BioNumerics Tutorial:

Analyzing an individual sequence in the Sequence Editor

1 Aim

The *Sequence editor* window is a convenient tool implemented in BioNumerics to edit and analyze nucleotide and amino acid sequences. The nucleotide sequence of L77616, the Oryza sativa (rice) dihydrodipicolinate synthase gene, will be used in the next paragraphs to illustrate the features of the *Sequence editor* window.

2 Preparing the database

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 ***Download sequences from internet*** under the ***Sequence type data*** item in the tree and click <***Import***>.
4. Enter the accession code **L77616** in the ***Accession codes*** input field and press <***Next***> (see Figure 1).

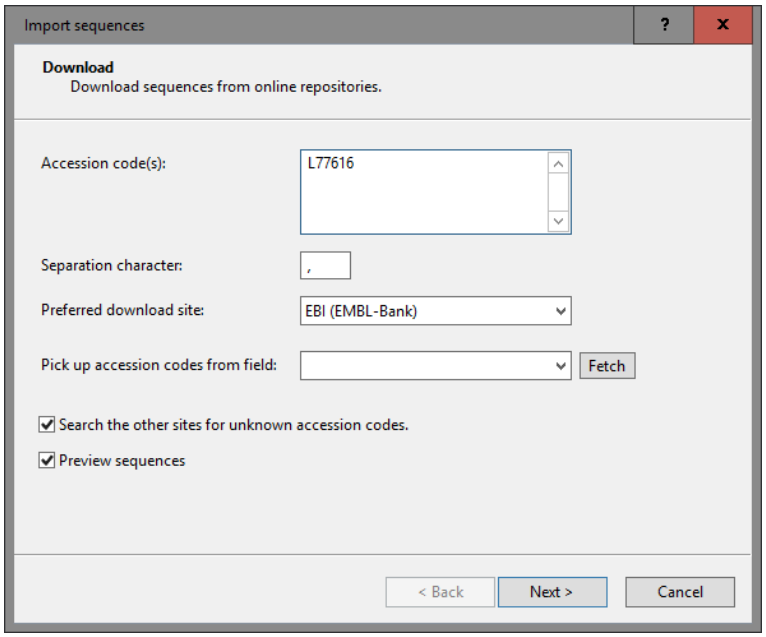

Figure 1: Download sequence from EBI.

The import routine fetches the sequence from the selected database and shows detailed information in the next step.

5. Press <***Next***>.

6. Make sure the option ***Create new*** is selected as ***Experiment type*** and press <***Next***> once more.

7. Specify a sequence type name (e.g. **dhdps**) and press <***OK***> and confirm the action.

8. Press <***Finish***>.

The sequence is imported in the database.

3 General features

1. Click on the green colored dot of entry **L77616** in the *Experiment presence* panel to open the *Sequence editor* window.

The upper panel shows the numbered sequence, with bases grouped in blocks of 10. Non-translated sequence parts are displayed in black lowercase, exons in blue uppercase, and introns in green lowercase.

The middle panel shows a graphical representation of the sequence. The zoom slider allows one to continuously zoom in or out on the graphical sequence view. Zooming can be done up to base level.

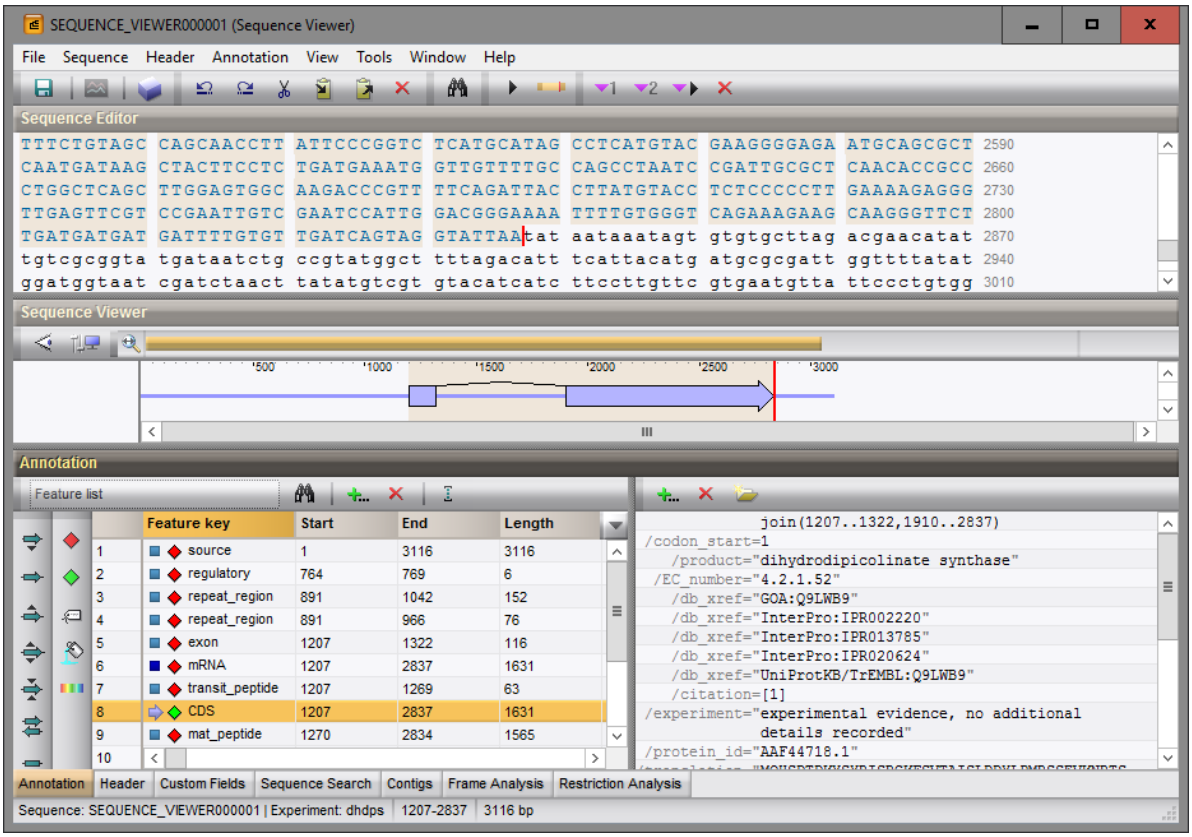

Figure 2: The *Sequence editor* window: the *Annotation* panel.

The feature information is stored in the *Annotation* panel (see Figure 2).

If the feature information is not shown, click on the arrow down button in the *Annotation panel* next to the annotation columns and select ***Restore default settings***.

2. Click on a particular feature, for example **CDS**.

The selected feature is highlighted with an orange background in the upper and middle panels and is highlighted in the feature list in the lower panel. The *qualifiers* associated with the selected feature are given in

the right panel.

3. Click on the *Header tab*.

The EMBL description fields are displayed in the *Header tab*.

4 Frame analysis

The frame analysis in the *Frame Analysis* panel analyzes a nucleotide sequence in function of its six possible translation frames.

1. Click on the *Frame Analysis tab* to put it into focus.

The frame analysis is performed using settings that can be accessed with ***Tools*** > ***Frame analysis*** > ***Filter settings...*** ( ).

2. Select ***Tools*** > ***Frame analysis*** > ***Show frame analysis in sequence viewer*** ( ) to map the reading frames on the sequence plot (see Figure 3).

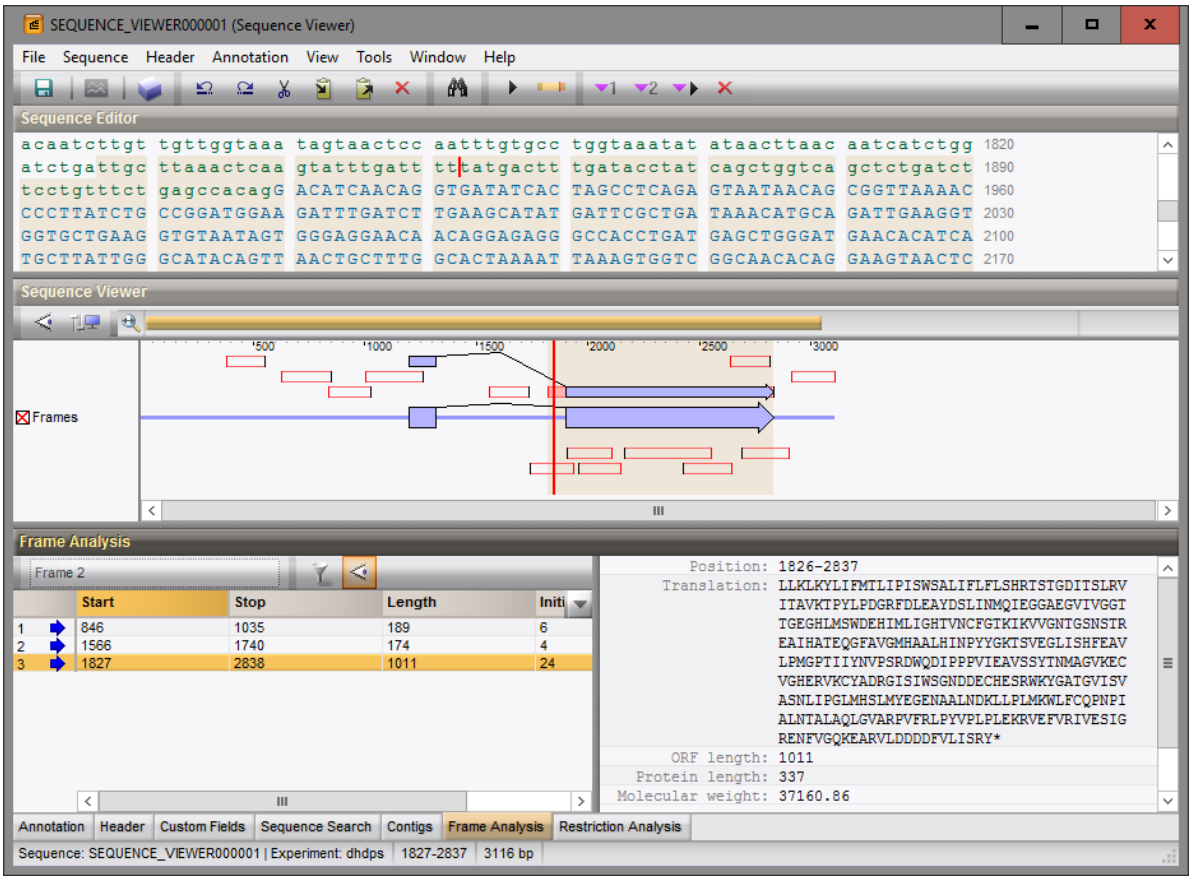

Figure 3: The *Frame Analysis* panel.

The *Sequence Viewer* panel shows a graphical view of the sequence along with its 6 possible reading frames. The three reading frames of the forward strand are mapped above the sequence plot, while the three reading frames of the reverse strand are mapped below.

3. The zoom slider in the *Sequence Viewer* panel allows zooming from full-length sequence view up to base level view.

In the *Frame Analysis* panel, the frame can be selected from the drop-down list. The ORFs that are found

in this frame are listed on the left. ORFs in the direct orientation are indicated with a blue arrow pointing to the right. Red arrows pointing left indicate ORFs in the reverse orientation.

4. Click on an ORF in the left panel.

Additional information about the ORF and the predicted protein is displayed on the right.

5 Restriction enzyme analysis

Using the restriction enzyme analysis tool, in silico digests of nucleic acids can be generated for a selected restriction enzyme or a combination of restriction enzymes. The resulting fragments can then be rendered as a virtual gel image.

1. Click on the *Restriction Analysis tab* in the lower panel.

2. Select ***Tools*** > ***Restriction analysis*** > ***Add enzyme map...*** ( ).

3. Select ***Select single enzyme from list***, select ***PstI*** and press <***OK***>.

4. Double-click on the enzyme in the *Restriction Analysis* panel or select ***Tools*** > ***Restriction analysis*** > ***Add selected enzyme to fragment list*** ( ) to map the restriction enzyme onto the sequence (see Figure 4).

5. Select ***Tools*** > ***Restriction analysis*** > ***Add enzyme map...*** ( ) again.

6. Select ***Select single enzyme from list***, select ***SpmI***, make sure ***Replace list*** is unchecked and press <***OK***>.

7. Double-click on the enzyme in the *Restriction Analysis* panel or select ***Tools*** > ***Restriction analysis*** > ***Add selected enzyme to fragment list*** ( ) to map the restriction enzyme onto the sequence.

The fragments from both enzymes are now displayed in the right panel (see Figure 4).

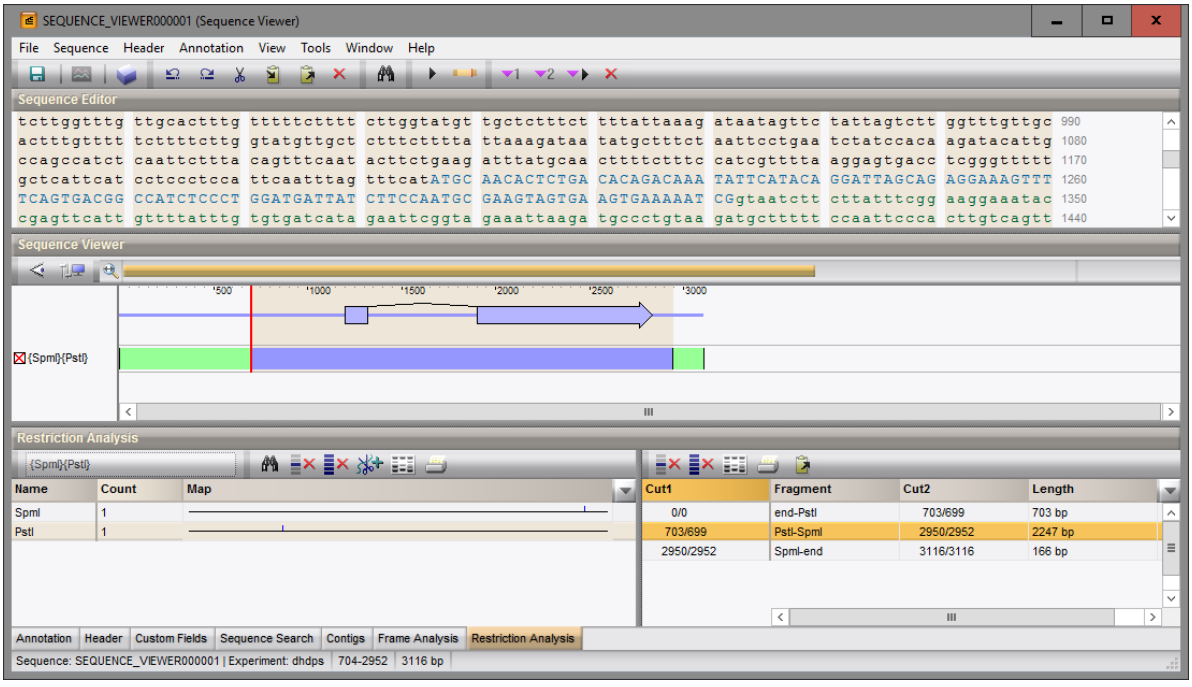

Figure 4: The *Restriction Analysis* panel.

8. Click on a fragment in the fragment list (right panel).

The fragment is highlighted in purple on the map image. Non-selected fragments are shown in green.

A virtual gel can be created as follows:

9. Highlight "PstI" and select ***Tools*** > ***Restriction analysis*** > ***Add selected enzyme to gel window*** ().

Hover with the mouse over a band to see its molecular size.

10. Highlight "SpmI" and select ***Tools*** > ***Restriction analysis*** > ***Add selected enzyme to gel window*** () to add it as a lane to the virtual gel.

11. Select one of the fragments in the right panel and select ***Tools*** > ***Restriction analysis*** > ***Add fragments to gel window*** () to add the fragments to the virtual gel.

12. In the *Fingerprint fragment lists* window, select ***Edit*** > ***Add marker*** to display a molecular weight marker pattern on the gel.

13. If desired, the position of a gel pattern can be changed with ***Edit*** > ***Move to the left*** (**Ctrl+LEFT**) or ***Edit*** > ***Move to the right*** (**Ctrl+RIGHT**).

14. Remove all fragments from the list with ***Tools*** > ***Restriction analysis*** > ***Clear fragment list.*** ().

15. Remove all enzymes from the map with ***Tools*** > ***Restriction analysis*** > ***Clear map list*** ().

Restriction enzymes can also be selected by using more general criteria.

16. Select ***Tools*** > ***Restriction analysis*** > ***Add enzyme map...*** ().

17. Check ***User filter settings*** and click on <***Change filter settings***>.

18. Under ***Recognition pattern and cleavage type***, uncheck ***Blunt end*** and ***3' protruding end*** and leave all other options checked.

19. Under ***Length recognition site***, check ***n=4*** and uncheck all other options.

20. Uncheck ***Include restriction enzymes not commercially available*** and make sure all other options are checked under ***Restriction enzymes*** and ***Methylation***.

21. Finally, set ***Minimal cuts*** to "1" and ***Maximal cuts*** to "4". Check ***Replace list*** and press <***OK***> twice to apply the settings.

The enzymes that fulfill the specified criteria are shown in the left panel.

22. Map an enzyme onto the sequence with ***Tools*** > ***Restriction analysis*** > ***Add selected enzyme to fragment list*** ().

6 Primer design

The primer analysis application in BioNumerics has been designed to calculate optimal primers and primer combinations for the amplification of a target region in function of various experimental parameters. When launching the primer analysis tool from the *Sequence editor* window, the target region can be adapted from the complete sequence, or from a (feature) selection.

1. In the *Annotation* panel of **L77616**, click on the CDS feature to select the corresponding sequence.

2. Select ***Tools*** > ***Primer design...*** to display the *Primer design* window.

The *Primer design* window is divided into an upper panel displaying the nucleotide sequence (the *Sequence viewer panel*), and a lower panel, giving information concerning the primers and primer combinations found (initially empty).

3. Use the zoom slider to zoom in on the sequence.

Zooming can be done up to base level. The translation of the CDS features – if present – is shown underneath the nucleotide sequence. The red vertical line indicates the cursor position on the sequence.

Before primers and primer combinations can be searched for, a target region needs to be defined on the plot:

4. Select ***Target*** > ***Edit...***.

A dialog box appears, asking for some specifications about the target region. In this example we will locate the primer regions up- and downstream of the selected CDS.

5. Make sure ***Selected feature*** is checked, press <***Extended selection***> and extend the selected CDS region with 300 bases upstream and 300 bases downstream. Check ***Exclude selection as priming target*** and press <***OK***> twice.

The target design is plotted on the sequence (see Figure 5).

6. Select ***File*** > ***Run calculation*** ( ) to display the Primer and PCR settings.

7. Leave the settings at their default values and press <***OK***>.

The left and right panels in the *Primers tab* list the most suitable forward and reverse primers respectively (see Figure 5).

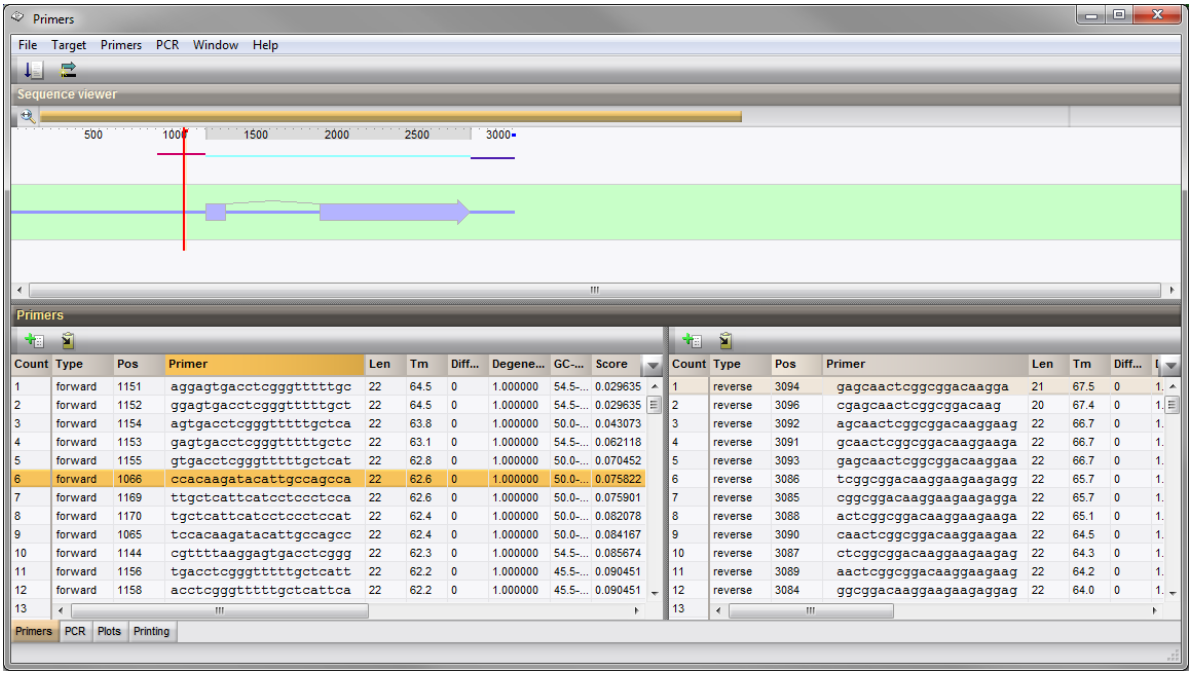

Figure 5: The *Primers tab*.

The primer sequence is displayed in the ***Primer*** column. The length of the primer is shown in the ***Len*** column and the start position in the ***Pos*** column. The melting temperature (calculated according to the nearest-neighbor thermodynamics published by SantaLucia using the default thermodynamic settings), the number of different bases, the degeneracies, and GC-content are displayed in the ***Tm***, ***Difference***, ***Degeneracy***, and ***GC-content*** columns respectively.

The primers are ordered according to a score, which is calculated from the primer calculation parameters (***Score*** column). The primers can also be sorted according to the other features, such as position, length, degeneracy, melting temperature, and number of different bases (in case discriminating primers were searched for). Arranging the primer list according to another feature (for example base difference) can help select suitable primers out of the full list.

8. To sort the list according to another feature, right-click with the mouse pointer on a header field and select ***Sort***.

9. Select a primer from the forward or reverse primer list.

The sequence of the selected primer is highlighted with an orange background in the *Sequence viewer panel*. A red line is shown in the sequence position bar when a forward primer is selected, a reverse primer is recognized by a blue line.

10. Information of forward or reverse primers can be copied to the clipboard with ***Primers*** > ***Forward*** > ***Export to clipboard*** ( ) or ***Primers*** > ***Reverse*** > ***Export to clipboard*** ( ), respectively. From the clipboard, this information can be pasted in other applications.

11. Alternatively, add the information of forward or reverse primers to the printing list with ***Primers*** > ***Forward*** > ***Add to printing list*** ( ) or ***Primers*** > ***Reverse*** > ***Add to printing list*** ( ). This list is displayed in the *Printing panel* and can be sent to a printer from there.

The *PCR panel* lists the obtained PCR products, ordered based on their scores. Their primer information can also be exported to the clipboard or added to the printing list.

The *Plots panel* shows on the left-hand side the degeneracy and melting temperature over the length of the sequence. On the right, the degeneracy of each possible primer (X-axis) is plotted against its melting temperature (Y-axis). If a primers falls within the settings specified, it is plotted purple, if not it is plotted gray.

7 Blast analysis

When looking at detailed sequence information in the *Sequence editor* window, one may want to start a BLAST search on the complete sequence or on a selected subsequence.

1. In the *Annotation* panel of **L77616**, click on the CDS feature to select the corresponding sequence.

2. With the subsequence selected, select ***Tools*** > ***BLAST analysis...***.

This action will launch the *BLAST search settings* dialog box (see Figure 6).

Before the BLAST analysis is launched, the BLAST program and BLAST screening database need to be defined in the *BLAST search settings* dialog box.

3. In this example, select ***blastn*** (or ***blastx***), leave the other settings unaltered and press <***OK***>.

A BLAST analysis is launched for the selected subsequence and the *BLAST* window is displayed. The analysis might take a couple of minutes, depending on the length of the selected sequence, the BLAST program and BLAST screening database.

Once the BLAST analysis has completed, the results are displayed in the *BLAST* window:

- The *BLAST Report* panel provides an overview of the BLAST screening settings i.e. the program that was used to perform the search and the related program settings, and also the database that was used for screening.

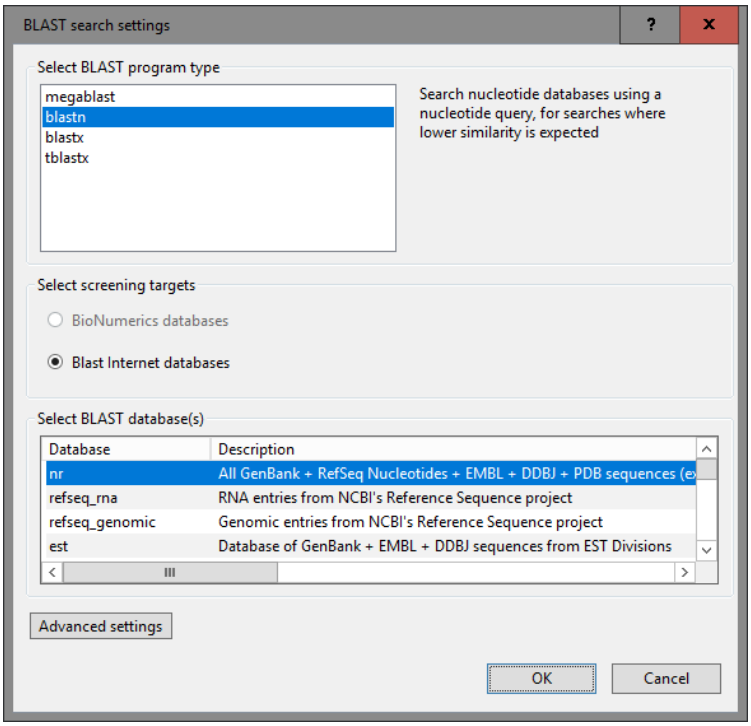

Figure 6: The *BLAST search settings* dialog box.

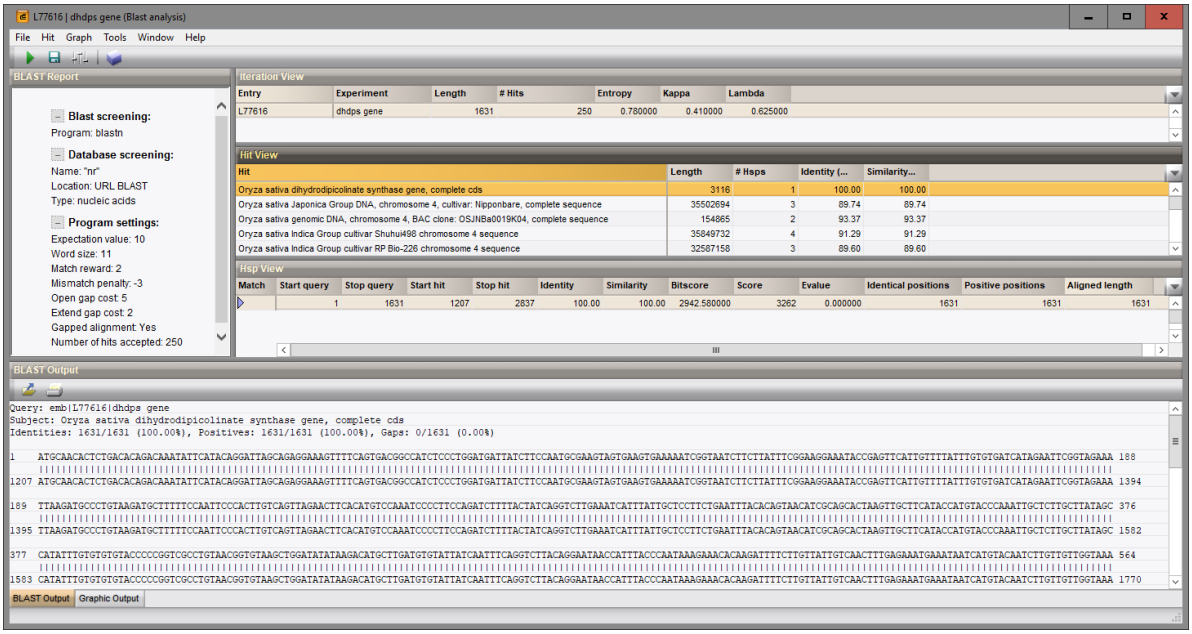

Figure 7: BLAST result.

- Information about the blasted (sub)sequence is given in the *Iteration View* panel, with information about the returned hits.
- In the *Hit View* panel, the different BLAST hits are listed by decreasing identity values. For each hit, the header definition of the sequence is displayed in the Hit column, followed by the length of the full hit sequence, the number of High-scoring segment pairs (HSPs) per hit, and the identity and similarity value of the HSP with the highest alignment score in a given search.
- Detailed Hsp information about the selected hit is displayed in the *Hsp View* panel.

- For the selected Hsp, the *BLAST Output* panel displays the alignment, preceded by the sequence identifier, the full subject definition line, and the number of identical residues in this alignment (Identities), the number of conservative substitutions (Positives), and if applicable, the number of gaps in the alignment. Beneath that, the actual alignment is shown, with the query on top, and the database match below.

4. (Optionally) save the BLAST analysis with ***File*** > ***Save...*** ( ). This opens the *Save BLAST object* dialog.

5. Specify a BLAST object name and press <***OK***> (see Figure 8).

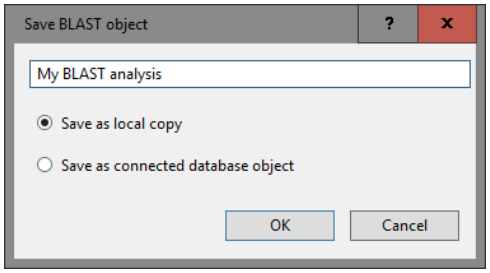

Figure 8: Save the BLAST project.

6. Close the *BLAST* window and *Sequence editor* window.

7. Click on the tab of the *BLAST projects* panel in the *Main* window.

The saved BLAST project is listed (see Figure 9) and can be consulted again by double-clicking the BLAST project name.

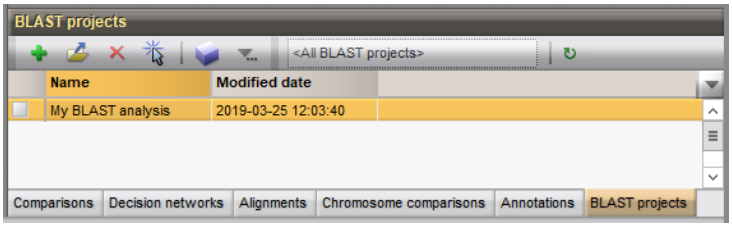

Figure 9: Saved BLAST project.

The BLAST functionality is also accessible from other windows in BioNumerics: *Main* window, *Sequence alignment* window, *Chromosome Comparison* window, and *Annotation* window. Please check the other tutorials on our website and the reference manual for detailed information.