User manual
AlignMiner

Darío Guerrero and Rocío Bautista and M. Gonzalo Claros

Plataforma Andaluza de Bioinformática

Universidad de Málaga

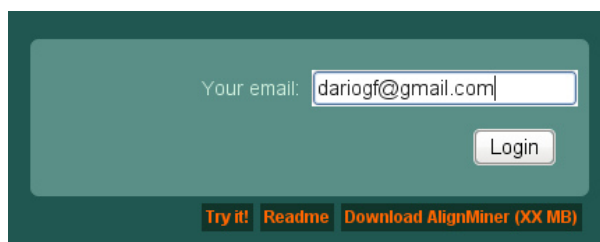Severo Ochoa, 34

29590 Málaga, Spain

September 23, 2009

1   Introduction

This document is only a quick guide to AlignMiner, if you need an in-depth explanation you should read the article "XXXXX".

1.1   Login in the system

The first action you need to do in order to use Alignminer is to do a login in the system. Enter any valid email in the form of the main page, as seen in the following picture:
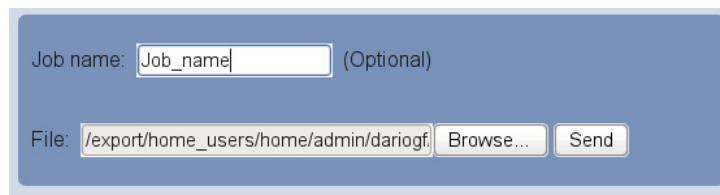

Once logged in, you can choose between submitting new jobs, or browse already sent ones.
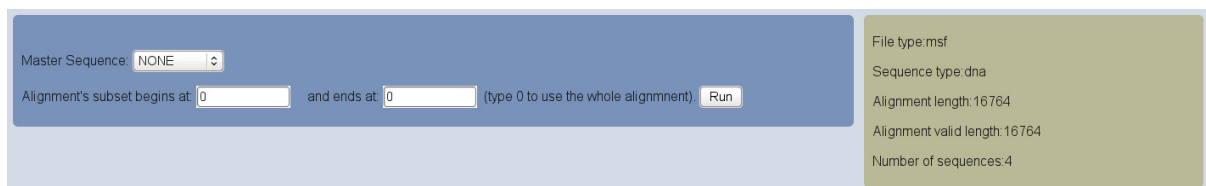
1.2   Submit a new job

To submit a new job for execution, you need to upload a file that contains a valid alignment. AlignMiner supports a wide variety of formats (fasta, clustal, msf, ...), so you can try to upload your alignment files without any conversion. Select the alignment to upload using the file field.

You may also provide a name for the job. If not an automatically generated number is assigned.

Click **send** button to submit the file.

When the file has been successfully uploaded to our servers, you will see a new form (and some useful information) where you can select some parameters:

- Master sequence: you can select a master sequence from the popup to be used by scoring methods. If you don't select a master sequence, then a consensus of the whole alignment is calculated.

- Alignment subset: with this parameter you can force AlignMiner to use only a portion of the alignment. If you leave both fields with zeroes, then an automatic range will be calculated, and weak ends (those with little information) of the alignment will be cropped.

Click **run** when you have modified the desired parameters. You have just sent the job for execution and the job should appear in the job list:

JOB ID	File name	Master seq.	File type	Seq. type	Length	N. Seqs.	Result size	STEP	STATUS	
job1	alinea.txt	NONE	msf	dna	16762 of 16764	4	2.1M	DONE (8/8)	DONE	
test	alignment.file	NONE	fasta	protein	1616 of 1630	3	388K	DONE (8/8)	DONE	
122185311449270	clustal	NONE	clustalw	protein	384	3	224K	DONE (8/8)	DONE	
121970766616315	fasta	NONE	fasta	protein	384	3	224K	DONE (8/8)	DONE	
121473530531072	alineamientoClustal.msf	NONE	msf	dna	16764	4	1.7M	DONE (8/8)	DONE	

Actually, the job is run in our supercomputers using a batch queue system, because of that you can view different status of execution in the job list:

- Waiting run: the job has just been uploaded and is waiting you to click the RUN button.

- Queued: the job has been sent to the batch system and is waiting for resources to be available for execution.

- Running: the job is running. In the STEP column you can see the completed phases.

- Done: the job is done and available for browsing.

- Error: there has been an error that prevents the job from executing correctly. Check the small info icon for more information.

Once a job is submitted you can close the system and come later to see its status, or wait it to finish.
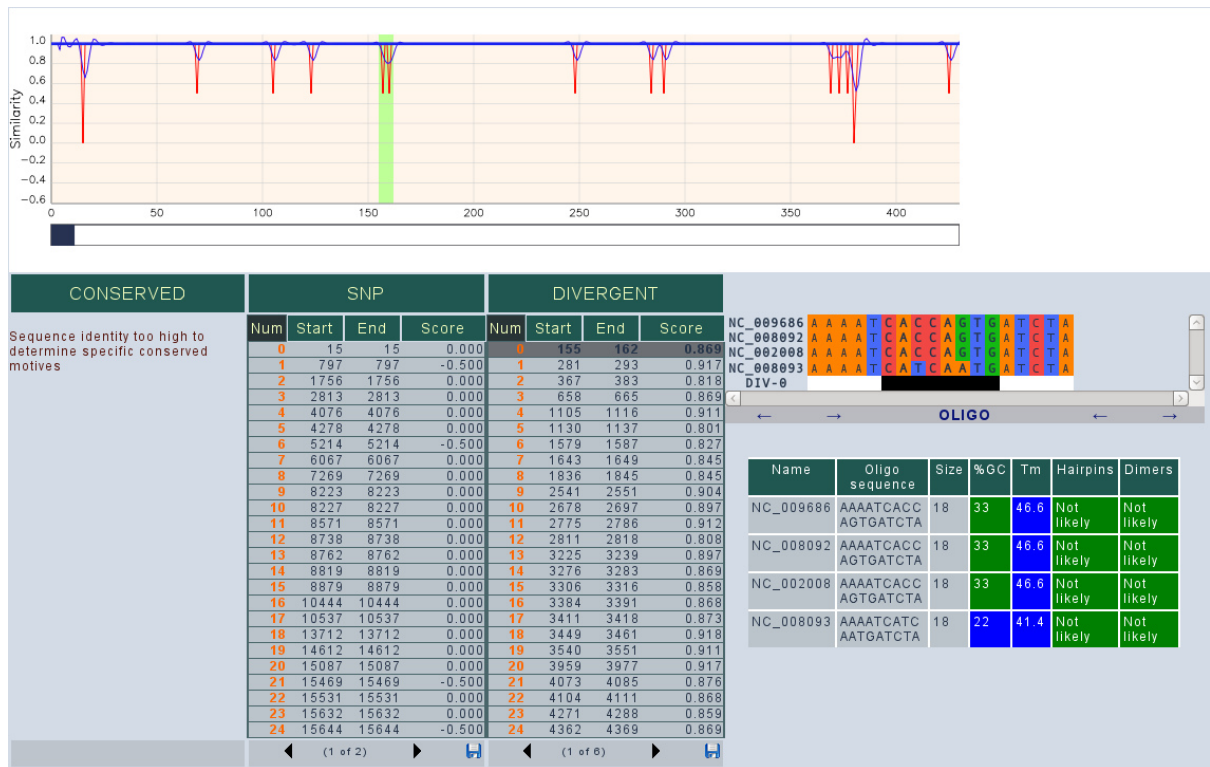
1.3   Browse existing jobs

Jobs are saved for later browsing, so you can access the system at any time and choose a job from the job list to browse its results.

If you click on a job name, you get a list of available scoring methods:

When a scoring method is selected, a graph is shown with the visual representation of the results. You can drag, zoom or click the graph to your desire:

CONSERVED
Sequence identity too high to determine specific conserved motives

SNP

Num	Start	End	Score
0	15	15	0.000
1	797	797	-0.500
2	1756	1756	0.000
3	2813	2813	0.000
4	4076	4076	0.000
5	4278	4278	0.000
6	5214	5214	-0.500
7	6067	6067	0.000
8	7269	7269	0.000
9	8223	8223	0.000
10	8227	8227	0.000
11	8571	8571	0.000
12	8738	8738	0.000
13	8762	8762	0.000
14	8819	8819	0.000
15	8879	8879	0.000
16	10444	10444	0.000
17	10537	10537	0.000
18	13712	13712	0.000
19	14612	14612	0.000
20	15087	15087	0.000
21	15469	15469	-0.500
22	15531	15531	0.000
23	15632	15632	0.000
24	15644	15644	-0.500

◄ (1 of 2) ►

DIVERGENT

Num	Start	End	Score
0	155	162	0.869
1	281	293	0.917
2	367	383	0.818
3	658	665	0.869
4	1105	1116	0.911
5	1130	1137	0.801
6	1579	1587	0.827
7	1643	1649	0.845
8	1836	1845	0.845
9	2541	2551	0.904
10	2678	2697	0.897
11	2775	2786	0.912
12	2811	2818	0.808
13	3225	3239	0.897
14	3276	3283	0.869
15	3306	3316	0.858
16	3384	3391	0.868
17	3411	3418	0.873
18	3449	3461	0.918
19	3540	3551	0.911
20	3959	3977	0.917
21	4073	4085	0.876
22	4104	4111	0.868
23	4271	4288	0.859
24	4362	4369	0.869

◄ (1 of 6) ►

```
NC_009686  A A A A T C A C C A G T G A T C T A
NC_008092  A A A A T C A C C A G T G A T C T A
NC_002008  A A A A T C A C C A G T G A T C T A
NC_008093  A A A A T C A T C A A T G A T C T A
DIV-0
```

← → **OLIGO** ← →

Name	Oligo sequence	Size	%GC	Tm	Hairpins	Dimers
NC_009686	AAAATCACCAGTGATCTA	18	33	46.6	Not likely	Not likely
NC_008092	AAAATCACCAGTGATCTA	18	33	46.6	Not likely	Not likely
NC_002008	AAAATCACCAGTGATCTA	18	33	46.6	Not likely	Not likely
NC_008093	AAAATCATCAATGATCTA	18	22	41.4	Not likely	Not likely

Below the graph, the interesting regions that have been found are presented in tables. You can sort the results by clicking on the column titles. When you click on a row (a region of interest), its corresponding alignment is show at the right of the page. In DNA alignments, characteristics of a putative oligo that uses this region are also calculated. You can change the oligo's width by clicking on the arrows, and its characteristics will be calculated again for you.