

alignment trimming

OVERVIEW

- Introduction
- Publications
- News
- trimAl team
- FAQ
- Mailing list

DOCUMENTATION

- Getting started with trimAl v1.2
- Use of the trimAl v1.2 command line
- Use of the trimAl v1.2 webserver interface
- Getting started with readAl v1.2
- Changelog

DOWNLOADS

OLDER VERSIONS

trimAl v1.2 command line version

Basic usage

-h

trimal -in <inputfile> -out <outputfile> -(other options)

Common options

```
--version
                         Print the trimAl version.
-in <inputfile>
                         Input file in several formats (clustal, fasta, NBRF/PIR, nexus
-compareset <inputfile>
                        Input list of paths for the files containing the alignments to
-matrix <inpufile>
                         Input file for user-defined similarity matrix (default is Blog
-out <outputfile>
                         Output alignment in the same input format (default stdout). (d
-htmlout <outputfile>
                         Get a summary of trimal's work in an HTML file.
-clustal
                         Output file in CLUSTAL format
-fasta
                         Output file in FASTA format
-nbrf
                         Output file in NBRF/PIR format
-nexus
                         Output file in NEXUS format
                         Output file in MEGA format
-mega
-phylip3.2
                         Output file in PHYLIP3.2 format
                         Output file in PHYLIP/PHYLIP4 format
-phylip
                         Get the complementary alignment.
-complementary
-colnumbering
                         Get the relationship between the columns in the old and new al
-select { n,l,m-k }
                         Selection of columns to be removed from the alignment. (see Us
-gt -gapthreshold <n>
                         1 - (fraction of sequences with a gap allowed).
-st -simthreshold <n>
                         Minimum average similarity allowed.
-ct -conthreshold <n>
                         Minimum consistency value allowed.
-cons <n>
                         Minimum percentage of the positions in the original alignment
                         Remove all positions with gaps in the alignment.
-nogaps
-noallgaps
                         Remove columns composed only by gaps.
-gappyout
                         Use automatic selection on "gappyout" mode. This method only \psi
                         Use automatic selection on "strict" mode. (see User Guide).
-strict
                         Use automatic selection on "strictplus" mode. (see User Guide)
-strictplus
-automated1
                         Use a heuristic selection of the automatic method based on sin
                         Minimum overlap of a positions with other positions in the col
-resoverlap
-seqoverlap
                         Minimum percentage of "good positions" that a sequence must ha
-w <n>
                         (half) Window size, score of position i is the average of the
                         (half) Window size only applies to statistics/methods based or
-qw < n>
                         (half) Window size only applies to statistics/methods based or
-sw < n>
                         (half) Window size only applies to statistics/methods based or
-cw <n>
                         Print gap percentage count for columns in the input alignment.
-sgc
                         Print accumulated gap percentage count.
-sgt
                         Print conservation values for columns in the input alignment.
-scc
                         Print accumulated conservation values count.
-sct
```

Print this information and show some examples.

Examples

-sfc

-sft

-sident

1) Removes all positions in the alignment with gaps in 10% or more of the sequences, unless this leaves less than 60%. In such case, print the 60% best (with less gaps) positions.

Print compare values for columns in the selected alignment from

Print accumulated compare values count for the selected alignment

Print identity statistics for all sequences in the alignemnt.

```
trimal -in <inputfile> -out <outputfile> -gt 0.9 -cons 60
```

2) As above but, the gap percentage is averaged over a window starting 3 positions before and ending 3 positions after the column.

```
trimal -in <inputfile> -out <outputfile> -gt 0.9 -cons 60 -w 3
```

3) Uses an automatic method to decide optimal thresholds, based in the gap percentage count over the whole alignment.

```
trimal -in <inputfile> -out <outputfile> -gappyout
```

4) Uses an heuristic method to decide which is the best automated method to trim the input alignment.

```
trimal -in <inputfile> -out <outputfile> -automated1
```

 Removes a set of columns as indicated by the user. Note that the numbering of the columns start by 0.

```
trimal -in <inputfile> -out <outputfile> -select { 0-7,12,15,23-40 }
```

6) Removes spurious sequences from the input alignment.

```
trimal -in <inputfile> -out <outputfile> -resoverlap 0.75 -seqoverlap 80
```

7) Gets the trimAl's trimming summary in an HTML file.

trimal -compareset <pathfile> -out <outputfile>

```
trimal -in <inputfile> -out <outputfile> -htmlout <htmlfile> -strict
```

8) Gets the relationship between the columns from the output and input alignments.

```
trimal -in <inputfile> -out <outputfile> -strictplus -colnumbering
```

9) Selects the most consistency alignment among a set of them. Note that the alignments

```
should have the same sequences in the same order.
```

10) Selects the most consistency alignment among a set of them. After that, applies some manual thresholds to the selected alignment.

trimal -compareset <pathfile> -out <outputfile> -gt 0.8 -st 0.7 -ct 0.3

Use of an user defined similarity-matrix

To compute a similarity score for protein alignments, trimAl uses the *Blosum62* Similarity Matrix meanwhile for DNA or RNA alignments, trimAl uses the *Identity* Similarity Matrix by default. However, in any case the user can use any other matrix.

Allowed combination

It is impossible to use all the trimAl's options at the same time. The first incompatibility is between -in (designed to work with a single file) and -compareset (designed to work with a set of files). Other incompatibility is between the statistical information options and the trimming methods; in order to use combination between those methods is necessary that an output file is defined. Other compatibility is between the output and html file, they should be different in order to avoid problem in their creation. It is important that a trimming is used when one or more among the options -colnumbering, -complementary and -htmlout is used.

use_of_the_command_line_trimal_v1.2.txt · Last modified: 2009/04/19 17:48 by scapella