

# BIONUMERICS Tutorial: Importing sequences from FASTA files

### 1 Aim

With the BIONUMERICS FASTA import routine, sequences in FASTA format can be imported into BIONUMERICS. In this tutorial you will learn how to use this import tool by importing sequences from an example file.

## 2 Example data

As an example we will import sequences from the FASTA file H1N1\_HA.txt into a new or existing BIONUMERICS database. This FASTA file contains the sequence of the HA (haemagglutinin) viral segment for a set of 200 Influenza A H1N1 strains. The example file can be found on the download page on our website (https://www.applied-maths.com/download/sample-data, "FASTA files").

### 3 The Import wizard

- 1. Create a new database (see tutorial "Creating a new database") or open an existing database.
- 2. Select *File* > *Import...* (, Ctrl+I) to open the *Import* dialog box.
- 3. Choose the option *Import FASTA sequences from text files* under the *Sequence type data* item in the tree and click <*Import*>.
- 4. The import wizard allows you to browse for one or more text files as data source. Press <*Browse*>, navigate to the folder, select the H1N1\_HA.txt file and press <*Open*> (see Figure 1).
- 5. With the option *Preview sequences* checked, press <*Next*>.

The import wizard now displays a preview of the sequence data in the FASTA file (see Figure 2). From this preview, it is clear that the first FASTA field contains the accession number, the second field the strain number, the third field the date of isolation and the fourth field the country of origin.

6. Press <*Next*>.

Import sequence	es	?	×
	A formatted sequences data to import.		
Select file(s):	C:\Users\Public\Documents\H1N1_HA.bt       Browse         Delete       Delete         1 file(s), less than 1 Mb		
	< Back Next >	Can	cel

Figure 1: Select the FASTA file(s).

N	r. File name	Length	Header	<u></u>
⊻ 1	H1N1_HA	1734	CY071135 A/Athens/INS389/2010 2010-01-13 Greece	
∠ 2	H1N1_HA	1734	CY071143 A/Athens/INS390/2010 2010-01-14 Greece	
☑ 3	H1N1_HA	1734	CY075059 A/Athens/INS392/2010 2010-01-06 Greece	
✓ 4	H1N1_HA	1734	CY075067 A/Athens/INS393/2010 2010-01-06 Greece	
⊻ 5	H1N1_HA	1734	CY071159 A/Athens/INS394/2010 2010-01-11 Greece	
☑ 6	H1N1_HA	1734	CY071167 A/Athens/INS396/2010 2010-01-21 Greece	
☑ 7	H1N1_HA	1734	CY071175 A/Athens/INS397/2010 2010-01-24 Greece	
28	H1N1_HA	1734	CY071183 A/Athens/INS398/2010 2010-01-24 Greece	
🗹 9	H1N1_HA	1734	CY071263 A/Athens/INS411/2010 2010-01-30 Greece	
✓ 10	D H1N1_HA	1734	CY071271 A/Athens/INS412/2010 2010-02-02 Greece	
🗹 11	I H1N1_HA	1734	CY071279 A/Athens/INS413/2010 2010-02-05 Greece	
✓ 1:	2 H1N1_HA	1739	CY071287 A/Athens/INS414/2010 2010-02-05 Greece	
✓ 1:	3 H1N1_HA	1734	CY075075 A/Athens/INS416/2010 2010-02-17 Greece	
1	4 H1N1 HA	1734	CY071295IA/Athens/INS417/2010I2010-01-18IGreecel	$\sim$

Figure 2: Sequence preview.

The next step of the import wizard lists the templates that are present to import sequence information in the database. As this is the first time we import FASTA formatted sequences in the database, we need to create a new import template by specifying *Import rules*.

- 7. Click < *Create new*> to create a new import template.
- Select *Field 1* in the list and click < *Edit destination*> or simply double-click on *Field 1*. Under *Entry info field*, select *Create new* and press < *OK*>. Change the suggested name for the new field into *Accession number* and confirm the action twice.
- Double-click on *Field 2*. Select *Create new* under *Entry info field* and click < *OK* >. Enter *Strain* as name for the new information field, press < *OK* > and confirm the creation of the new field with < *Yes* >.

- 10. Using the Ctrl-key, highlight *Field 3* and *Field 4* in the list. Click < *Edit destination*> or doubleclick on *Field 4*.
- 11. Highlight *Entry info field* and press < OK >.
- 12. Change the suggested names into *Date* and *Country* for Field 3 and Field 4, respectively.

Create new		?	×
Source	Destination	tields:	
Field 3	Date		
Field 4	Country		
	ОК	Canc	el

Figure 3: Create new entry information fields.

13. Press < OK > and confirm the creation of the new information fields with < Yes >.

The grid is updated (see Figure 4).

Source type	Source	Destination type	Destination	
ASTA field	Field 1	Entry information : Entry info field	Accession number	
ASTA field	Field 2	Entry information : Entry info field	Strain	
ASTA field	Field 3	Entry information : Entry info field	Date	
ASTA field	Field 4	Entry information : Entry info field	Country	
ASTA field	Field 5	<none></none>	<none></none>	
ASTA field	Field 6	<none></none>	<none></none>	
ASTA field	Field 7	<none></none>	<none></none>	
ASTA field	Field 8	<none></none>	<none></none>	
ASTA field	Field 9	<none></none>	<none></none>	
ASTA field	Field 10	<none></none>	<none></none>	
ASTA field	Field 11	<none></none>	<none></none>	
ASTA field	Field 12	<none></none>	<none></none>	
ASTA field	Field 13	<none></none>	<none></none>	
ASTA field	Field 14	<none></none>	<none></none>	
Edit destination Preview	]			

Figure 4: Import rules.

- 14. Optionally, you can press < *Preview* > to obtain a preview of the data you are about to import.
- 15. Click <*Next*>.
- 16. Do not select an *Entry link field*. Press < *Finish*>.
- 17. Specify a template name, (e.g. **FASTA**), and optionally enter a description. Press < **OK** >.

18. Highlight the newly created template and select *Create new* as *Experiment type* (see Figure 5).

nport sequences			?	×
Import template Specify how to import data into	the database.			
Import templates:				
<default> FASTA</default>	FASTA	Create	new	
		Edi	it	
		Previ	ew	
		Сор	w	
		00	· ···	
	]			
Experiment type: <create new=""></create>	~			

Figure 5: Import template.

- 19. Press <*Next*>.
- 20. Specify a sequence type name (e.g. **HA** or **haemagglutinin**) and press < **OK** > and confirm the action.

The *Database links* wizard page will indicate that 210 new entries will be created during import (see Figure 6).

mport sequences	;						?	$\times$
	orted reco	rds to database en o get an overview.						
Overview In 'All levels'		reate 210 entries	and 🗌	update 0 en	tries			
Select modifie	d entries							
				< Back	Finish	_	Cance	al

Figure 6: The Database links wizard page.

21. Press <*Finish*> to start the import into the database.

For 210 strains, strain information and sequences for the HA genome segment are imported in the database (see Figure 7).

Import FASTA - BioNumerics							- 🗆 X
File Edit Database Analysis Scripts Window Help							
Experiment types	Database entries						Comparisons
ଡ଼ିଶ + Pੈ⊗ ଲି¦ ଲି ସ…   ↑	\$ + ₺ ⊗ ₿		<all entries=""></all>	0			+ 🖻 🛛 🗟 🛛 🗠
		Accession number	Strain	Date	Country		
	Key						
Action 1 HA Sequence types	MPORT_FASTA0000001	CY071135	A/Athens/INS389/2010	2010-01-13	Greece	•	·
	MPORT_FASTA000002 MPORT_FASTA0000003	CY071143 CY075059	A/Athens/INS390/2010 A/Athens/INS392/2010	2010-01-14 2010-01-06	Greece		
	MPORT_FASTA0000004	CY075059 CY075067		2010-01-06			
	MPORT FASTA0000004	CY075067 CY071159	A/Athens/INS393/2010 A/Athens/INS394/2010	2010-01-08	Greece Greece		
	MPORT FASTA0000006	CY071167	A/Athens/INS396/2010 A/Athens/INS396/2010	2010-01-11	Greece		
	MPORT FASTA0000007	CY071175	A/Athens/INS397/2010	2010-01-21	Greece		
	MPORT FASTA0000008	CY071183	A/Athens/INS398/2010	2010-01-24	Greece		< >
Entry fields Database design	MPORT FASTA0000009	CY071263	A/Athens/INS411/2010	2010-01-30	Greece		Identification projects Decision networks
	MPORT FASTA0000010	CY071271	A/Athens/INS412/2010	2010-02-02	Greece		
+ 1 ⊗ 🗟   🗗 🖾 1 ↓ «Alle	MPORT FASTA0000011	CY071279	A/Athens/INS413/2010	2010-02-05	Greece		2 + 13 ⊗ 8,   61 マ
Name Field type 🗨	MPORT FASTA0000012	CY071287	A/Athens/INS414/2010	2010-02-05	Greece		Name Modified date 💌
Att Accession number Fixed	MPORT FASTA0000013	CY075075	A/Athens/INS416/2010	2010-02-17	Greece		
All Strain Fixed	MPORT_FASTA0000014	CY071295	A/Athens/INS417/2010	2010-01-18	Greece		
At Date Fixed	MPORT_FASTA0000015	CY071303	A/Athens/INS419/2010	2010-01-25	Greece		
And Country Fixed	MPORT_FASTA0000016	CY071311	A/Athens/INS420/2010	2010-01-27	Greece		
	MPORT_FASTA0000017	CY071319	A/Athens/INS421/2010	2010-02-03	Greece		
	MPORT_FASTA0000018	CY071327	A/Bangkok/INS424/2010	2010-02-23	Thailand		
×	MPORT_FASTA0000019	CY071335	A/Bangkok/INS425/2010	2010-02-25	Thailand	•	< >
	MPORT_FASTA0000020	CY071343	A/Bangkok/INS426/2010	2010-02-25	Thailand		Alignments BLAST projects Chrom. Comp.
Fingerprint files Power assemblies Annotations	MPORT_FASTA0000021	CY071351	A/Bangkok/INS427/2010	2010-03-04	Thailand	•	
- 月 + P ⊗ 🗟   🗗 🗠 - «All Fingerpi	MPORT_FASTA0000022	CY071359	A/Bangkok/INS428/2010	2010-03-04	Thailand	•	+乃⊗良∣品 ऱ
File name Experiment type Link 🔻	MPORT_FASTA0000023	CY071375	A/Berlin/INS430/2010	2010-02-05	Germany	•	Name Modified date 🔻
	MPORT_FASTA0000024	CY062979	A/Bonn/INS178/2010	2010-01-05	Germany	•	
	MPORT_FASTA0000025	CY066447	A/California/VRDL1/2010	2010-01-09	USA	•	
	MPORT_FASTA0000026	CY063179	A/California/VRDL10/2010	2010-02-11	USA	•	
	MPORT_FASTA0000027	CY063187	A/California/VRDL11/2010	2010-01-26	USA	•	
	MPORT_FASTA0000028	CY063195	A/California/VRDL12/2010	2010-01-27	USA	•	
	MPORT_FASTA0000029	CY063203	A/California/VRDL13/2010	2010-01-25	USA	•	
	MPORT_FASTA0000030	CY063211	A/California/VRDL14/2010	2010-01-25	USA	•	
< >	<ul> <li></li> </ul>				,	• •	· · · · · · · · · · · · · · · · · · ·
Database: Import FASTA (_DefaultUser_) Entries: Loaded=210, View=2	10, Selected=210 1 experiments	C:\Users\Public\Docume	ents\BioNumerics\Data BN8\Impor	t FASTA This is	a time limited package	valid until 2020-12-30	:

Figure 7: The Main window after import of the sequences.

## 4 Conclusion

In this tutorial you have seen how easy it is to import FASTA formatted sequences in BIONUMER-ICS. The sequences can now be analyzed in BIONUMERICS. More information can found in the analysis tutorials on our website.