

Comparison of Cost Functions in Sequence Alignment

Ryan Healey

Use of Cost Functions

- Used to score and align sequences
- Mathematically model how sequences mutate and evolve.
 - Evolution and mutation can be dependent on the source and other conditions of each sequence. Cost functions can be context dependent.
- Small changes can have significant effects (as measured by sensitivity)

Common Methods: Gap Functions

- Simple
 - C
- Affine
 - $C + C'L$
- Logarithmic
 - $C + C'\text{Log}(L)$
- Affine-Logarithmic
 - $C + C'L + C''\text{Log}(L)$

Other Methods

- Stochastic / Probabilistic
- Weighting by sequence, position, nucleotide, etc.
- Structural Homology
- Guidance
- And more

Questions to Answer

- When (if ever) is each method preferred?
- How are parameter values chosen?
- How do the parameter values affect performance?
- What limitations may make each method insufficient?

Methods of Comparison

- Popularity
- Speed-Complexity
- Alignment Accuracy
- Divergence
- Others?

Sources

- [1] Altschul, Stephen. "Gap Costs for Multiple Sequence Alignment." *Gap Costs for Multiple Sequence Alignment - ScienceDirect*. Journal of Theoretical Biology, n.d. Web.
- [2] Cartwright, Reed A. "Logarithmic Gap Costs Decrease Alignment Accuracy." *BMC Bioinformatics*. BioMed Central, 05 Dec. 2006. Web.
- [3] Cartwright, Reed A. "Problems and Solutions for Estimating Indel Rates and Length Distributions." *Molecular Biology and Evolution*. Oxford University Press, 28 Nov. 2008. Web.
- [4] Fan, YanHui, Qi Shi, JinFeng Chen, WenJuan Wang, HongXia Pang, JiaoWei Tang, and ShiHeng Tao. "The Rates and Patterns of Insertions, Deletions and Substitutions in Mouse and Rat Inferred from Introns." *SpringerLink*. SP Science in China Press, 17 Sept. 2008. Web.
- [5] Liu, Kevin, and Tandy Warnow. "Barking Up The Wrong Treelength: The Impact of Gap Penalty on Alignment and Tree Accuracy - IEEE Xplore Document." *Barking Up The Wrong Treelength: The Impact of Gap Penalty on Alignment and Tree Accuracy - IEEE Xplore Document*. IEEE, 2008. Web.
- [6] Keightley, Peter D., and Toby Johnson. "MCALIGN: Stochastic Alignment of Noncoding DNA Sequences Based on an Evolutionary Model of Sequence Evolution." *Genome Research*. Cold Spring Harbor Lab, 01 Jan. 1970. Web.
- [7] Kim, Jaebum, and Saurabh Sinha. "Indelign: A Probabilistic Framework for Annotation of Insertions and Deletions in a Multiple Alignment." *Bioinformatics*. Oxford University Press, 15 Nov. 2006. Web.
- [8] Liu, Kevin, and Tandy Warnow. "Treelength Optimization for Phylogeny Estimation." *PLOS ONE*. Public Library of Science, 19 Mar. 2012. Web.
- [9] Lunter, Gerton. "Probabilistic Whole-genome Alignments Reveal High Indel Rates in the Human and Mouse Genomes." *Bioinformatics*. Oxford University Press, 01 July 2007. Web.
- [10] Ogden, T. Heath, and Michael S. Rosenberg. "Alignment and Topological Accuracy of the Direct Optimization Approach via POY and Traditional Phylogenetics via ClustalW + PAUP*." *Systematic Biology*. Oxford University Press, 01 Apr. 2007. Web.
- [11] Phillips, Aloysius, Daniel Janies, and Ward Wheeler. "Multiple Sequence Alignment in Phylogenetic Analysis." *Multiple Sequence Alignment in Phylogenetic Analysis - ScienceDirect*. Molecular Phylogenetics and Evolution, Sept. 2000. Web.
- [12] Redelings, Benjamin. "Erasing Errors Due to Alignment Ambiguity When Estimating Positive Selection." *Molecular Biology and Evolution*. Oxford Academic, 27 May 2014. Web.
- [13] Rivas, Elena, and Sean R. Eddy. "Parameterizing Sequence Alignment with an Explicit Evolutionary Model." *BMC Bioinformatics*. BioMed Central, 10 Dec. 2015. Web.
- [14] Shafee, Thomas M. A., Andrew J. Robinson, Nicole Weerden, and Marilyn A. Anderson. "Structural Homology Guided Alignment of Cysteine Rich Proteins." *SpringerPlus*. Springer International Publishing, 12 Jan. 2016. Web.
- [15] Thompson, Julie D., Desmond G. Higgins, and Toby J. Gibson. "CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific Gap Penalties and Weight Matrix Choice." *Nucleic Acids Research*. Oxford University Press, 11 Nov. 1994. Web. 12 Apr. 2017.
- [16] Varón, Andrés; Wheeler, Ward; and Bar-Noy, Amotz, "TR-2008015: An Efficient Heuristic for the Tree Alignment Problem" (2008). *CUNY Academic Works*.
- [17] Varón, Andrés, and Ward C. Wheeler. "The Tree Alignment Problem." *BMC Bioinformatics*. BioMed Central, 2012. Web.
- [18] Yamane, Kyoko, Kentaro Yano, and Taihachi Kawahara. "Pattern and Rate of Indel Evolution Inferred from Whole Chloroplast Intergenic Regions in Sugarcane, Maize and Rice." *DNA Research*. Oxford University Press, 01 Jan. 2006. Web.
- [19] Zhang, Jia, Li Xiao, Yufang Yin, Pierre Sirois, Hanlin Gao, and Kai Li. "A Law of Mutation: Power Decay of Small Insertions and Small Deletions Associated with Human Diseases." *SpringerLink*. Humana Press Inc, 10 Oct. 2009. Web.