NAME

aln3nn − progressive sequence alignment tool based on three−way sequence comparison

SYNOPSIS

aln3nn [−LwWtTgGuUSsCMvqh] [−−distance] [−−score] [−−[D|A]matrix STRING] [−−[D|A]match FLOAT] [−−[D|A]mismatch FLOAT] [−−[D|A]go FLOAT] [−−[D|A]ge FLOAT] [−−trans FLOAT] [−−wobble FLOAT] [−−psi FLOAT] [−−seqtype STRING] [fastafile] [alignmentfile]

DESCRIPTION

aln3nn is a progressive alignment tool that computes approximate exact solutions of the multiple alignment problem by the subsequent examination of sequence triplets. It is therefore predestinated to calculate more reliable alignments compared to the pairwise progressive sequence alignments approach. The alignment order is determined by means of a phylogenetic network that is constructed using the NeighborNet package. To limit processor and memory usage we apply a standard divide−&−conquer approach.

The **aln3nn** software tool is designated to align both nucleic acid and amino acid sequences. If RNA sequences are aligned the tool provides a special feature that is to use additional structural information. This makes sense because many ncRNA evolve very quickly and therefore sequence information is not always reliable. Structure in contrast evolves slowly and is thus is more stable.

OPTIONS

The following options consider distance calculation only.

−−Dmatrix STRING
: Specifies the used scoring matrix. For nucleotide sequences valid arguments for STRING are MAMMA (Match/Mismatch scores) and IUB (IUB scoring matrix). Valid matrices for amino acid sequences are PAM20, PAM60, PAM120, PAM250, PAM350, GONNET40, GONNET80, GONNET120, GONNET160, GONNET250, GONNET300, GONNET350 as well as BLOSUM30, BLOSUM40, BLOSUM45, BLOSUM62, BLOSUM80. You can specify a valid substitution matrix family such as PAM, BLOSUM or GONNET. The program then chooses a proper fitting matrix from that family on its own.

−−Dmatch FLOAT
: Score contribution for matching residues.

−−Dmismatch FLOAT
: Score contribution for mis−matching residues.

−−Dgo FLOAT
: Specify the gap open penalty for opening a gap. FLOAT should be positive.

−−Dge FLOAT
: Specify the gap extension penalty for extending a gap. FLOAT should be positive.

The following options consider alignment calculation only.

−−Amatrix STRING
: The same options as for −−Dmatrix are possible.

−−Amatch FLOAT
: Score contribution for matching residues.

--Amismatch FLOAT
Score contribution for mis-matching residues.

--Ago FLOAT
Specify the gap open penalty for opening a gap. FLOAT should be positive.

--Age FLOAT
Specify the gap extension penalty for extending a gap. FLOAT should be positive.

If the preceding â/—Dâ/— or â/—Aâ/— character is omitted, the provided ments for the switches --matrix, --match, --mismatch, --go, and --ge apply both for distance and alignment calculation.

The given match and mis-match scores apply only for the MAMMA scoring matrix. If another matrix type has been chosen these scores do not take effect.

--trans FLOAT
Gives transistions (i.e. purine-purine or pyrimidine-pyrimidine substitutions) a score between mismatch and match. A value of zero scores transitions as mismatches, while a value of one scores transitions as matches. Values smaller than zero are treated as zero, whereas values larger than one are treated as one. The transistion score is only applicable for DNA or RNA sequences.

--wobble FLOAT
Gives RNA wobble pairs (A-C and G-U pairs) a score between mis-match and match. A value of zero scores wobble pairs as mis-matches, while a value of one scores wobble pairs as matches. Values smaller than zero are treated as zero, whereas values larger than one are treated as one. The wobble score is only applicable for RNA sequences.

--psi Assigns the structure sequence balancing term a value between zero and one. A value of zero fully weights the structure, while a value of one fully weights the sequence.

--seqtype STRING
Specifies the sequence type explicitly. Set STRING to DNA for nucleotide sequences and to PROTEIN for amino acid sequences. In most cases you do not need to provide **aln3nn** with an sequence type because it tries to determine it automatically.

--distance
Calculates the mean sequence distance of the given sequences only.

--score
Calculates the alignment score of a given alignment only.

For score and distance calculation only sequence information is consid-ered.

--dclength
Length threshold for the Divide&Conquer algorithm.

-w, --weightseqs
Do weight sequences in alignment.

-W, --noweightseqs
Do NOT weight sequences in alignment.

-g, --extendedgap
Use the complete insertion/deletion model for three sequences. This setting handles all possible combinations of gap-open and gap-extensions than can occur if three sequences are considered. Enabling this option leads to more reliable alignments.

-G, --simplegap
Use the same insertion/deletion model as for the pairwise alignment approach. Saves some computation time but does not exploit the full potential of the three-way alignment approach.

-t, --weightterms
Do weight terminal gaps.

-T, --noweightterms
Do NOT weight terminal gaps.

-u, --structure
Use structural information.

-U, --nostructure
Do NOT use structural information.

-s, --sort
Sort alignment with respect to order of input sequences.

-S, --nosort
Leave alignment unsorted.

-C, --clustal
Use CLUSTAL output format.

-M, --msf
Use MSF output format instead of CLUSTAL format.

-v --verbose
Increase verbose level.

-q --quiet
Generate no output except warning and error messages.

-h, --help
Prints the help screen.

EXAMPLES

aln3nn --matrix PAM --go 6.5 --ge 2.0 --psi 0.5 sequences.fa sequences.aln

Aligns the sequences stored in file sequences.fa and save the alignment to file sequences.aln. If the sequences read are interpreted as amino acid sequences the best fitting scoring matrix from the PAM family is chosen. The gap open and gap extension penalties are set to 6.5 and 2.0 respectively. Structure and sequence are equally weighted.

aln3nn --match 2.0 --mismatch 0.0 --trans 0.5 --go 6.5 --ge 2.0 --psi 0.5 sequences.fa sequences.aln

Same as the above example with the difference that the match and mismatch scores are set explicitly. The score for pyrimidine-pyrimidine or purine-purine transitions are set to 0.5 times the match score plus 0.5 times the mismatch score

aln3nn --Dmatrix BLOSUM62 --Amatrix BLOSUM --psi 0.5 sequences.fa sequences.aln

Example where different scoring matrices are used for distance and

alignment calculation. For the distance calculation the BLOSUM62 scoring matrix is used whereas for the alignment calculation of proper fitting scoring matrix from the BLOSUM family is chosen.

aln3nn --matrix BLOSUM62 --Dgo 3.0 --Ago 1.6 --psi 0.5 sequences.fa sequences.aln

Here the BLOSUM62 scoring matrix is used for both distance and alignment calculation but different gap open penalties are chosen.

aln3nn --match 2.0 --mismatch 0.0 --trans 0.5 --go 6.5 --ge 2.0 -s sequences.aln

Calculates the alignment score of the alignment given in file sequences.aln using the given alignment parameters.

aln3nn --match 2.0 --mismatch 0.0 --trans 0.5 --go 6.5 --ge 2.0 -d sequences.fa

Calculates the mean pairwise distance of the sequences given in file sequences.fa using the given alignment parameters.

REFERENCES

Bryant, D. & Moulton, V. (2002). NeighborNet: An agglomerative method for the construction of planar phylogenetic networks. In WABI â/—02: Proceedings of the Second International Workshop on Algorithms in Bioinformatics. Springer–Verlag, London, UK, 375–391.

Hofacker, I., Fekete, M. & Stadler, P. (2002). Secondary structure prediction for aligned RNA sequences. J. Mol. Evol., 319, 1059–1066.

Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, L., Tacker, M. & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte fÃ¼r Chemie, 125, 167–188.

Stoye, J. (1997). Multiple sequence alignment with the divide–and–conquer method. Gene Combis, 211, 45–56.

VERSION

This man page concern version V0.6.0 of **aln3nn**.

AUTHORS

Matthias Kruspe and Peter F. Stadler.

BUGS

Comments and Bug reports should be sent to <matthias@bioinf.uni–leipzig.de>.