

# Alignancer User's Guide

v0.6; 2012-09-07

Copyright © 2012 Fabian Kilpert

[Bioinformatics at AWI](#)

Alignancer is a tool for alignment enhancing. It is performing additional alignments on masked sections of hmalign (HMMER 3.0) multiple alignments. ClustalW (default) or Muscle is being utilized to automatically align all sections which were not covered by a profile HMM and therefore were not aligned. In using additional aligners to complement hmalign, compulsory loss of information may be avoided keeping the entire alignment for subsequent analyses, e.g. the calculation of phylogenetic trees.

Just to recall from the [HMMER User's Guide](#) (Eddy 2010): hmalign uses upper case letters and '-' (dash) characters for aligned columns and lower case letters and '.' (period) characters for assigned columns! The latter do not match to a HMM profile and just contain unaligned nucleotides. Instead of real gaps ('-' deletions with cost), '.' characters are just for padding and have no cost at all. In the majority of cases unaligned columns are discarded in order to result useful alignments.

But having no match state in a HMM profile does not necessarily mean that there is no useful phylogenetic information at all. Here is where Alignancer comes in, taking those columns and aligning them with alternative aligners, i.e. ClustalW (Larkin 2007) and Muscle (Edgar 2004). All columns already aligned by HMM profile remain untouched.

It is important to note that an on top alignment with Alignancer – though preserving potentially more information – must not always result in a “better” alignment. However, as Alignancer focuses on less conserved columns it complements the profile alignment very well.

## License and availability

Alignancer is released under GNU GPL v3. <http://www.gnu.org/licenses/>

The software is downloadable from the AWI web-site. [www.awi.de/en/go/bioinformatics](http://www.awi.de/en/go/bioinformatics)

## Mode of operation

Let's say you just produced an alignment with hmmlalign:

```
hmmlalign --trim --outformat afa --allcol --mapali ref.sto
-o example.fas ref.hmm3 queries.fas
```

Then Alignancer is run:

```
python alignancer.py -i example.fas
```

Alignancer now reads in the input FASTA file, in this case example.fas. It determines the nucleotides and builds 'blocks' of subsequently aligned or assigned (unaligned) columns. Those columns containing lower case letters and '.' (period) characters are potential candidates for an on top alignment. Blocks of at least 2 column lengths (and with nucleotides in at least two sequences) are aligned with ClustalW or Muscle and replace the initial unaligned block. The output file of Alignancer combines all unchanged blocks with the aligned blocks.

## Requirements

Alignancer is developed for the Unix shell and was tested with Python 2.6.5 on Ubuntu 10.04 LTS. External aligners CLUSTAL 2.0.10 and MUSCLE v3.7 are supported and must be installed beforehand, best configured for being used by the path variable.

## Installation

Download the alignancer.tar.gz file and extract its content to your folder of choice.

## Usage

Help (show options):

```
python alignancer.py -h
```

Run Alignancer on an hmmlalign alignment (default aligner is ClustalW):

```
python alignancer.py -i example.fas
```

Using Muscle:

```
python alignancer.py -i example.fas -a muscle
```

If there is no appropriate path variable set for the aligner, the path has to be specified on the command line as follows:

```
python alignancer.py -i example.fas -p /my/path/to/clustalw
```

```
python alignancer.py -i example.fas -a muscle -p /my/path/to/muscle
```

Alignancer is running the external aligners with their default options. If you prefer optimized settings use Alignancer's `--optstr` option. The option string itself must contain valid [ClustalW](#) or [Muscle](#) options of course, e.g., saving the ClustaW log to file:

```
python alignancer.py -i example.fas --optstr "--stats=clustalw.log"
```

By default all aligned columns (using ClustalW and Muscle) have lower case letters in the output file. This allows inspecting the new alignment columns more easily. If you don't need that you can output upper case nucleotides only:

```
python alignancer.py -i example.fas -u
```

Some more output, showing e.g. the command line called to run the external aligner:

```
python alignancer.py -i example.fas -v
```

Using parallel processing (e.g. 5 threads):

```
python alignancer.py -i example.fas -t 5
```

Showing the license, GPL v3:

```
python alignancer.py --license
```

## Acknowledgements

This work was supported by SPP 1158 Antarktisforschung (DFG: FR 2134/1-1, VA 105/17-1). Additional thanks go to Stephan Frickenhaus and Bank Beszteri for helpful discussions.

## References

Eddy SR (2010). HMMER User's Guide.

<ftp://selab.janelia.org/pub/software/hmmer3/3.0/Userguide.pdf> : 26-27.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5) : 1792-1797.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.

(2007). Clustal W and Clustal X version 2.0. *Bioinformatics.* 23 : 2947-2948.