

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
- Changelo

DOWNLOADS

OLDER VERSIONS

Getting Started with trimAl v1.2

Thank you for choosing **trimAl v1.2** to trim your alignments. In this version we have solved some bugs from the previous version and we have also added new functionality to the program. You will see that it is very easy to get familiar with the program. The first thing you need to do to start is to decide whether you will be using trimAl in its command line version or through the web interface. The command line version is faster and has more possibilities, so it is recommended if you are going to use trimAl extensively.

The trimAl webserver included in Phylemon 2.0 provides a friendly user interface and the opportunity to concatenate your trimmed alignment to many different phylogenetic analysis.

Input and Output formats

Alignment formats, including Phylip, Fasta, Clustal, NBRF/Pir, Mega and Nexus. The program detects automatically the input format and generates the output file in the same format. Alternatively, the user can select a different format for the output. Moreover, trimAl can provide as an output the complementary MSA, that is, the columns that would otherwise be removed by the specified parameters (option *-complementary*). Finally, to facilite the visualization of trimAl's trimming, the program can generate an html file in which selected and trimmed columns are colored differently (*-htmlout*).

Besides MSAs, trimAl can optionally produce other outputs, which have been deemed of

interest. For instance, to facilitate the tracking of the correspondences between the

columns in the original and the trimmed alignment, trimAl can return the relationship

trimAl reads and renders protein or nucleotide alignments in several Multiple Sequence

between their column numbers (option *-colnumbering*). trimAl can provide information on gap and/or conservation scores in a MSA. This information can be relative to each column or the distribution of these values along the alignment (options *-sgc* for gaps and *-scc* for conservation values) (options *-sgt* and *-sct* for gaps and conservation distribution, respectively). When comparing several alignments, trimAl can also offer statistical information about their consistency score (options *-sfc* for each column and *-sft* for whole alignment). Finally, trimAl can provide a comparison matrix summarizing the percentage of identities between each pair of sequences in the alignment, their averages and the highest identity pair for each sequence (option *-sident*).

You would need first to install **trimAl v1.2** on your computer. On the following link

Command line version

downloads you have the necessary files and instructions to get trimAl properly installed on your computer, whether you are a Linux, MacOS or Windows user.

Once you have trimAl v1.2 installed, just type "trimal" on your prompt to get the basic

In this section, we are going to use one (dataset/example1.phy) of the files included into

the trimAl package. We also are going to use the trimAl -htmlout option to show you the

A very common way of using **trimAl v1.2** to trim an alignment is to use just a gap threshold (the minimum fraction of sequences without a gap that you require to consider a column of "enough quality")

.

For example:

trimAl's performance.

will remove all columns with any gap (equivalent to *-nogaps* option)

trimal -in example1 -out output1 -htmlout output1.html -gt 1

Selected Residue / Sequence

Deleted Residue / Sequence 20 50 60 ------FPWNGLQIHMMGII Spl7 -----FAYTAPD---LLLIGFLLKTVA-T-FG--DTWF-----QLWQGLDLNKMPVF Spl0 Sp26 AAAAAAAA----ALL---TYL-GLFLGTDY-----EN---FAAAAANAWLGLEINMMAQI -----QAWGGLEINKQAIL Sp33 -----ASGAI---LTL-GIYLFTLC-AVIS--VSWY-----LAWLGLEINMMAII Sp6 trimal -in example1 -out output2 -htmlout output2.html -gt 0.8 -st 0.001

will remove all columns with gaps in more than 20% of the sequences or with a similarity

Selected Residue / Sequence

Deleted Residue / Sequence 10 20 30 50 ======+=====+=====+=====+ ------GLGKV---IVY-GIVLGTKS-DQFSNWVVWL-----FPWNGLQIHMMGII Spl7 -----FAYTAPD---LLLIGFLLKTVA-T-FG--DTWF-----QLWQGLDLNKMPVF Spl0 AAAAAAAA----ALL---TYL-GLFLGTDY----EN---FAAAAANAWLGLEINMMAQI Sp26 Sp33 -----QAWGGLEINKQAIL -----ASGAI---LTL-GIYLFTLC-AVIS--VSWY-----LAWLGLEINMMAII If you feel that, for some alignments this will be too strict and prefer to use a *minimum*

percentage of the columns in the original alignment) you can do it as follows:

trimal -in example1 -out output3 -htmlout output3.html -gt 0.8 -st 0.001 -cons 60

will remove all columns with gaps in more than 20% of the sequences or with a similarity

coverage in the trimmed alignment (that is, the trimmed alignment will retain a given

score lower than 0.001, unless this removes more than 40% of the columns in the original alignment, we want to conserve at least 60% of them. In such cases **trimAl v1.2** will add the necessary number of columns (in decreasing order of scores) so that the minimum coverage is respected.

-----FAYTAPD---LLLIGFLLKTVA-T-FG--DTWF-----QLWQGLDLNKMPVF

20

Selected Residue / Sequence Deleted Residue / Sequence

Sp8

Spl7

Sp10 ------DPAVL----FV--IMLGTIT-K-FS--SEWF-----FAWLGLEINMMVII
Sp26 AAAAAAAA----ALL---TYL-GLFLGTDY-----EN---FAAAAANAWLGLEINMMAQI
Sp33 --------PTIL---NIA-GLHMETDI-N-FS--LAWF-----QAWGGLEINKQAIL
Sp6 -------ASGAI---LTL-GIYLFTLC-AVIS--VSWY-----LAWLGLEINMMAII

Yet another threshold that you can use is based on the comparison of different alignments.

Sometimes one does not know which alignment algorithm will perform best (or which parameters, e.g gap penalties). A way out is to just produce different alignments with the different algorithms and then choose the alignment that contains the most consistent

You can then trim the output alignment, the most consistency one, with other algorithms

20

-----GLGKVIV-YGIVLGTKSDQFSNWVVWLFPWNGLQIHMMGII

-- FAYTAPDLLL - IGFLLKTV - ATFG - - DTWFQLWQGLDLNKMPVF

30

40

trimAl v1.2 can do this for you, just provide in the input file a list of the paths for the different alignments (in this case, we are going to use the file dataset/fileset1). Just type

Will trim such the most consistency alignment removing all columns with a consistency score lower than 0.5

trimal -compareset fileset1 -out output5 -htmlout output5.html -ct 0.5

or/and trim it based on the consistency values, for instance:

Selected Residue / Sequence Deleted Residue / Sequence

Sp17

and similarity scores.

Sp8 Sp17

Spl0

Sp26

Sp33

Sp6

Sp8

Sp17

Spl0

Sp8

Sp8

analyses of molecular sequence data.

residue-pairings, that is the residue pairs that are recovered by most algorithm.

Sp10 ------DPAVL--FVIMLGTI-TKFS--SEWFFAWLGLEINMMVII
Sp26 AAAAAAAAALLTYLGLFLGTDYENFA--AAAANAWLGLEINMMAQI
Sp33 -------PTILNIAGLHMETD-INFS--LAWFQAWGGLEINKQAIL
Sp6 -----ASGAILT-LGIYLFTLCAVIS--VSWYLAWLGLEINMMAII

trimal -compareset fileset1 -out output5 -ct 0.5 -gt 0.75 -st 0.001

Will trim such the most consistency alignment removing all columns with a consistency score lower than 0.5, with gaps in more than 25% of the sequences and with a similarity score lower than 0.001. In this case, it is impossible to generate an html file in order to

track the trimAl's trimming because we are applying two consecutive trimming methods. The first one is related to the consistency score and the second one is related to the gaps

Moreover, you can use one of the implemented methods to set up, depending on the alignment features, the different thresholds. Among these methods, you can find - gappyout, -strict and -strictplus as automated methods that uses gaps and similarities distribution to fix the thresholds. You also can find the -nogaps and -noallgaps methods that let you remove columns with, at least, one gap and columns with only gaps respectively. Finally, you can find the heuristic method -automated1 that it is used to decide which is the best automated method to trim your input alignment depending on its

features. If you want to know about these methods, please, see our **publication** for more details.

trimal -in example1 -out output6 -htmlout output6.html -gappyout

will remove columns from the input alignment using the gappyout method.

Selected Residue / Sequence
Deleted Residue / Sequence
10 20 30 40 50 60

-----GLGKV---IVY-GIVLGTKS-DQFSNWVVWL----FPWNGLQIHMMGII

-----FAYTAPD---LLLIGFLLKTVA-T-FG--DTWF-----QLWQGLDLNKMPVF

------QAWGGLEINKQAIL

-----ASGAI---LTL-GIYLFTLC-AVIS--VSWY-----LAWLGLEINMMAII

======+

-----GLGKV---IVY-GIVLGTKS-DQFSNWVVWL----FPWNGLQIHMMGII

-----FAYTAPD---LLLIGFLLKTVA-T-FG--DTWF-----QLWQGLDLNKMPVF

------DPAVL----FV--IMLGTIT-K-FS--SEWF-----FAWLGLEINMMVII

-----GLGKV---IVY-GIVLGTKS-DQFSNWVVWL-----FPWNGLQIHMMGII

50

60

Selected Residue / Sequence Deleted Residue / Sequence

trimal -in example1 -out output7 -htmlout output7.html -strict

will remove columns from the input alignment using the strict method.

Sp26 AAAAAAAA----ALL---TYL-GLFLGTDY-----EN---FAAAAANAWLGLEINMMAQI
Sp33 ---------PTIL---NIA-GLHMETDI-N-FS--LAWF-----QAWGGLEINKQAIL
Sp6 -------ASGAI---LTL-GIYLFTLC-AVIS--VSWY-----LAWLGLEINMMAII

trimal -in examplel -out output8 -htmlout output8.html -automated1

will remove columns from the input alignment using the heuristic automated1 method to decide which is the best automated method to trim the alignment between gappyout and strict ones.

Selected Residue / Sequence
Deleted Residue / Sequence

Sp17 ------FAYTAPD---LLLIGFLLKTVA-T-FG--DTWF-----QLWQGLDLNKMPVF
Sp10 -------DPAVL----FV--IMLGTIT-K-FS--SEWF-----FAWLGLEINMMVII
Sp26 AAAAAAAA----ALL---TYL-GLFLGTDY-----EN---FAAAAANAWLGLEINMMAQI
Sp33 --------PTIL---NIA-GLHMETDI-N-FS--LAWF-----QAWGGLEINKQAIL
Sp6 -------ASGAI---LTL-GIYLFTLC-AVIS--VSWY-----LAWLGLEINMMAII

Finally, trimAl can remove spurious sequences from your alignment. For that purpose, it is

important to define the *-resoverlap* and *-segoverlap* properly.

achieve an overlap, with the rest of the sequences, of 0.75.

Selected Residue / Sequence
Deleted Residue / Sequence
10 20 30 40 50 60

-------FPWNGLQIHMMGII

trimal -in example1 -out output9 -htmlout output9.html -resoverlap 0.75 -seqoverlap 80

will remove all those sequences that not earn, at least, to the 80% of residues that

Sp17 -------FAYTAPD---LLLIGFLLKTVA-T-FG--DTWF-----QLWQGLDLNKMPVF
Sp10 -------DPAVL----FV--IMLGTIT-K-FS--SEWF-----FAWLGLEINMMVII
Sp26 AAAAAAAA----ALL---TYL-GLFLGTDY-----EN---FAAAAANAWLGLEINMMAQI
Sp33 --------PTIL---NIA-GLHMETDI-N-FS--LAWF-----QAWGGLEINKQAIL
Sp6 -------ASGAI---LTL-GIYLFTLC-AVIS--VSWY-----LAWLGLEINMMAII

Webserver version

Alternatively you can use trimAl v1.2 through its user-friendly web interface,

implemented in **Phylemon**, an online platform for phylogenetic and evolutionary

a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypothesis testing

BUN EMOND

Instructions on how to use it can be found here.