



BIONUMERICS Tutorial:

Setup and analysis using your own custom MLST scheme

1 Introduction

This tutorial illustrates the steps needed to set up your own MLST schema, e.g. in case you are studying a less common organism for which no MLST data is available yet.

2 Example data

We will set up an MLST schema for the fictitious organism *Ambiorix belgiensis*, based on an – equally fictitious – set of seven housekeeping genes named *wtlF*, *frt*, *mslS*, *orvL*, *mnkP*, *beeR*, and *wffL*. The trace files for this tutorial are generated by an Applied Biosystems 3130XL capillary sequencer and can be downloaded from the Applied Maths website (<https://www.applied-maths.com/download/sample-data>, click on "MLST sample AB1 trace files"). There is a pair of AB1 files (forward and reverse primer) for each entry / housekeeping gene combination. The optimal trimming positions for each of the loci are already determined and can be found in the `Custom MLST.txt` text file.



To manually determine the optimal trimming positions, first import and assemble the trace files without trimming and then perform a sequence multiple alignment to find conserved regions that cover a sufficiently large DNA sequence.

3 Creating a new database

To illustrate the complete setup of the *MLST online plugin* for the creation of your own MLST schema, we will start by creating a new, empty database.

1. Double-click on the BIONUMERICS icon () on the desktop.
2. In the *BIONUMERICS Startup* window, press the  button to enter the *New database* wizard.
3. Enter a database name, e.g. "Ambiorix belgiensis".
4. Click <Next>.

A new dialog box pops up, asking whether to create a new relational database for data storage or to use an existing one.

5. Leave the default option **Create new** enabled and press **<Next>**.

The next dialog asks which database engine should be used for storing data.

6. Select the default option and press **<Finish>**.

The *Main* window opens with an empty database.

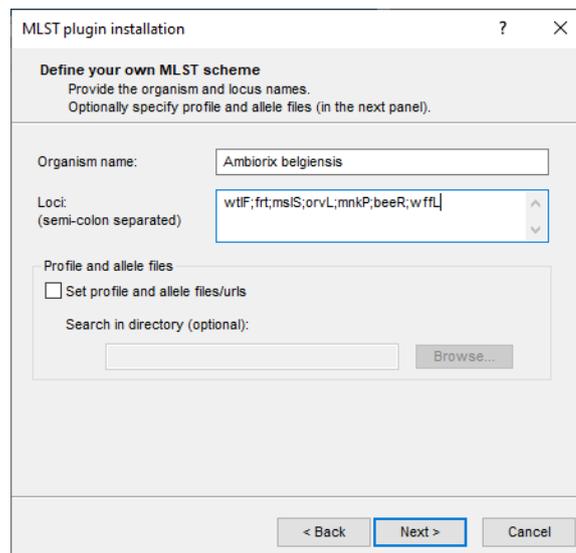
4 Installing the MLST online plugin

In this section we will install the *MLST online plugin* in the **Ambiorix belgiensis** database and set it up to create our own MLST schema.

1. Select **File > Install / remove plugins...** () in the *Main* window to display the *Plugins* dialog box.
2. Highlight the *MLST online plugin* in the list in the *Applications tab* and press the **<Activate>** button.

The next dialog asks to confirm the installation of the *MLST online plugin*. Installation of the plugin requires administrator privileges on the relational database.

3. Confirm the installation of the *MLST online plugin*.
4. Choose the option **Define your own MLST scheme** and press **<Next>**.
5. Enter “Ambiorix belgiensis” for **Organism name** and enter “wtlF;frt;mslS;orvL;mnkP;beeR;wffL” (see the *Custom MLST.txt* file for the loci names) for **Loci** (see Figure 1).



MLST plugin installation

Define your own MLST scheme
Provide the organism and locus names.
Optionally specify profile and allele files (in the next panel).

Organism name:

Loci:
(semi-colon separated)

Profile and allele files
 Set profile and allele files/urls

Search in directory (optional):

< Back Cancel

Figure 1: The *Define your own MLST scheme* dialog box for specifying organism name and a list of loci.

In case you should have text files containing profile and allele information for your organism (e.g. if they were sent to you by a colleague who works on the same organism), you should check the option **Set profile and allele files/URLs**. With this option checked, profile and allele file locations can be specified in the next step. The FASTA text files should be formatted as follows:

> [ID]

ACTG...

With [ID] an integer allele identifier (without brackets). Alternatively, the identifier line could also be:

>-[ID], >-[ID], >[LOCUS]-[ID], >[LOCUS]-[ID], >[LOCUS] [ID].

In this tutorial, we will leave the option **Set profile and allele files/URLs** unchecked since we do not have text files containing profile and allele information for our organism.

6. Make sure the option **Set profile and allele files/URLs** is unchecked and press <Next> to proceed to the *Allele trimming patterns* dialog box.
7. Copy and paste the start and stop trimming patterns from the Custom MLST.txt file for each of the loci and press <Next>.

Locus	Start pattern	Stop pattern
wtIF	GGGGAAAGGGACTCAGGCTC	CAAAGAAGCRGAAGCGGGTA
frt	CGAGCGCCATTGKCAAGCG	CCTCACGTAGCGAACTGGC
msIS	GGTCTGCACGGCGTTGGTGT	TGGCTTCCAGGAAAACATCT
orvL	CGACGCTGCAGTCGAGAAAG	ATACGCAATTGCYACGAYC
mnkP	GGYGTAGCGGTAAACCGGG	TCKGCAACCCTGTCTATGGGC
beeR	ATAACGCGCGTGAGAAAGCG	AAAGCTTACTMACTCGTGT
wffL	CGCACGTAAACTGGGCGTCG	GGTATCAACTCTACGGYGA

Calculate trimming patterns automatically

Trim pattern length:

Update trimming patterns at database startup

< Back Next > Cancel

Figure 2: Start and stop patterns.

In the final step of the wizard, the program prompts for database information fields to store the **Sequence types** and **Clonal complexes** information.

8. For this exercise, use the default "MLST ST" and "MLST CC" fields and press <Next>.
9. Pressing <Finish> starts the installation of the *MLST online plugin* according to the settings specified.

As no information needs to be downloaded from an online repository, the plugin installation will not take long. When the *MLST online plugin* is successfully installed, a confirmation message is displayed.

10. Press <OK> twice.
11. Press <Exit> to close the *Plugins* dialog box.
12. Close and reopen the database to activate the features of the *MLST online plugin*.

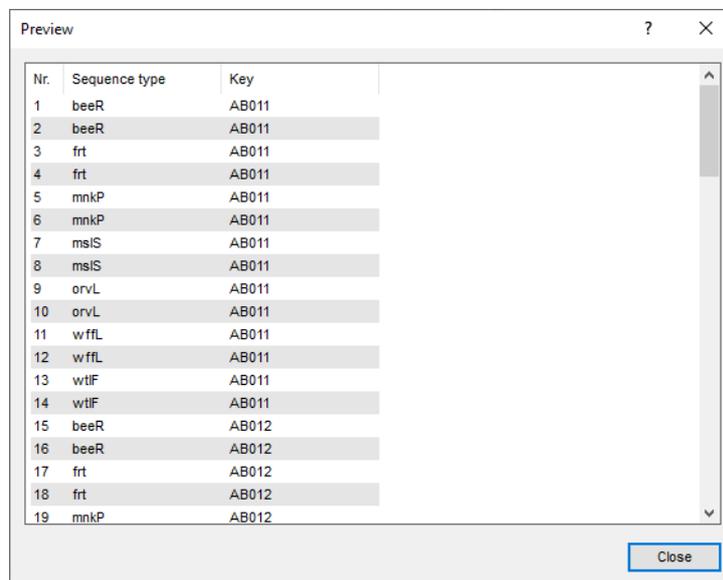
The plugin is now set up and we are ready to start assembling allele sequences.

5 Importing and assembling trace files

1. Select **File** > **Import...** (📁, **Ctrl+I**) to call the *Import* dialog box.
2. Select **Import and assemble trace files** under **Sequence type data** and press <**Import**>.
3. Press the <**Browse**> button, select all sequence trace files and press <**Open**>.
4. Press <**Next**> to proceed to the *Import template* wizard page.

The trace file names contain the strain number, locus name and primer name, each time separated by an underscore (.) character. For this naming convention, an example import template is provided.

5. Highlight **Example import 2** from the list of import templates and press <**Preview**> to check if the file names are correctly parsed.



The screenshot shows a 'Preview' window with a table containing 19 rows. The table has three columns: 'Nr.', 'Sequence type', and 'Key'. The rows are as follows:

Nr.	Sequence type	Key
1	beeR	AB011
2	beeR	AB011
3	frit	AB011
4	frit	AB011
5	mnkP	AB011
6	mnkP	AB011
7	msIS	AB011
8	msIS	AB011
9	orvL	AB011
10	orvL	AB011
11	wffL	AB011
12	wffL	AB011
13	wtlF	AB011
14	wtlF	AB011
15	beeR	AB012
16	beeR	AB012
17	frit	AB012
18	frit	AB012
19	mnkP	AB012

A 'Close' button is located at the bottom right of the window.

Figure 3: Preview.

The locus names are listed in the preview under **Sequence type** and the strain numbers under **Key**, which is what we need (see Figure 3).

6. Press <**Close**> to close the preview.
7. With the **Example import 2** import template highlighted in the list, press <**Next**>.

The *Database links* wizard page indicates that 6 new entries will be created during import.

8. Press <**Next**> to open the *Processing* wizard page.

The trimming settings as specified during the installation of the plugin will be used. The assembly and report settings can be left at their default values for this exercise.

9. Press <**Finish**> to have the sequences automatically assembled.

When the assemblies are processed, the *Batch sequence assembly report* window appears. This report displays the entries (keys) as rows and the experiments as columns (see Figure 4). This window can also be displayed from the *Main* window with **Analysis** > **Sequence types** > **Batch assembly reports....**

Key	beeR	frt	mnkP	mslS	orvL	wffL	wtIF
AB011	OK	warning	OK	OK	OK	OK	OK
AB012	OK	OK	OK	OK	OK	OK	OK
AB024	OK	OK	OK	OK	OK	OK	OK
AB025	OK	OK	OK	OK	OK	OK	OK
AB075	OK	OK	OK	OK	OK	OK	OK
AB111	OK	OK	OK	OK	OK	OK	OK

Code	Message	Status	Comment
info	Created new assembly Report for AB011 / beeR		

Figure 4: Batch Report.

6 Checking the sequence assemblies

Before alleles can be identified on the assembled sequences, each assembly in the *Batch sequence assembly report* window that has a warning or error needs to be further checked upon. For this batch of assembled sequences, there is only a single assembly that needs to be verified (i.e. **AB011/frt**) (see Figure 4).

1. Click on the **AB011/frt** warning message cell to update the *Details* panel for this key/experiment combination.

The detailed report lists a single alignment inconsistency at position 485.

2. Double-click on the warning message in the detailed report.

This will open the sequence in the *Contig assembly* window, with the corresponding position in focus.

3. In the *Contig assembly* window click on the *Aligned traces* panel (see Figure 5).
4. To obtain an optimal view of the curves, use the zoom sliders in the *Traces* panel or use the zoom buttons.
5. Press the **Del**-key on the keyboard to delete the G at position 485 in the reverse trace file (on top). The consensus will be automatically updated.
6. Select **Batch sequence assembly > Set report to solved, save and close (Ctrl+Shift+S)** in the *Contig assembly* window.
7. Close the *Batch sequence assembly report* window.

7 Identifying alleles and profiles

Since we are building our own MLST schema, we start out with an empty alleles and profiles database. By screening more and more strains, alleles and profiles will be added to the database.

Optionally, to illustrate that currently no alleles can be assigned, you can try the following:

1. With all strains in the database selected, use **MLST > Identify alleles**.

The *MLST identify errors window* now lists an identification error for each strain/locus combination.

2. Close the *MLST identify errors window* again.

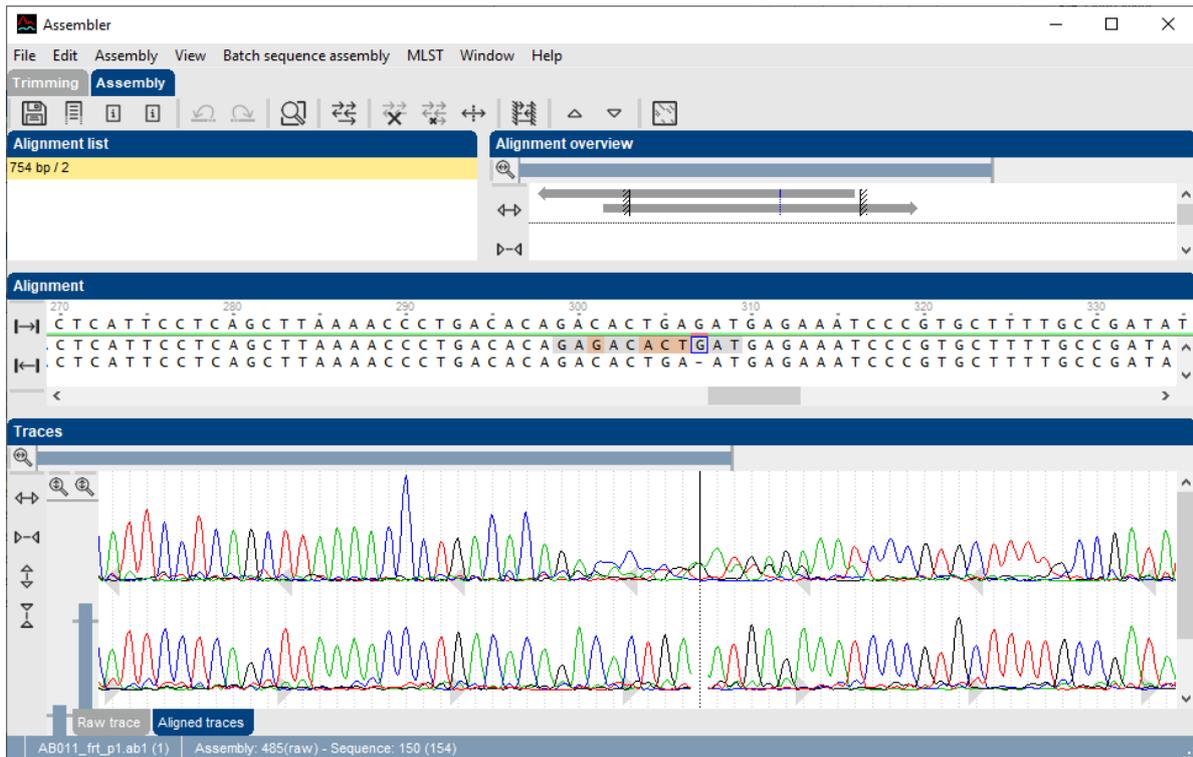


Figure 5: The *Contig* assembly window.

3. Make sure the 6 entries are selected in the *Main* window and select **MLST > Update alleles and profiles**.
4. In the *Alleles* tab of the *Update MLST alleles/profiles dialog box*, press <**Screen selection**>.

In total for all seven loci, 27 new alleles were found among the selected entries and these are now listed in the *Update MLST alleles/profiles dialog box* (see Figure 6).

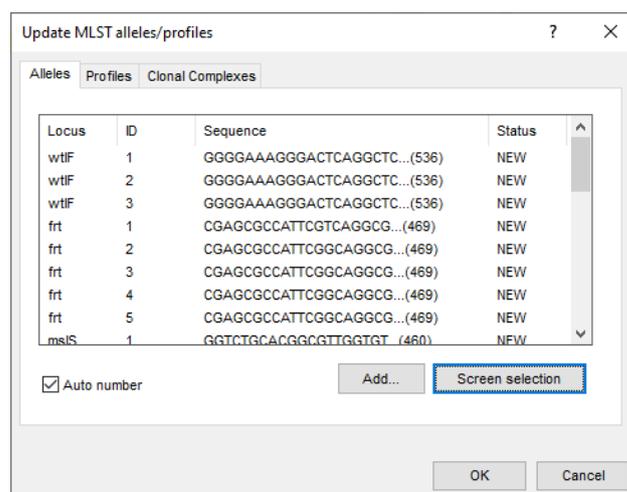


Figure 6: The *Update MLST alleles/profiles dialog box* showing new alleles found in the selection.

5. Press <**OK**> in the *Update MLST alleles/profiles dialog box* to add the new alleles to the database.

Now we can successfully assign alleles for all entries in the database:

6. Make sure the 6 entries are selected and select **MLST > Identify alleles**.

The matched allele IDs are stored as character values in the **MLST** character type, as can be seen in the *Experiment card* window:

7. Click on the green dot in the **MLST** column of the *Experiment presence* panel to open the character *Experiment card* window for an entry (see Figure 7).

Character	Value	Mapping
wtIF	3	<+>
frt	4	<+>
msIS	3	<+>
orvL	3	<+>
mnkP	3	<+>
beeR	1	<+>
wffL	3	<+>

Figure 7: Character experiment card.

8. Close the *Experiment card* window by clicking in the small triangle-shaped button in the left upper corner.

Similar as for alleles, there is currently no profile information present in the database. Therefore, we should add this info by screening the strains:

9. Make sure the 6 entries are selected in the *Main* window and select **MLST > Update alleles and profiles**.

10. In the *Profiles* tab of the *Update MLST alleles/profiles dialog box*, press <**Screen selection**>.

Five profiles were found among the selected entries and these are now listed in the *Update MLST alleles/profiles dialog box* (see Figure 8).

ST	wtIF	frt	msIS	orvL	mnkP	beeR	wffL	Status
5	3	5	4	4	4	3	4	NEW
4	3	4	3	3	3	1	3	NEW
3	2	3	2	2	2	2	2	NEW
2	1	2	1	1	1	1	1	NEW
1	1	1	1	1	1	1	1	NEW

Figure 8: The *Update MLST alleles/profiles dialog box* showing new profiles found in the selection.

11. Press <**OK**> in the *Update MLST alleles/profiles dialog box* to add the new profiles to the database.

Now we can successfully assign sequence types for all entries in the database:

12. Make sure the 6 entries are selected and select **MLST** > **Identify profiles**.

The matched sequences types are displayed in the 'MLST ST' information field.

To visualize and compare the allelic profiles of all entries, we can open a comparison:

13. With all six entries still selected, highlight the *Comparisons* panel and select **Edit** > **Create new object...** (+).
14. In the *Experiments* panel of the *Comparison* window, click on the eye icon (👁) that proceeds **MLST** to display the allelic profiles in the *Experiment data* panel.
15. Select **Characters** > **Show values** (📄) to show the actual allele IDs (character values) for all entries in the comparison (see Figure 9).

For the first entry screened (**AB011**), all allele sequences received the allele ID **1**. The next entry (**AB012**) had the same sequences for all loci, except locus **frt**, for which an allele ID of **2** was assigned, etc..

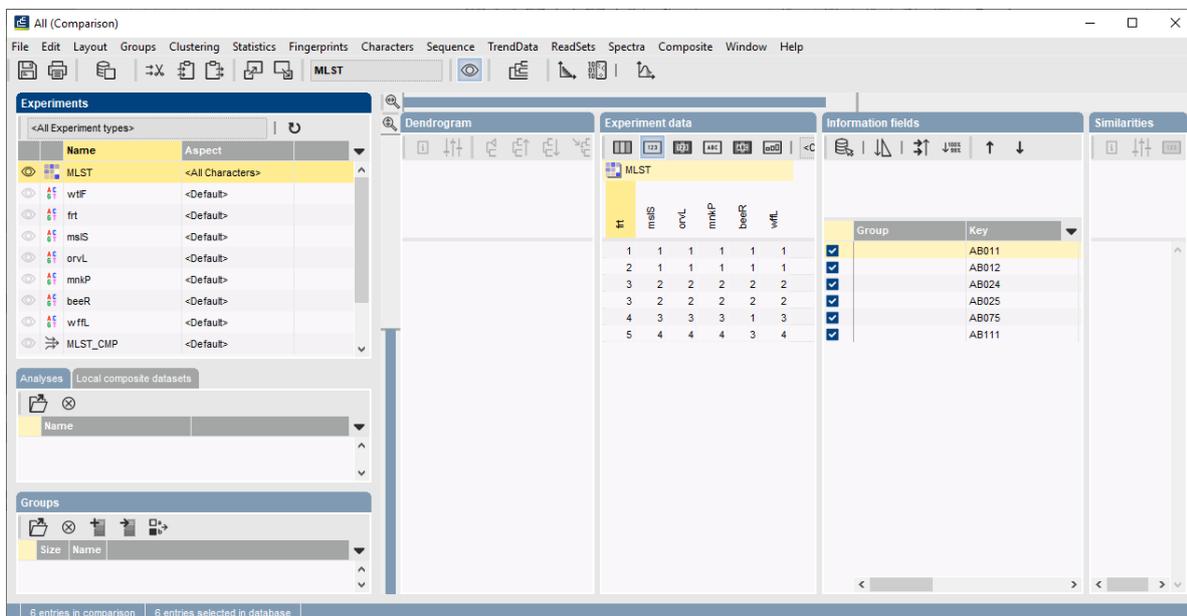


Figure 9: The *Comparison* window.

8 Sharing your MLST scheme

Once your database contains a certain amount of information, you might want to share your MLST schema with coworkers to ensure that a consistent nomenclature is used. To make this possible, the profiles and alleles files need to be exported and made available via the LAN or over the internet.

1. Select **MLST** > **Export alleles and profiles**.
2. Press <**Browse**> to change the export folder and select e.g. your desktop.
3. Export the seven FASTA-formatted allele files and the single profiles file with <**OK**>.

To make the MLST schema available to a colleague, the files should be copied to a location that he or she has access to. Depending on the situation, this might be on a file server in the LAN, the intranet or the internet.

During installation of the *MLST online plugin* in a new database that uses the same MLST schema, follow the steps as described in 4, but this time the option **Set profile and allele files/urls** in the *Define your own MLST scheme* dialog box needs to be checked. In the *Define profile and allele files* dialog box, the profiles file and an allele file for each of the loci can then be browsed for (see Figure 10).

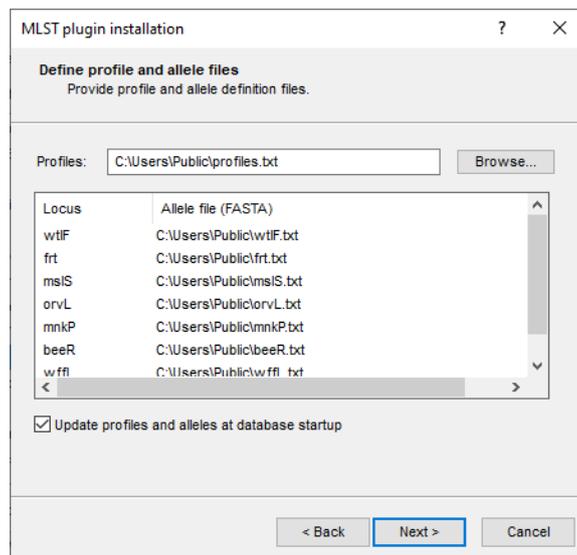


Figure 10: The *Define profile and allele files* dialog box.