A Review of Recent Developments in Protein Sequence Alignment

Allyson Julian

Objectives

- Review of Literature (so no independent evaluation/benchmarking)
- Summarize use of MSA for Protein Sequence Alignment
- Application of MSA algorithms/methods
- Implementation of MSA algorithms/methods
- Improvements/enhancements included in implementation

Why do we care?

- Protein alignment informs structural analysis
- Evolutionary conservation
- Large peptide sequence reads (millions – even trillions in size)
- How do we address increasing size of datasets?
- Constraints of current hardware
- Limitations of current software
- Availability of publicly accessible peptide/nucleotide databases

Bioinformatics Background

- Sequence alphabet: 21 amino acids
- Alphabet above used to form string representation of peptides
- Pairwise alignments
- (Dis)simmilarity scores
- Profile HMM
- MAFFT
- T-coffee

What will be summarized

- Input
- Methodology
  - MSA Algorithms
- Enhancements
  - Domain-specific knowledge
  - Protein/genomic external databases
- Output
  - Results
  - Accuracy (according to authors; will not be evaluating this myself)
  - Running time (also according to authors)

Software to be explored

ConSurf 2016 MAFFT

PROSOMALS3D (2014)

PSI/T-Coffee (2016)

Primary References

[1]H. Ashkenazy et al., “ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules,” Nucleic Acids Res, vol. 44, no. W1, pp. W344–W350, Jul. 2016.

[2]J. Pei and N. Grishin, “PROMALS3D: Multiple Protein Sequence Alignment Enhanced with Evolutionary and Three-Dimensional Structural Information,” in Multiple Sequence Alignment Methods, D. J. Russell, Ed. Humana Press, 2014, pp. 263–271.

[3]E. W. Floden, P. D. Tommaso, M. Chatzou, C. Magis, C. Notredame, and J.-M. Chang, “PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases,” Nucleic Acids Res, vol. 44, no. W1, pp. W339–W343, Jul. 2016.