SequenceAlignment
User's Guide

Elisa Del Fabbro, Paolo Magni

Università degli Studi di Pavia
Pavia, Italy

SequenceAlignment is a simple Matlab tool expressly designed for the quality control check of new BioBrick. It allows the user to align the results of the sequencing procedure with the aspected sequences built up assembling simple parts according to the rules of Standard Assembly (modality *Compare Sequence/BioBrick*).

It offers also the possibility of simply align two sequences (modality *Compare two Sequences*).

SequenceAlignment executable files for Windows platforms can be downloaded free of charge for non-profit institutions from the following web site: http://aimed11.unipv.it/SequenceAlignment.

SequenceAlignment - a simple example

Modality *Compare Sequence/BioBrick*

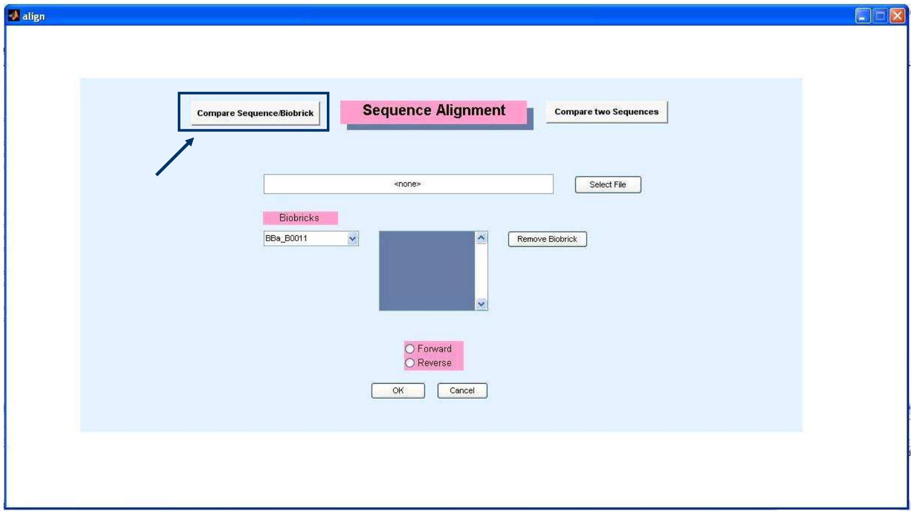


Figure 1: SequenceAlignment main page, modality of use *Compare Sequence/BioBricks*.

1. Choose the file containing the result of the sequencing by browsing the file system clicking the Select File button. Only FASTA files are allowed.

2. Select from the *Biobricks*[1] menu the parts, in order to obtain the expected sequence of the new BioBrick assembled by the tool according to the rules of Standard Assembly. The selected part can be removed by clicking the Remove Biobrick button.
3. Select the type of sequencing you are analyzing by choosing between the Forward and Reverse radiobuttons.
4. Click the OK button to start the alignment algorithm and to produce the corrispondent report.

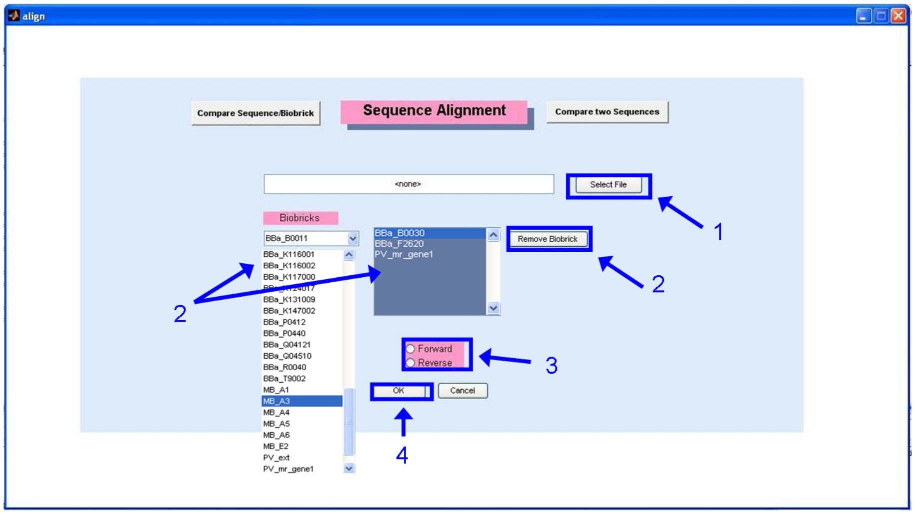


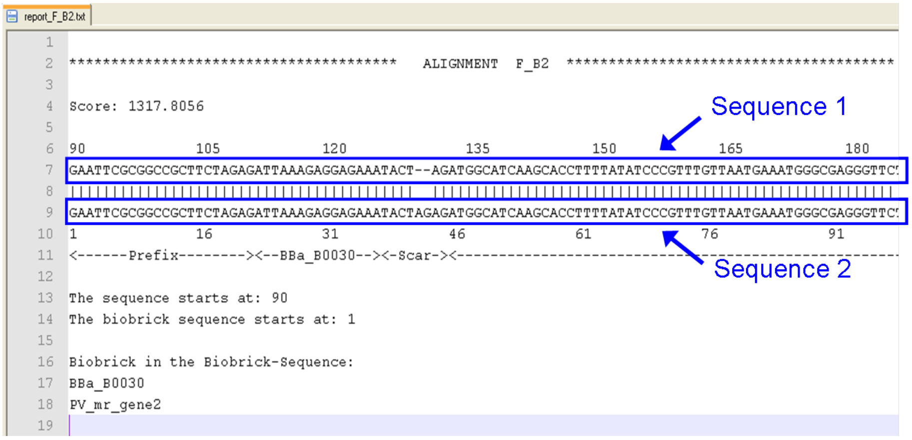
Figure 2: SequenceAlignment, modality of use *Compare Sequence/BioBricks*, example of use.

The software proceeds connecting to the Internet and searching the sequences of the selected BioBricks in the *Registry of Standard Biological Parts* website (http://partsregistry.org/Main_Page). Then, it assembles to create the expected sequence (including prefix, suffix and scars between the parts). The algorithm used for the alignment is the Smith-Waterman, provided by the Matlab Bioinformatics Toolbox.

The report created by the tool is showed in Fig.3: it is a text file and it is automatically opened with the text editor defined by the user in the file *config.txt*.

---

[1] See the section *How to create the file "lista Bricks.xls"* for more details.



```
****************************************   ALIGNMENT  F_B2  ****************************************

Score: 1317.8056

90             105            120            135            150            165            180
GAATTCGCGGCCGCTTCTAGAGATTAAAGAGGAGAAATACT--AGATGGCATCAAGCACCTTTTATATCCCGTTTGTTAATGAAATGGGCGAGGGTTC
|||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
GAATTCGCGGCCGCTTCTAGAGATTAAAGAGGAGAAATACTAGAGATGGCATCAAGCACCTTTTATATCCCGTTTGTTAATGAAATGGGCGAGGGTTC
1              16             31             46             61             76             91
<------Prefix--------><--BBa_B0030--><-Scar-><-----------------------------------------------------

The sequence starts at: 90
The biobrick sequence starts at: 1

Biobrick in the Biobrick-Sequence:
BBa_B0030
PV_mr_gene2
```

Figure 3: SequenceAlignment, modality of use *Compare Sequence/BioBricks*, example of report.

The sequence 1 is that loaded by the user, the sequence 2 is the expected one.

Useful labels identify the subparts recognized in the sequence by the alignment (BioBricks, external sequences, typical parts of the Standard Assembly), the number allow an easy localization of the parts in the two sequences.

Modality *Compare two Sequences*

1. Choose the files to be aligned (containing the results of the sequencing procedure) by browsing the file system clicking Select File button. Only FASTA files are admitted.
2. Select the type of sequencing analysed for both sequences by choosing between the Forward and Reverse radiobuttons.
3. Click the OK button to start the alignment algorithm and to produce the corrispondent report.

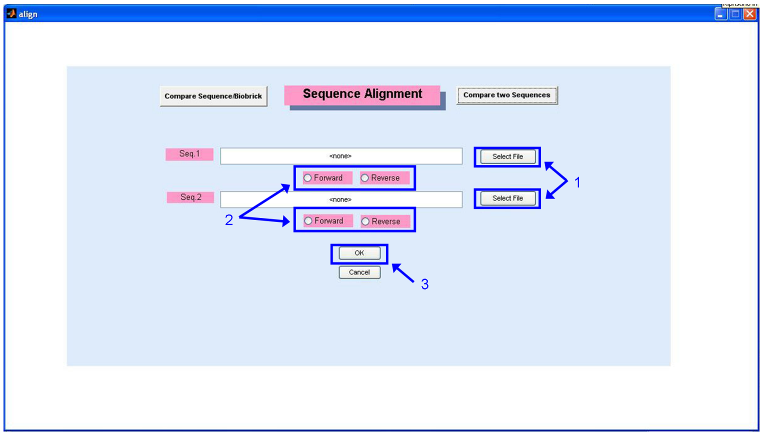
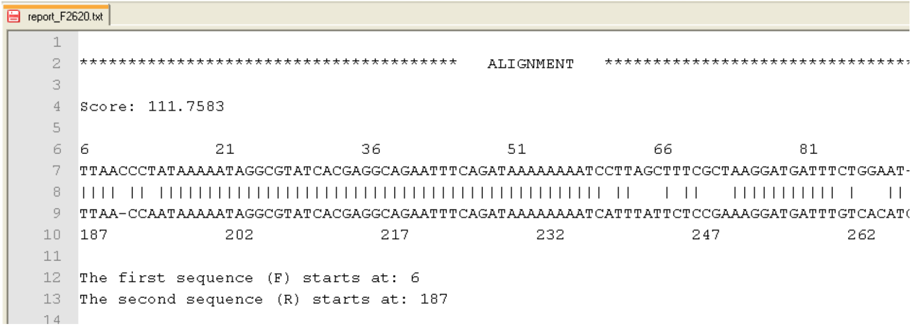


Figure 4: SequenceAlignment, modality of use *Compare two Sequences*, example of use.

The report created by the tool is showed in Fig.5: it is a text file and is automatically opened with the editor defined by the user in the file *config.txt;* it contains simply the alignment of the two sequences.



```
***************************************   ALIGNMENT   ********************************

Score: 111.7583

6              21             36             51             66             81
TTAACCCTATAAAAATAGGCGTATCACGAGGCAGAATTTCAGATAAAAAAAATCCTTAGCTTTCGCTAAGGATGATTTCTGGAAT-
|||| || |||||||||||||||||||||||||||||||||||||||||||||| ||   | ||   ||||||||||| |   ||
TTAA-CCAATAAAAATAGGCGTATCACGAGGCAGAATTTCAGATAAAAAAAATCATTTATTCTCCGAAAGGATGATTTGTCACAT(
187            202            217            232            247            262

The first sequence (F) starts at: 6
The second sequence (R) starts at: 187
```

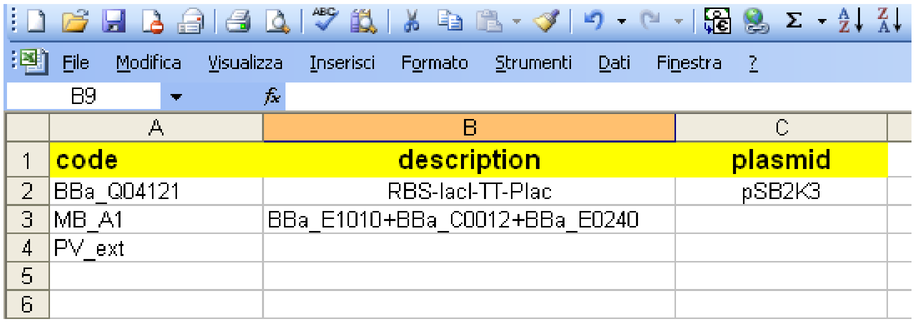
Figure 5: SequenceAlignment, modality of use *Compare two Sequences*, example of report.

How to create the file “BioBrick_list.xls”

In this file you have to define the BioBricks involved in your project.

It offers also the possibility of import sequences, different from the standard parts present in the *Registry of Standard Biological Parts.*

You can even define and name your assembled parts.



	A	B	C
1	**code**	**description**	**plasmid**
2	BBa_Q04121	RBS-lacI-TT-Plac	pSB2K3
3	MB_A1	BBa_E1010+BBa_C0012+BBa_E0240	
4	PV_ext		
5			
6			

Figure 7: layout file *BioBrick_list.xls,* an example.

The “Code” rules

- BBa_*xxx*: standard name of a single BioBrick present in the *Registry of Standard Biological Parts*, the other attributes are completely free (row 1 in Fig.5).
- MBa_*myname*: assembled BioBricks defined by the user in the attribute description, which is composed by the standard names of the BioBricks, in the correct order, separated by a “+” without spaces (row 2 in Fig.5).

- myfile: name of an external file “.txt”, saved in the working directory, conteining a sequence of interest not present in the *Registry of Standard Biological Parts* (row 3 in Fig.6). The layout of that file must be the following one (the sequence must be written as a single string, without any spaces inside):

  >NAME

  AAATTTTGCATACGG