

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 BIONU-MERICS 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** >.

Create new Alignment	?	×
Enter the name of the new Alignment		
MyAlignment		
ОК	Car	icel

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:

- 1. Make sure **HA** is selected from the drop-down list.
- 2. Select Alignment > Calculate > Multiple alignment... (iii).

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)...* (ﷺ) to call the *Clustering settings (multiple alignment)* dialog box.

MyAlignment (Alignment)		-		×
File Edit Clustering Alignment Mutations Options	Tools Window Help			
B € ∽ ∵ I;× C; Q @				
(Q)		1		
Dendrogram	Sequence display 1	1		
90.15% 92 94 98 98 100	10 120 130 140 Key 🗸 0 20 40 60	80	100	
F	SCTTCTTTTTGCAATAGTCAGTCTTGTTAAAAGTGATCAGATTTGC 🗹 infla005 100 94 90 93 94 97 97 95	98 97		^
	-CTTCTTCTTGCAATAGTCAGCCTTGTCAAAAGCGATCAGATTTGC 🗹 inflA006 94 100 89 92 95 94 94 93	94 95		
	GCTTCTTCTTGCAATAGTTAGTCTTGTTAAAAGTGACCAGATTTGC	88 90		
	GCTCCTTTTTGCAATAGTCAGTCTTGTTAAAAGTGATCAGATTTGC 2 33 32 90 100 98 93 92 92	92 93		
		93 94		
		98 97		
	GTTTTTTTGCAATAATCAGTTTTGTTAAAAGTGATCAGATTTGC	95 95		
	GCTTTTTTTTGCAATAGTCAGTATTGTTAAAAGTGATCAGATTTGC 🗹 inflA004 58 94 88 92 93 97 99 95 0	00 98		
Ľ	GCTTCTTCTTGCAATAGTCAGTCTTGTTAAAAGTGATCAGATTTGC V infA001 97 95 90 93 94 97 97 95	98 100		
				\sim
Sequence display 2	Mutation listing			
(A)				
()		_	_	
	Entry Alignment Position r Position Type NA change AA change	Quality		
				^
	v			~
<	Sequence search results Mutation listing Bookmarks			
		_		

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* ($\frac{1}{2}$) (see Figure 3).

Consensus definition	? ×	
Maximal fraction of gaps at a defined alignment position	40 %	
Minimal fraction of a specific nucleotide at a defined consensus position	20 %	
Reset to default OK	Cancel	

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.

Sequence display 1
scttyttytt <mark>ic</mark> caatartca28tyttgttaa32agtgatcag2tttgcattg8ttaccatgc22aacaactcg22cag4
BCTTCTTTTTGCAATAG <mark>TCAGTCTTGTTAAAAGTGATCAGATTTGCATTGGTTACCATG</mark> CAAACAACTCGACAGA
- CTTCTTCTTGCAATAGTCAGCCTTGTCAAAAGCGATCAGATTTGCATTGGTTACCATGCAAACAACTCGACAGA
GCTTCTTCTTGCAATAGTTAGTCTTGTTAAAAGTGACCAGATTTGCATCGGTTACCATGCAAACAACTCGACAGA
BCT CCTTTTTGC AAT AG <mark>T CAGT CTTGT TAAAAG TGA TCAGAT TTG CAT TGG TTA CC</mark> ATGC AAA CAA CTC GAC AGA
3CTTYTTTTTGCAATAGTCAGTYTTGTTAAAAGTGATCAGATTTGCATTGGTTACCATGCAAACAACTYGACAGA
BCTTYTTYTTGCAATA <mark>GTCAGTYTTGTTAAAAGTGATCAGATTTGCATTGGTTACCATGCAAACAACTCGACAG</mark> A
BCC TYTTYTTGC AATAAT CAGTCTTGTTAAAAG TGATCAGATTTGCATTGGTTA CCATGC AAACAACTC GAC AGA
GTTTTTTTTTGC AAT AA <mark>T CAGTTTTGTTAAAAG TGA TCAGATTTGCATTGGTTA CCATGCAAA CAACTCGACAG</mark> A
3CTTTTTTTTGC AAT AG <mark>T CAGTAT TGT TAAAAG TGA TCAGAT TTGCAT TGG TTA CCA TGC AAA CAACTC GAC AG</mark> A
GCTTCTTCTTGCAATAG <mark>TCAGTCTTGTTAAAAGTGATCAGATTTGCATTGGTTACCATGCAAACAACTCGACAG</mark> A
<

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* ($\frac{11}{110}$) to display the translation (see Figure 8).

Select translation ta	ble			
[1] The Standard	Code			
(1) The Standard ([2] Vertebrate Mit [3] Yeast Mitocho [4] Mold, Protozo [5] Invertebrate M [6] Ciliate, Dasyclc [9] Echinoderm an [10] Euplotid Nuc [11] Bacterial and [12] Alternative Ye [13] Ascidian Mitt [14] Alternative Fl	ochondrial Code ndrial Code an, Coelenterate Mitoch litochondrial Code dacean and Hexamita N da Flatworm Mitochond lear Code Plant Plastid Code east Nuclear Code schondrial Code atworm Mitochondrial C	ondrial Code and t Nuclear Code rial Code	he Mycc	op v
Define translation f	rame O Frame 2	◯ Frame 3		
Keep nucleotide	e sequence displayed			

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...* (1).

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* (E, 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.

Find mutations	?	\times
Entries to be screened	Select all	
inflA005 inflA006 inflA007 inflA002 inflA008 inflA009 inflA010 inflA003 inflA004 inflA001		
Types of mutations to be searched		
Intergenic mutations (silent mutation)		>
Synonymous mutations (silent mutation)		>
Non-synonymous mutations (missense mutation)		>
Indels (insertions and deletions)		>
Screen focus		
Screen selected alignment		
○ Screen all alignments		
Implement IUPAC code		
Map mutations in alignment viewer		
Find	Cancel	

Figure 7: Mutation settings.



Figure 8: Mutation listing.