



## BIONUMERIC Tutorial:

# Importing sequences from GenBank/EMBL files

## 1 Aim

---

With the BIONUMERIC GenBank/EMBL import routine, sequences in GenBank/Embl format can be imported into BIONUMERIC. 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 Influenza A sequences of the hemagglutinin (HA) gene into a new or existing BIONUMERIC database. The example GenBank file (H1N1 HA gene.gb file) can be found on the download page on our website (<https://www.applied-maths.com/download/sample-data>, "GenBank file"). Note that the steps for the import of an EMBL formatted file are the same.

## 3 The Import wizard

---

1. Create a new database (see tutorial "Creating a new database") or open an existing database.
2. In the *Main* window, select **File > Import...** (📄, **Ctrl+I**) to open the *Import* dialog box.
3. Choose the option **Import EMBL/GenBank sequences from text files** under the **Sequence type data** item in the tree and click **<Import>**.
4. Press **<Browse>**, select the H1N1 HA gene.gb file and press **<Open>**.
5. With the option **Preview sequences** checked, press **<Next>**.

The import wizard now displays a preview of the sequence data in the file (see Figure 1).

6. Press **<Next>**.

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 GenBank 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.

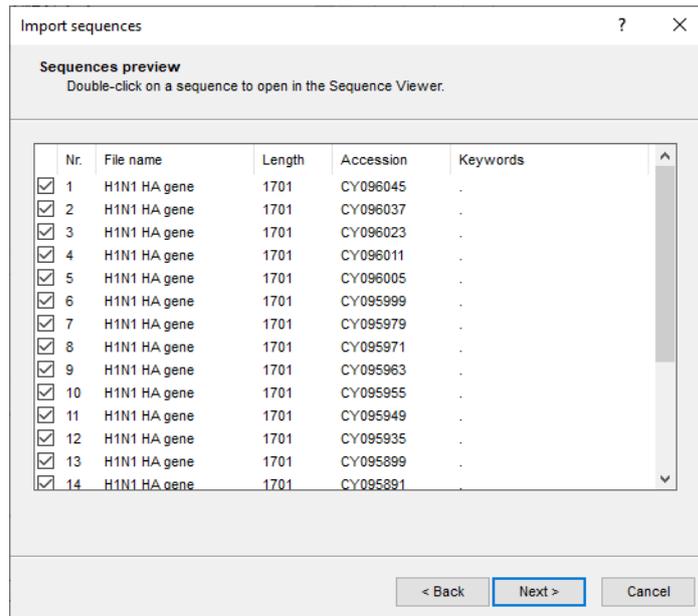


Figure 1: Preview.

8. Select **OS - SOURCE** in the list and click **<Edit destination>**.

9. Select **Create new** under **Entry info field** and click **<OK>**.

10. Enter "Strain" as name for the new information field, press **<OK>** and press **<Yes>**.

The accession number is valuable information which is specific for the sequence (not for the strain), therefore we will store the accession number as a sequence information field.

11. Select **AC - ACCESSION** in the list and click **<Edit destination>** or double-click on **AC - ACCESSION**. Under **Sequence info field**, select **Create new** and press **<OK>**.

12. Change the suggested name for the new information field into **Accession number**, press **<OK>** and confirm with **<Yes>**.

The grid is updated (see Figure 2).

13. Optionally, you can press **<Preview>** to obtain a preview of the data you are about to import.

14. Click **<Next>** to go to the next step.

15. Do not select an **Entry link field**. Press **<Finish>**.

16. Specify a template name (e.g. **GenBank**) and optionally enter a description. Press **<OK>**.

17. Highlight the newly created template and select **Create new** as **Experiment type** (see Figure 3).

18. Press **<Next>**.

19. 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 20 new entries will be created during import (see Figure 5).

20. Press **<Finish>**.

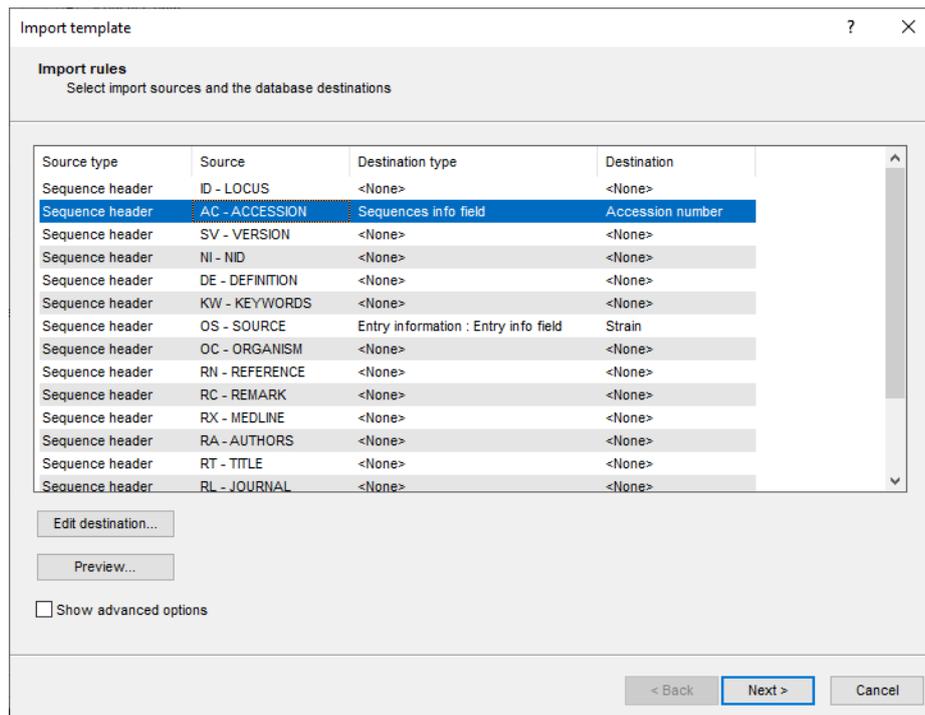


Figure 2: Import template

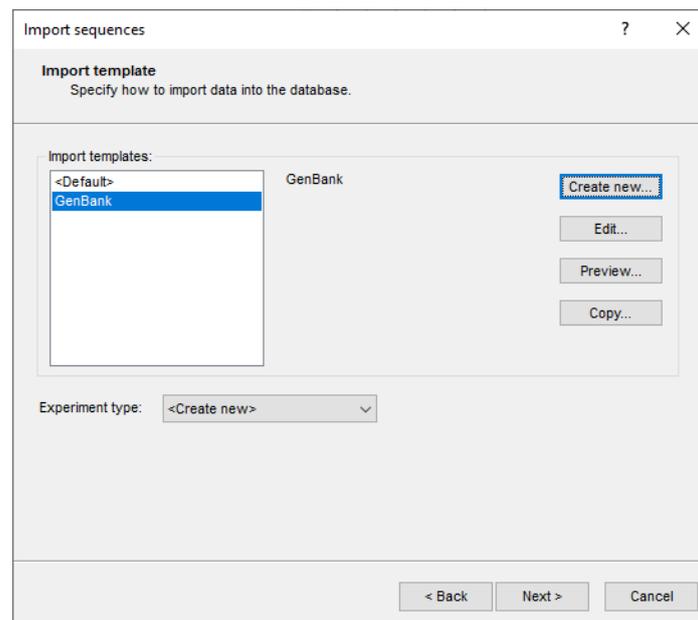


Figure 3: Import template.



Figure 4: Create a new sequence type.

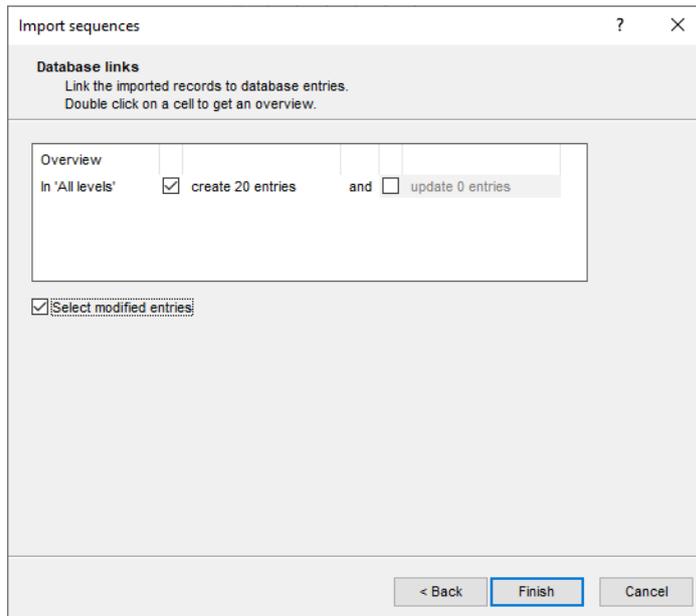


Figure 5: The *Database links* wizard page.

Twenty sequences are imported in the database (see Figure 6). All entries for which information was imported are automatically selected.

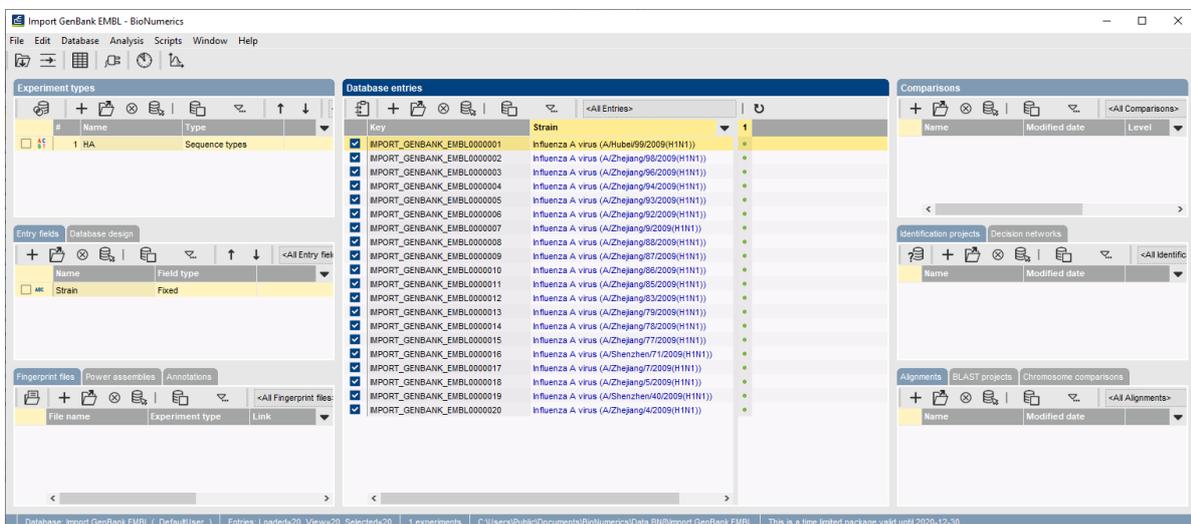


Figure 6: The *Main* window.

21. Click on a green colored dot in the *Experiment presence* panel to open the *Sequence editor* window.

The sequence is displayed in the upper panel and a graphical representation of the sequence is displayed in the panel below. The *Annotation* panel (see Figure 7) holds the GenBank features, the header information is stored in the *Header* panel and the accession number is stored in the *Custom Fields* panel.

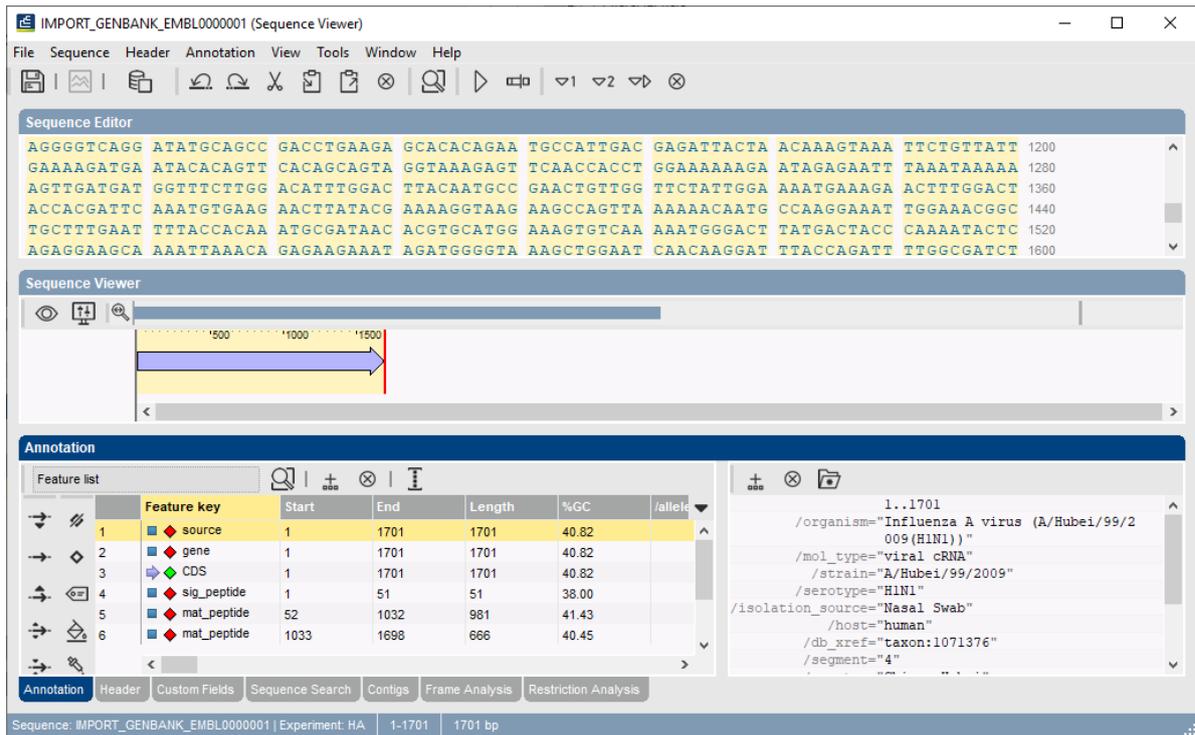


Figure 7: The *Sequence editor* window.

## 4 Conclusion

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