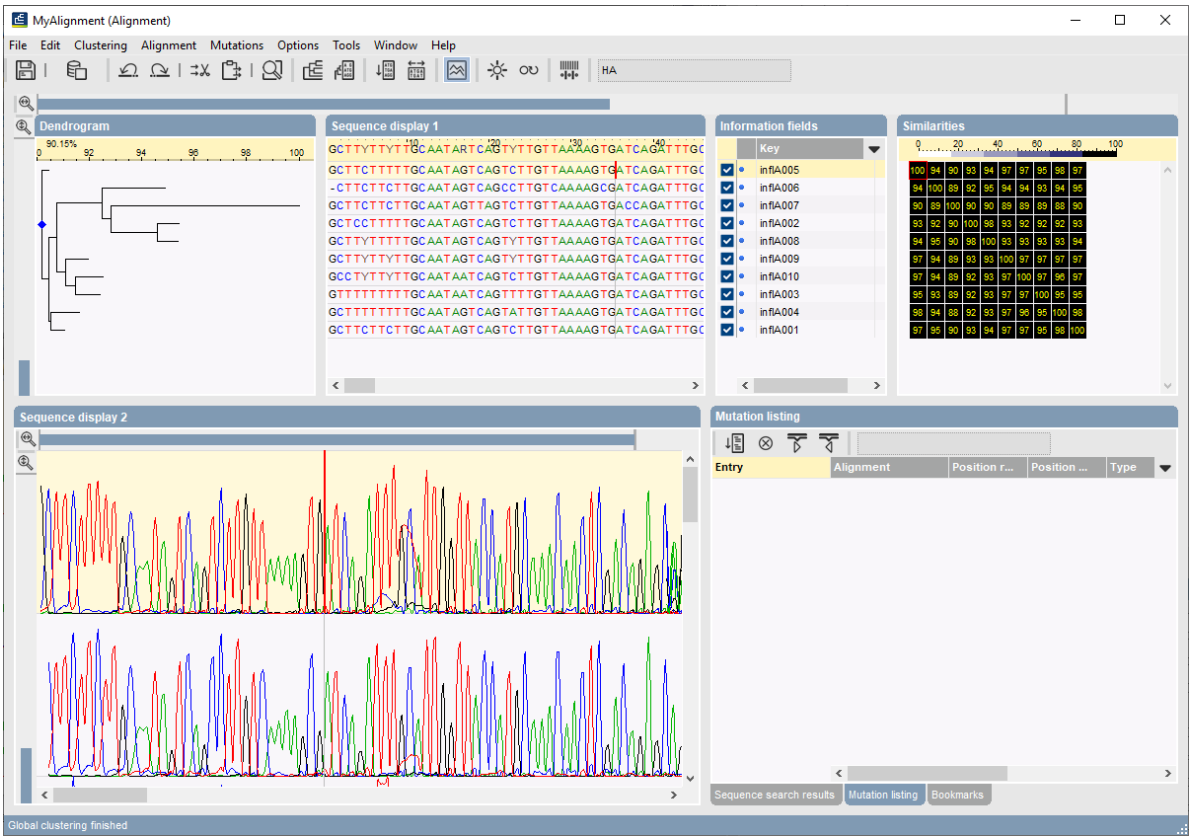

Figure 4: Curves loaded into the alignment project.

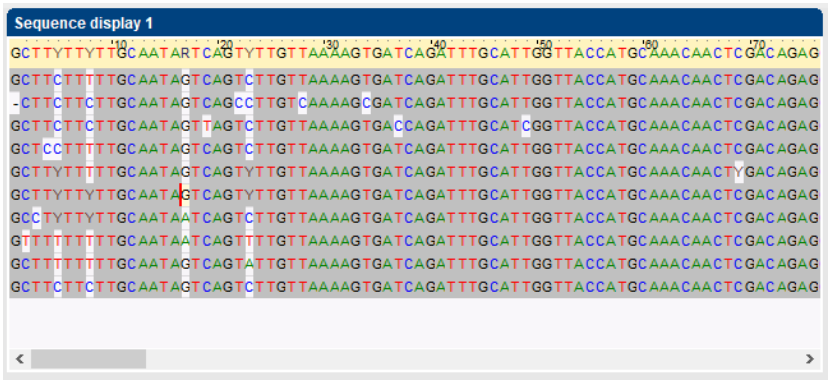

Figure 5: Identity with consensus.

lated amino acid sequence is displayed in the sequence alignment.

1. Select ***Alignment*** > ***Translation*** > ***Define...*** to call the *Translation settings* dialog box (see Figure 6).

2. For the **HA** experiment type select ***Frame 1*** and press <***OK***> to accept the settings.

3. Select ***Alignment*** > ***Translation*** > ***Show/Hide*** ( ) to display the translation (see Figure 8).

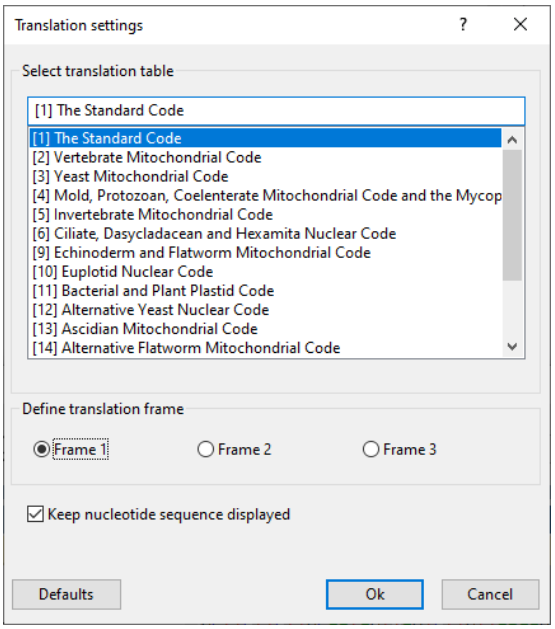

Figure 6: The *Translation settings* dialog box.

8 Mutation search

The mutation search tool is designed to detect mutations in individual sequences based on comparison with a consensus. This consensus can be derived from a single sequence or a set of sequences. Therefore, in order to perform a mutation search, a consensus sequence should first be calculated.

1. Make sure **HA** is selected from the drop-down list in the main toolbar and make sure a consensus sequence is present.

2. Select ***Mutations*** > ***Search...*** ( ).

This calls the *Find mutations* dialog box (see Figure 7).

3. Leave all settings at their default and press <***Find***> to start the mutation search.

The results are displayed in the *Mutation listing* panel (see Figure 8).

4. Click on any of the mutations listed in the *Mutation listing* panel.

The cursor will jump to the corresponding position on the alignment (in the *Sequence display 1* panel) and to the corresponding position on the curves (in the *Sequence display 2* panel, if displayed).

5. Select ***File*** > ***Save project*** (, **Ctrl+S**) to save the alignment project.

When an alignment project is saved, all calculations done on the sequences it contains will be stored along.

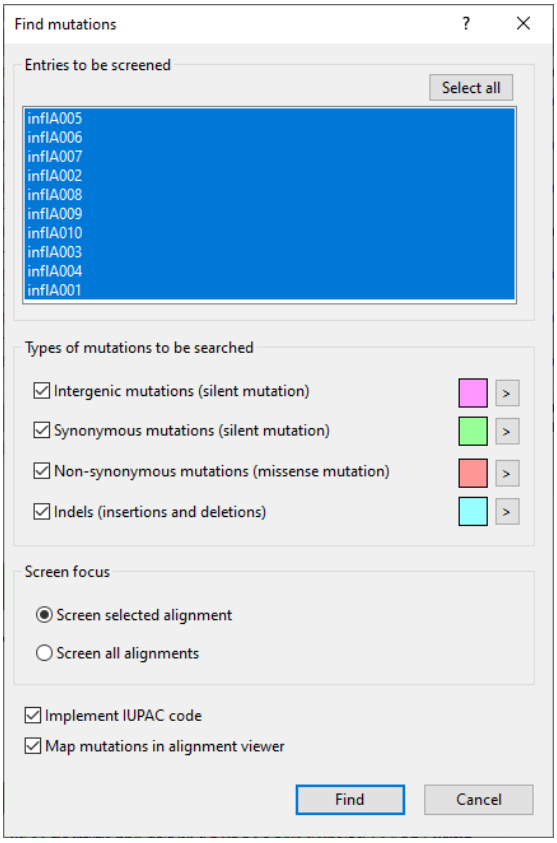

Figure 7: Mutation settings.

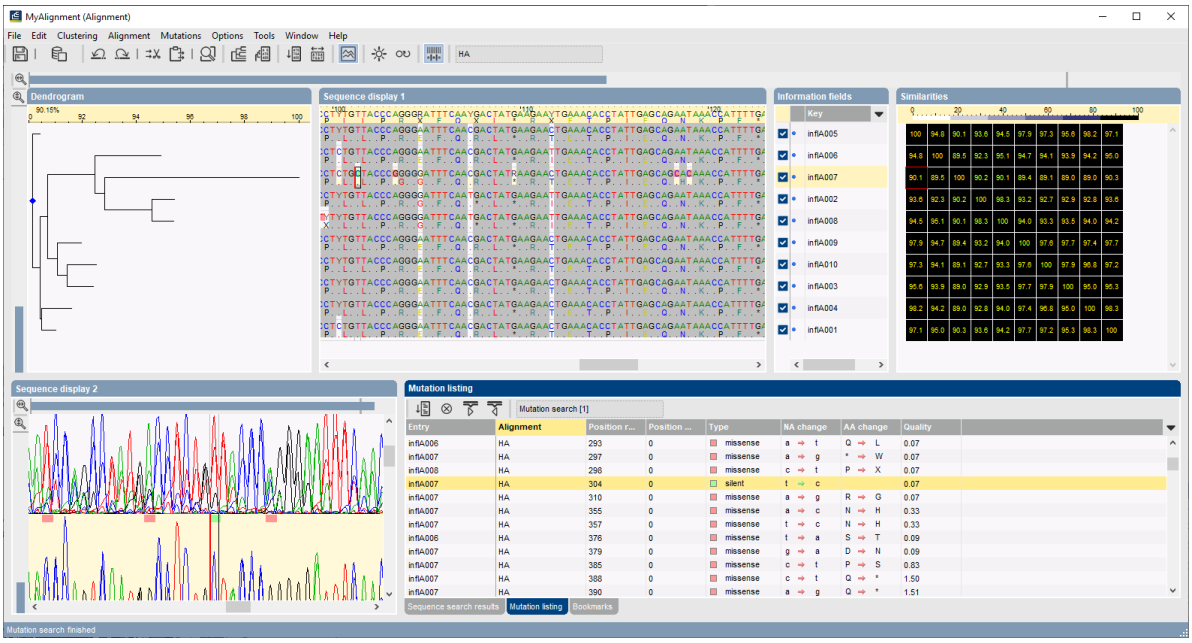

Figure 8: Mutation listing.