

# **BIONUMERICS** Tutorial:

# Setup and analysis using a publicly available MLST scheme

# 1 Introduction

In this tutorial, we will illustrate the most common usage scenario of the *MLST online plugin*, i.e. when you want to perform MLST using an already published MLST schema, made available via <a href="http://pubmlst.org">http://pubmlst.org</a> or <a href="http://pubmlst.org">http://pubm

## 2 Creating a new database

To illustrate the complete setup of the *MLST online plugin* for using a publicly available MLST schema, we will start by creating a new, empty database.

- 1. Double-click on the BIONUMERICS icon (
- 2. In the *BIONUMERICS Startup* window, press the 🗟 button to enter the *New database* wizard.
- 3. Enter a database name, e.g. "Neisseria database".
- 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.

#### 3 Installing the MLST online plugin

In this section we will install the *MLST* online plugin in our database and we will set it up to use the publicly available *Neisseria* MLST schema.

- 1. The *Plugins* dialog box is called from the *Main* window by selecting *File* > *Install / remove plugins...* (, ).
- 2. Select the *MLST online plugin* from 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.

We are asked to select the organism source. In the next section we will import sequences from the organism *Neisseria* in our sample database. Since this organism has an MLST repository online we will use this online database to retrieve the allele, sequence type and clonal complex information for our sample data set.

4. Choose the option *Select organism from on-line list* and press <*Next*>.

Any organism for which an MLST repository is available online, is listed in the next step.

5.	Select	"Neisseria	spp." fro	om the l	ist and	press -	<next></next>	(see Fig	gure 1	).
								( •	<b>J</b> /	/

MLST plugin installation				?	×
Select an organism Select an organism	from the on-	-line	list.		
Mycoplasma bovis Mycoplasma hyopneum Mycoplasma hyorhinis Mycoplasma pneumonia Mycoplasma synoviae Neisseria spp. Orientia tsutsugamushi Ornithobacterium rhinoti Paenibacilius larvae Pasteurella multocida Pasteurella multocida Padiococcus pentosace <	e racheale	~	Organism: Neisseria spp. Loci: abcZ; adk; aroE; fumO pgm Database: http://rest.pubmist.org eisseria_seqdef contains 14368 profile	/db/pubmlst_n	
			< Back Next >	Cano	el

Figure 1: Select organism from the list.

It is possible to specify a location for the profile and allele definition files, which can be located on your own computer, local area network or on the internet. However, the default location should point to the correct files already. The option **Update profiles and alleles at database startup** can also be checked, to avoid having to do a manual update.

- 6. Press <*Next*> to continue.
- 7. In the next step, leave *Calculate trimming patterns automatically* checked and press <*Next*>.

In the final step, 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 with the installation of the *MLST online plugin*.



All remotely stored MLST information (allele numbering, sequence types and clonal complexes) for the selected organism will be downloaded in the BIONUMERICS database during the installation of the plugin. This might take several minutes. When querying for allele numbers, sequence types and clonal complexes in BIONUMERICS, this locally stored information will be used.

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 *MLST online plugin* installs menu items in the main menu of the software under *MLST* (see Figure 2). In the *Main* window, the *MLST online plugin* has installed following items:

- Extra information fields in the Database entries panel (default names: MLST ST; MLST CC).
- One character type called **MLST**, one composite dataset called **MLST\_CMP**, and seven sequence types, each named after a housekeeping gene.

File       Edit Database Analysis       MLST       Scripts       Window       Help							
Image: Second Secon	🖆 Neisseria MLST database - BioNu	umerics					- 🗆 ×
Sectionative region     Identify profile   Identify profile   Identify profile   Identify profile   Export alleles and profiles   Update letters and profiles <td>File Edit Database Analysis M</td> <td>ILST Scripts Window Help</td> <td></td> <td></td> <td></td> <td></td> <td></td>	File Edit Database Analysis M	ILST Scripts Window Help					
Seperiment types     Identify subles   Identify subles   Identify subles   Export lales and profiles   Update lates and profiles   Update		Identify alleles					
Identify alleles and profiles   Exponential eles and profiles Update alleles and profiles Update alleles and profiles Update into patterns Issue Issue Update into patterns Issue Issue Issue Issue Update into patterns Issue <		Identify profiles					
Export alleles and profiles   Update alleles and profiles   Update alleles and profiles   Update tim patterns   15   2 abc2   Update tim patterns   15   2 abc2   Update tim patterns   15   15   2 abc2   Update tim patterns <td></td> <td>Identify alleles and profiles</td> <td></td> <td>-</td> <td></td> <td></td> <td></td>		Identify alleles and profiles		-			
Image: Control of the second state       Image: Contro	⊗   +  ∆ ⊗  -		2' ⊗ €,   €	All Entries>	U	+ /∄ ⊗ 🖳 ∣ 🛍 🗠	<all comparisons=""></all>
I det time patterns I det time patterns I det time patterns I det stock MLST Search I of a stock I of a stock MLST CLUP Compose data sets I of a stock <			Level	Modified date 🗸	• <mark>1</mark> 2 3 4 5 6 7 8 9	Name Modified date	Level 🔻
1     3     ask     MLST search       2     4     8 rote       3     5     func       3     5     func       4     7       5     5       4     7       7     7       6     gala       5     5       6     gala       7     7       7     gala       7     7       8     gala       7     gala       8     gala       8     gala       9     MLST_CUP       Composite data sets       7     gala       8     gala       8     gala       8     gala       9     MLST_CUP       Composite data sets       7     gala       8     gala       8     gala       8     gala       8     gala       9     MLST_CUP       Composite data sets       7     gala       8     gala       8     gala       8     gala       9     Must state       9     Must state       9     Must state       9	1 MLST	Update alleles and profiles					
Image: Stand Control Security		Update trim patterns					
Fright field type Field type Field type Field type Curposed data sets Curposed data	3 adk	M CT search					
1     6     gpn     Seturce types       1     7     pdc     Sequence types       2     9     MLST_CMP     Composte data sets       2     >     4     0       1     8     pgn     Sequence types     1       2     >     4     0       2     >     4     0	5 fumC	WEST SERVICE.					
Image: Sequence types       Image: Sequence type       Image: Sequence types	G gdh	Settings				<	>
Image: Provide assembles     Automations						Identification projects Decision nativorka	
Image: Tropperson field date							
Cathy feelds     Didabases design       +     Processory       istance     Field type       istance     Field type       istance     Exact projects       Chromosome compatisons       istance       Experiment field type       istance       Experiment field       istance       Experiment field       istance       Experiment field       istance       Experiment field       istance       istance <td< td=""><td>□ ⇒ 9 MLST_CMP</td><td></td><td></td><td></td><td></td><td></td><td>All Identific</td></td<>	□ ⇒ 9 MLST_CMP						All Identific
Entry fields     Distabase design       +	/	Ň				Name Modified date	
Image: Field type     Image: Field type       Image: Field type		,					
Name     Field type       W     NLST ST       Field type     Algomenta       ILST CC     Fixed       Figerprot flex     Power assembles       Annotations     Algomenta       File name     Experiment type       Link       C     X	Entry fields Database design						
Hame     Field type       W     HLST ST       Field type     Image: Stress Stresst	+ 13 ⊗ 6.   6.						
44     MLST OC     Fixed       Fingerprint flex     Power assembles     Annotations       Image: Prover assembles     Image: Prover assembles       Image: Prover assembles     Image: Prover assembles       Image: Prove assembles     Image: Prove assembles       Image: Prove assembles     Image: Prove assembles       Image: Prove assembles     Image: Prove assembles       Image: Prove assembles <t< td=""><td>Name F</td><td>ield type 📃 🗸</td><td></td><td></td><td></td><td></td><td></td></t<>	Name F	ield type 📃 🗸					
Fingerprit files     Power assembles     Annolations       Image: Prover assembles     Annolations       Image: Prove assem	🗌 🕊 MLST ST 🛛 🛛 F	ixed A					
Fingerprint files     Power assembles     Annotatons       Image: Prover assembles     Annotatons       File name     Experiment type       Link       <	ANC MLST CC F	ixed				Alignments BLAST projects Chromosome com	parisons
Fingerprint files     Power assembles     Annotatons       Image: Prover assembles     Annotatons       File name     Experiment type       Link       <							<all alignments=""></all>
File name     Experiment type     Link        >		×					
File name     Experiment type     Link       C     >	Fingerprint files Power assemblies	Annotations					
File name     Experiment type     Link       C     >	月 + 乃 ⊗ 島」	All Fingerpr					
Database: Nelserie MLST database / DefaultUser ) Entres: Loaded-0. Vew=0. Selected=0 9 excerments C. Users/Public/Documents/Bit/Numerics/Data BNR/Nesseria MLST database / This is a time limited opc/sace valid until 2020-12-30	<			>	< >>		
	Database: Neisseria MLST database	( DefaultUser ) Entries: Loaded=0. View=0	Selected=0 9 experiments (	:\Users\Public\Documents\BioNumerics\Data	BN8Weisseria MLST database	This is a time limited package valid until 2020-12-30	

Figure 2: The *Main* window after installation of the MLST online plugin.

Upon installation of the plugin, the trimming patterns of *Neisseria* were automatically calculated. Since a lot of sequences exist for each housekeeping gene in the online repository, the trimming patterns of the *Neisseria* housekeeping genes contain a lot of degenerated positions. The presence of these degenerated positions might result in the assignment of the wrong trimming positions on our sample sequences. Less degenerated patterns can be found in the MLST trimming

patterns.txt file that can be downloaded from the Applied Maths website (https://www.applied-maths. com/download/sample-data, click on "MLST sample SCF trace files").

- 13. Select *MLST* > *Settings* to call the *MLST plugin settings* dialog box.
- 14. Click on the Trim patterns tab.
- 15. Uncheck *Calculate trimming patterns automatically* and copy the patterns from the MLST trimming patterns.txt file to the correct genes in the grid (see Figure 3).

.ST plugi	n settings					?	
rganism	Profiles/alleles	Trim patterns	Info fie	lds			
Locus	Start pattern	1	:	Stop patte	rn		
abcZ	TTYRAYAY	YRTNKSYRAAF	G	VHRNAW	HRTHRW	MHWHRAH	IC
adk	GARGCRAA	AAAAAWYAT	YGA	TYGAYG	GHACYC	RRSHDGT	2
aroE	DHTBRGNY	TRDYSRRYGA	A ·	TCVAAYA	YATRAR	AGHBHTR	
fumC	RRRDCHYT	RGGHRRRCRHO	BA I	DHAAAA	TYGGYT/	YGRRAA	Y
gdh	RTDTTYRA	RYCNHTDTKGR	A	TYGANRI	OYRRYYI	RHNCYKHF	۶
pdhC	ATRCYNRM	ARGYGYRRAR	SA	TGRBYD	NYTRDC	NRANCRR	
pgm	VTVDYBRC	CAMARASRRY	MW	YHDCNM	AHRCCVI	IRTTYDAV	/
Calcul	ate trimming patte	erns automatica	lly				
Trim	pattern length:	20					
Update trimming patterns at database startup							
					ОК		Cance

Figure 3: Trimming patterns after correction.

16. Close the *MLST plugin settings* dialog box.

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

#### 4 Importing and assembling trace files in batch

A set of *Neisseria* trace files can be downloaded from the Applied Maths website (https://www.applied-maths.com/download/sample-data, click on "MLST sample SCF trace files") and are used in this guide to explain the work flow of the *MLST online plugin*.

- 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. Select the *<Browse>* button, navigate to the correct path, select all the sequence trace files and press *<Open>*.

The *Import sequence traces* wizard page is updated (see Figure 4).

4. Press < *Next*> to go the next step.

The way the information should be imported in the database can be specified with an import template. In the example data set, the *Key* and *Sequence experiment name* are provided in the trace file name.

5. Select the predefined template *Example import 2* and press the *<Preview>* button to check the rules applied on the selected files (see Figure 5).

Import sequenc			?	×
Select the	data to import.			
Select file(s):	C:\Users\Public\Docum\\SO1_abcZ-3_F.scf C:\Users\Public\Docum\\SO1_abcZ-3_F.scf C:\Users\Public\Docum\\SO1_adk-3_F.scf C:\Users\Public\Docum\\SO1_adk-3_F.scf C:\Users\Public\Docu\\SO1_adk-3_R.scf C:\Users\Public\Docu\\SO1_arcE-15_F.scf C:\Users\Public\Docu\\SO1_arcE-15_R.scf C:\Users\Public\Docu\\SO1_fumC-2_F.scf C:\Users\Public\Docu\\SO1_fumC-2_R.scf C:\Users\Public\Docu\\SO1_fumC-2_R.scf C:\Users\Public\Docu\\SO1_gdh-9_F.scf C:\Users\Public\Docu\\SO1_gdh-9_R.scf C:\Users\Public\Docu\\SO1_gdh-9_R.scf C:\Users\Public\Docum\SO1_gdh-9_R.scf C:\Users\Public\Docum\SO1_gdh-9_R.scf C:\Users\Public\Docum\SO1_gdh-9_R.scf C:\Users\Public\Docum\SO1_gdh-9_R.scf C:\Users\Public\Docum\SO1_gdh-9_R.scf C:\Users\Public\Docum\SO1_g	Browse Delete Delete All		
Template file:		Browse		
	< Back	Next >	Car	icel

Figure 4: Select trace files.

revie	2W		?	1
Nr.	Sequence type	Кеу		_
1	abcZ-3	ISO1		
2	abcZ-3	ISO1		
3	adk-3	ISO1		
4	adk-3	ISO1		
5	aroE-15	ISO1		
6	aroE-15	ISO1		
7	fumC-2	ISO1		
8	fumC-2	ISO1		
9	gdh-9	ISO1		
10	gdh-9	ISO1		
11	pdhC-6	ISO1		
12	pdhC-6	ISO1		
13	pgm-9	ISO1		
14	pgm-9	ISO1		
15	abcZ-2	ISO2		
16	abcZ-2	ISO2		
17	adk-3	ISO2		
18	adk-3	ISO2		
19	aroE-4	ISO2		
			Clo	se



In the preview we see that the sequence type name is not correctly parsed from the file names. In a next step, we will modify the rule for the *Sequence type*.

- Make sure the predefined template *Example import 2* is still selected and press the <*Edit*> button to edit the rules.
- 7. Check the option *Show advanced options*, make sure the first row is selected in the grid panel and press the *<Edit parsing>* button.
- 8. In the Data parsing dialog box, update the data parsing string: "\*\_[DATA]-\*".

This parsing string will only take into account the text occurring between the first underscore (\_) and the hyphen (-). The asterisk (\*) serves as a wildcard, meaning that all characters before the first underscore and after the hyphen will be ignored.

9. Press the *Preview* button and press *OK* when the parsing is correct (see Figure 6).

Edit data p	arsing				?	×
Parse	component	find the comp	onent '[DA	TA]', use '*' as	wildcar	d
Regula	ar expressio	on: match the e	xpression	and use the s	subexpre	ssion
Data parsir	ng string:	*_[DATA]-*			~	
Data decor	ation:	[DATA]			~	
Preview						
Data:	ISO1_abo	:Z-3_F		Preview		
Output:	abcZ					
				ОК	Can	icel

Figure 6: Preview parsing.

- 10. Press <*Next*> and <*Finish*>.
- 11. Make sure the updated template *Example import 2* is selected and press the *<Preview>* button.

The preview should now look like Figure 7.

revie	w		? >
Nr.	Sequence type	Кеу	
1	abcZ	ISO1	
2	abcZ	ISO1	
3	adk	ISO1	
4	adk	ISO1	
5	aroE	ISO1	
6	aroE	ISO1	
7	fumC	ISO1	
8	fumC	ISO1	
9	gdh	ISO1	
10	gdh	ISO1	
11	pdhC	ISO1	
12	pdhC	ISO1	
13	pgm	ISO1	
14	pgm	ISO1	
15	abcZ	ISO2	
16	abcZ	ISO2	
17	adk	ISO2	
18	adk	ISO2	
19	aroE	ISO2	1

Figure 7: Preview of the parsing.

- 12. Close the preview.
- 13. Make sure the updated template *Example import 2* is selected and press < *Next*>.
- 14. Press < *Next* > to confirm the creation of 3 new entries (see Figure 8).

The *Processing* wizard page opens (see Figure 9).

In the *Reports panel*, the *Maximum# of unresolved bases reported* can be specified (default value 20). Likewise, the *Maximum # of align inconsistencies reported* can be entered (default value 20). Align inconsistencies are positions where the consensus is resolved, but where one or more sequences are different from the consensus.

15. Press < *Trimming settings*> to pop up the *Assembly trimming settings* dialog box.

Import sequence trace files	?	×
Database links Link the imported records to database entries. Double click on a cell to get an overview.		
Overview         In 'All levels'       Image: create 3 entries         Image: create 3 entries       and image: update 0 entries         Image: create 3 entries       Image: create 3 entries         Image: create 3 entries       <		
O Update existing assemblies		
< Back Next >	Car	ncel

Figure 8: Database links.

		?	
< Back	Finish	Can	
	< Back	< Back Finish	< Back Finish Can

Figure 9: The *Processing* wizard page.

The trimming patterns entered in the *Trim patterns tab* (see Figure 3) are shown in the *Start pattern* and *Stop pattern* columns (see Figure 10).

Double-click on the <*Edit*> button for experiment *abcZ* to call the *Assembly trimming settings* dialog box (see Figure 11).

In the *Assembly trimming settings* dialog box, a number of additional settings can be specified for each individual sequence experiment:

- Minimum # of sequences specifies the minimum number of trace sequences that should contribute to the subsequence on the consensus that matches the trimming targets. For example, if "2" is entered, a trimming target will only be set if the matching region on the consensus is *fully* defined by at least 2 sequences.
- For both the *Start position* and *Stop position*, a *Trim pattern* is displayed. The use of IUPAC code for ambiguous positions is supported. The *Tolerance* defines the number of

Experiment	Start pattern	Stop pattern	
abcZ	TTYRAYAYYRTNKSYRAARG	VHRNAWHRTHRWMHWHRAHC	Edit
adk	GARGCRAAAAAAWYATYGA	TYGAYGGHACYCRRSHDGTR	Edit
aroE	DHTBRGNYTRDYSRRYGAYA	TCVAAYAYATRARAGHBHTR	Edit
fumC	RRRDCHYTRGGHRRRCRHGA	DHAAAATYGGYTAYGRRAAY	Edit
gdh	RTDTTYRARYCNHTDTKGRA	TYGANRDYRRYYRHNCYKHR	Edit
pdhC	ATRCYNRMARGYGYRRARSA	TGRBYDCNYTRDCNRANCRR	Edit
pgm	VTVDYBRCCAMARASRRYMW	YHDCNMAHRCCVMRTTYDAV	Edit
ogm	VTVDYBRCCAMARASRRYMW	YHDCNMAHRCCVMRTTYDAV	Edit

Figure 10: The Assembly trimming settings dialog box.

Assemb	ly trimming settings for 'abcZ'			?	×			
Minimum number of sequences contributing to consensus:								
	Trim pattern	Tolerance	Offset Sea	irch range				
Start:	TTYRAYAYYRTNKSYRAARG	0	0	-				
Stop:	VHRNAWHRTHRWMHWHRAHC	0	0	-				
			ОК	(	Cancel			

Figure 11: The Assembly trimming settings dialog box.

mismatches allowed for a sequence to be recognized as a trim pattern. With the *Offset*, one can specify that the consensus is trimmed at a certain offset from the start and end trimming target positions. If no offset is specified (zero), the trimming targets are included in the trimmed consensus. With the *Search range* one can restrict the search to certain regions on the consensus, e.g. to prevent incidental matches inside the targeted consensus sequence.

The entered trim patterns will be searched on the consensus sequence in both directions, i.e. on the consensus as it appears as well as on its complementary strand. In case the trim patterns match the complementary strand of the consensus, it will be automatically invert-complemented. If the *Trim pattern* text boxes are left empty, no preference sense is available.

- 17. Leave the predefined settings unaltered and press < *OK* > and < *Close* > to close dialog boxes.
- Press the <*Assembly settings*> button to call the *Assembly settings* dialog box (see Figure 12).
- Double-click on the <*Edit*> button for experiment *abcZ* to call the *Assembly settings* dialog box (see Figure 13).

The Assembly settings are grouped in tabs per settings dialog box in *Assembler*: **Quality** assignment, **Assembly** and **Consensus** determination. For a detailed description of the Assembler program settings, see the BIONUMERICS manual. In the last tab the Assembly settings can be copied from or to another sequence type experiment.

- 20. For this exercise, do not change the settings and press *<OK>* and *<Close>*.
- 21. Make sure the option *Open assembly overview report* is checked and press <*Finish*> to assemble the selected trace files from the example dataset into separate contig projects.

Experiment		
abcZ	Edit	
adk	Edit	
aroE	Edit	
fumC	Edit	
gdh	Edit	
pdhC	Edit	
pgm	Edit	

Figure 12: The Assembly settings dialog box.

Assembly settings for 'abcZ'	7	?	×
Quality Assembly Consensus Copy settings			
Curve sliding window: 5	bp		
Minimum good/bad peak ratio: 1.30			
Minimum short/long peak distance ratio: 0.50			
Base calling sliding window: 41	bp		
Minimum resolved positions: 30	bp		
Minimum consecutive good bases: 15	bp		
Minimum length of usable sequence: 50	bp		
Minimum fraction of good bases: 25	%		
Reset to default OK		Cancel	I

Figure 13: The Assembly settings dialog box.

## 5 Checking the sequence assemblies

When the assemblies are processed, an interactive report window appears (see Figure 14). This window can also be displayed from the *Main* window with *Analysis* > *Sequence types* > *Batch assembly reports...*.

ć	Assembly	report: Batch	- 2020-04-20	10:12:08							-		Х
ile	Overview	/ Details \	Window Hel	lp									
Over	rview								Details				
	]								$\approx$				
	Кеу	abcZ	adk	aroE	fumC	gdh	pdhC	pgm	Code	Message	Status	Con	nme
~	ISO1	warning	ОК	ок	warning	warning	warning	error	warning	Inconsistency in alignment at position 551 in sequence(s) 1	new		
~	ISO2	ОК	ОК	error	error	warning	error	ок	warning	Inconsistency in alignment at position 550 in sequence(s) 1	new		
<b>~</b>	ISO3	ОК	ок	error	error	error	ок	error	warning	Inconsistency in alignment at position 546 in sequence(s) 1	new		
									warning	Inconsistency in alignment at position 541 in sequence(s) 1	new		
									warning	Inconsistency in alignment at position 525 in sequence(s) 1	new		
									warning	Inconsistency in alignment at position 507 in sequence(s) 1	new		
									info	Created new assembly			
										Report for ISO1 / abcZ			
										<			
										<			

Figure 14: The Batch sequence assembly report window.

The *Overview* panel displays the entries (keys) as rows and the experiments as columns. Each cell in the grid, corresponding to a key/experiment pair, provides information about the current

status of the contig project. Only for those entries that have a green (= **OK** or **Solved**) or orange (= **Warning**) status, the allele IDs can be assigned.

- 1. Click a cell, e.g. *ISO1/pgm* to update the *Details* panel on the right-hand side.
- 2. In the *Details* panel double-click on the first message.

This will open the sequence in the *Contig assembly* window, with the corresponding position in focus. The position can now be examined and - if needed - the base calling can be changed manually. In the Assembler project of the pgm of ISO1, three unresolved bases are present in the consensus sequence. For each of these unresolved bases, an error is listed in the *Details* panel.

- 3. In the *Contig assembly* window click on the *Aligned traces* panel.
- 4. To obtain an optimal view of the curves, use the zoom sliders in the *Traces* panel or use the zoom buttons.

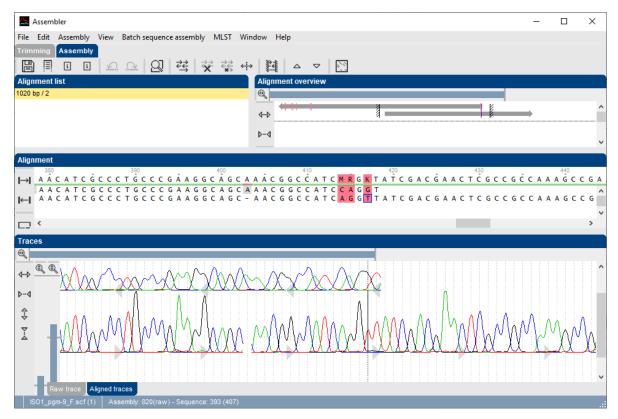


Figure 15: The Aligned traces panel.

The three unresolved bases in the consensus sequences are displayed in pink. The consensus sequence can be screened for (nearest) allele matches:

5. In the Assembly view, select **MLST** > **Identify allele**.

The consensus sequence is screened against the downloaded allele information of the selected organism.

The *MLST identification plot window* pops up. The sequence type of the sequence that is shown in Assembler is displayed in white. The other sequence types are shown in gray (see Figure 16). The identification result is shown below the sequence type boxes.

6. In our example, four mismatches are reported with allele ID 9 (see Figure 16).

🖆 ISO1: MI	LST identification	i				-		×
File View	Window Help							
Q Plot								
								^
	Entry ISO1	I: MLST identific	ation					
							1	
	abcZ	adk aroE	fumC	gdh	pdhC	pgm		
	abcZ pgm:	adk aroE	fumC Difference			pgm		
		adk aroE Mismatches		s: Allele		pgm		
	pgm:		Difference	s: Allele	ID 9	pgm		
r.	pgm: Allele ID	Mismatches	Difference Assembly	s: Allele Contig	ID 9 Allele	pgm		
	pgm: Allele ID 9	Mismatches	Difference Assembly 417	s: Allele Contig K	ID 9 Allele T	pgm		
[	pgm: Allele ID 9 32	Mismatches 4 5	Difference Assembly 417 415	S: Allele Contig K R	ID 9 Allele T G	pgm		
[	pgm: Allele ID 9 32 70	Mismatches 4 5 5	Difference Assembly 417 415 414	s: Allele Contig K R M	ID 9 Allele T G A	pgm		

Figure 16: The MLST identification plot window: Best match with allele ID 9.

7. Select the first mismatch from the list.

The focus in Assembler is updated.

- 8. Change the G at position 820 in the reverse sequence to a T. The consensus sequence at position 820 is automatically updated.
- Update the consensus sequence according to other editing suggestions: enter G at position 818, enter A at position 817, and remove A at position 806. Save the contig project with *File* > *Save* (E, Ctrl+S).
- Select *MLST* > *Identify allele*. The consensus should now have a perfect match with allele ID 9.
- 11. Select *Batch sequence assembly* > *Set report to solved, save and close* (Ctrl+Shift+S) in the *Contig assembly* window.

The corresponding key/experiment cell in the *Overview* panel is updated and displayed in green. The status "solved" is displayed in the cell and in the *Status* column of the *Details* panel.

	embly rep												- -
File Ov			2020-04-20 10:1	2:08								- 1	⊐ ×
	rview	Details Wir	ndow Help										
Overviev	"									Details			
										$\approx$			
Key	ai	ibcZ	adk	aroE	fumC	gdh	pdhC	pgm	-	Code	Message	Status	Comme
ISO1	1 w	warning	ок	ок	warning	warning	warning	solved		error	Unresolved base in consensus at position 820	solved	
ISO2	2 0	ж	ок	error	error	warning	error	ок		error	Unresolved base in consensus at position 818	solved	
IS03	з о	Ж	ок	error	error	error	ок	error		error	Unresolved base in consensus at position 817	solved	
										warning	Inconsistency in alignment at position 806 in sequence(s) 1	solved	
										info	Created new assembly		
											Report for ISO1 / pgm		
											<		



- 12. Click on the **ISO1**/abcZ warning cell in the *Overview* panel to update the *Details* panel on the right-hand side.
- 13. In the *Details* panel double-click on the first warning message to open the *Contig assembly* window.

Six alignment inconsistencies are reported. False peaks in the trace files result in the insertion of six false bases in the consensus sequence. These bases can be deleted from the consensus sequence as follows:

- 14. Click on the first alignment inconsistency in the *Details* panel. The focus is automatically updated.
- 15. Press the **Del**-key on the keyboard to delete the base from the consensus sequence.
- 16. Repeat this step for the other alignment inconsistencies.
- 17. Save the contig project with *File* > *Save* (B, Ctrl+S).
- Select *MLST* > *Identify allele*. The consensus should now have a perfect match with allele ID 3.
- 19. Select *Batch sequence assembly* > *Set report to solved, save and close* (Ctrl+Shift+S) in the *Contig assembly* window.
- 20. Check all other warnings for ISO 1. In the *Assembly view*, select *MLST* > *Identify allele* to get an idea of the (closest) matches.
- 21. As an extra exercise, check the errors and warnings of the ISO 2 and ISO 3 samples.

#### 6 Identifying alleles and profiles

After all sequence assemblies are checked, we can use the *MLST* online plugin for the identification of our sequences.

- 1. Make sure all three entries in the database are selected.
- 2. In the Main window, 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:

3. 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 18).

abcZ         3         <+>            adk         3         <+>             aroE         15         <+>             fumC         2         <+>             gdh         9         <+>             pdhC         6         <+>	adk     3 <+>       aroE     15 <+>       fumC     2 <+>       gdh     9 <+>	adk     3 <+>       aroE     15 <+>       fumC     2 <+>       gdh     9 <+>       pdhC     6 <+>	Character	Value	Mapping	•
aroE         15 <+>           fumC         2 <+>           gdh         9 <+>           pdhC         6 <+>	aroE         15 <+>           fumC         2 <+>           gdh         9 <+>           pdhC         6 <+>	aroE         15 <+>           fumC         2 <+>           gdh         9 <+>           pdhC         6 <+>	abcZ	3	<+>	
fumC         2 <+>           gdh         9 <+>           pdhC         6 <+>	fumC         2 <+>           gdh         9 <+>           pdhC         6 <+>	fumC         2 <+>           gdh         9 <+>           pdhC         6 <+>	adk	3	<+>	
gdh 9 <+> pdhC 6 <+>	gdh 9 <+> pdhC 6 <+>	gdh 9 <+> pdhC 6 <+>	aroE	15	<+>	
pdhC 6 <+>	pdhC 6 <+>	pdhC 6 <+>	fumC	2	<+>	
-	-	-	gdh	S	) <+>	
0.00	pgm 9 <+>	pgm 9 <+>	pdhC	e	i <+>	
pgin 5 <+>			pgm	9	<+>	



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

Next we will screen the allelic profiles of the entries against the sequence type and clonal complex information of *Neisseria*.

5. In the *Main* window, select *MLST* > *Identify profiles*.

The matched sequences types and clonal complexes are displayed in the MLST information fields "MLST ST" and "MLST CC", respectively.