

Fall 2015

Alignment Information

Sarah O'Leary-Driscoll

Illinois Mathematics and Science Academy, soleary@imsa.edu

Follow this and additional works at: http://digitalcommons.imsa.edu/bioinfo_seq_align

 Part of the [Bioinformatics Commons](#), [Curriculum and Instruction Commons](#), and the [Science and Mathematics Education Commons](#)

Recommended Citation

O'Leary-Driscoll, S. (2015). Alignment Information.

Retrieved from: http://digitalcommons.imsa.edu/bioinfo_seq_align/3

This Resources is brought to you for free and open access by the Bioinformatics at DigitalCommons@IMSA. It has been accepted for inclusion in Sequence Alignments by an authorized administrator of DigitalCommons@IMSA. For more information, please contact pgarrett@imsa.edu, jean@imsa.edu.

Alignments:

	Pairwise DNA	Multiple DNA	Pairwise Protein	Multiple Protein
When do we use this?	When we want to compare two things from an evolutionary standpoint; this focuses on the entire DNA rather than on merely the exons like proteins.	Comparing multiple DNA strands. This is very useful for looking at cross-species evolution (the similarities of various species). Often used for Phylogenetic trees.	Comparing two proteins. This is used largely for predicting protein folding structure for an unknown molecule.	Comparing multiple proteins. This has many different uses such as predicting protein structure.
Programs?	BLAST http://blast.ncbi.nlm.nih.gov/Blast.cgi	T-Coffee http://tcoffee.crg.cat/apps/tcoffee/do:regular Muscle http://www.ebi.ac.uk/Tools/msa/muscle/ Clustal X http://www.clustal.org/clustal2/	Dotle http://myhits.isb-sib.ch/cgi-bin/dotlet Chimera https://www.cgl.ucsf.edu/chimera/download.html	T-Coffee http://tcoffee.crg.cat/apps/tcoffee/do:regular Muscle http://www.ebi.ac.uk/Tools/msa/muscle/ Clustal Omega http://www.ebi.ac.uk/Tools/msa/clustalo/
Matrices/variables	BLOSUM (higher numbers are used for more variance), PAM (higher numbers are used for more closely conserved,	BLOSUM (higher numbers are used for more variance), PAM (higher numbers are used for more closely conserved, ratio of time (2/k or 2-5/k)	BLOSUM (higher numbers are used for more variance), PAM (higher numbers are used for more closely conserved, Wunsch-Needleman,Smith-Waterman	BLOSUM (higher numbers are used for more variance), PAM (higher numbers are used for more closely conserved), Logodds
Other	Kinds: Dot Matrix, Dynamic, and Text-based	Longer sequences create a better tree	Kinds: Dot Matrix, Dynamic, and Text-based	DNA might be more useful from an evolutionary standpoint, but proteins show functional changes more clearly

Type of Alignment: Pairwise DNA alignment

Pairwise DNA alignment is frequently used to identify similar regions that will show how two sequences have functional or structural similarities. It can also be used to show how exons and introns change between different sequences and whether they have an effect on the final structure of the RNA after the DNA is processed within a cell.

Pairwise DNA alignment can be done online without having to download additional software.

NCBI has the ability to perform DNA alignments using the word method.

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=blast2seq. It can output a dot matrix format of the data, but the main method it uses is BLAST which searches for similar subsequences, or words, in between the two sequences. The user can change the length of the word, as a longer word will mean the program will search for stronger similarities between sequences. The NCBI YouTube channel has tutorials for using the alignment. NCBI alignment is useful to see exactly which regions of the two DNA sequences are conserved, and how strongly the similarities between the sequences are.

Type of Alignment: Multiple Sequence Alignment DNA

What would you use this type of alignment for?

- Easy phylogenetic analysis and identification of orthologs
- Easy predictions of protein structure and functions
- Helps in identifying appropriate degenerate primers for families of genes
- Variety of results:

- Multiple sequence alignment checks for alignment
 - alignment can indicate shared ancestry
 - could indicate similar protein products
- Phylogenetic trees show relation
- Different programs are better for different results:
 - T-Coffee is good all around for just looking at the DNA alignment
 - “T-coffee is fast” - Bioinformatics for Dummies
 - ClustalX is good for looking at alignment, but it isn’t available online
 - “Everyone uses it” - Bioinformatics for Dummies
 - MUSCLE makes phylogenetic trees, but the alignments are difficult to interpret.
 - “MUSCLE is accurate” - Bioinformatics for Dummies

What programs would you use to do this alignment?(List them with their web links. Put a description with each of what you can do and the output)

- T-coffee: <http://tcoffee.crg.cat/apps/tcoffee/do:regular>
- MUSCLE: <http://www.ebi.ac.uk/Tools/msa/muscle/>
- Download the program for ClustalX at: <http://www.clustal.org/clustal2/#Download>

What matrices/variables do you use for your alignment? PAM? Blosum? Which ones under what conditions?

- T-coffee: uses PAM and BLOSUM matrices (BLOSUM are supposedly more accurate)
- Any other sources that help with directions on how to use programs, etc.? Link them here.

T-coffee manual (has a lot of useful information on T-coffee):

Type of Alignment: Multiple Sequence Alignment Protein

What would you use this type of alignment for?

- Looking at similarities between different protein structures
- protein domain identification

What programs would you use to do this alignment?(List them with their web links. Put a description with each of what you can do and the output)

Each of these programs requires each sequence to be input into the analyzer in the FASTA format. The output of each of the programs is virtually identical.

- Clustal Omega: <http://www.ebi.ac.uk/Tools/msa/clustalo/>
- Muscle: <http://www.ebi.ac.uk/Tools/msa/muscle/>
- T-COFFEE: <http://www.ebi.ac.uk/Tools/msa/tcoffee/>

What matrices/variables do you use for your alignment? PAM? Blosum? Which ones under what conditions?

PAM

- used when compared sequences are more conserved
- a higher-numbered PAM is used for sequences with less similarities, vice versa.

BLOSUM

- used in divergent alignments
- a higher-numbered BLOSUM is used for sequences with more similarities, vice versa.

Any other sources that help with directions on how to use programs, etc.? Link them here.

Youtube Multiple Sequence Alignment: <https://www.youtube.com/watch?v=WbdsfkzaVqk>