DeltaProt: Molecular comparison of proteins based on sequence alignments.

Steinar Thorvaldsen[1], Tor Flå[1] and Nils P. Willassen[2]
[1]Dept. of Mathematics and Statistics, University of Tromsø, [2]Norwegian Structural Biology Centre,
9037 Tromsø - Norway.
steinart@math.uit.no

1 Abstract

In the toolbox *DeltaProt* we present statistical methods, trend-tests and visualizations that are useful when the protein sequences in alignments can be divided into two or more groups based on known phenotypic traits such as preference of temperature, pH, salt concentration or pressure. The algorithms have been successfully applied in the research on extremophile organisms.

2 Objectives

We want to identify proteome-wide, and protein-specific, characteristics of cold adaptation (psycrophily). This is done by using comparative genomics on cold-adapted organisms, and similar genes from organisms with normal growth temperatures (mesophiles). In particular we have studied gamma-Proteobacteria from the order *Vibrios,* where many genomes with different optimum growth temperature (T_{opt}) are already completed.

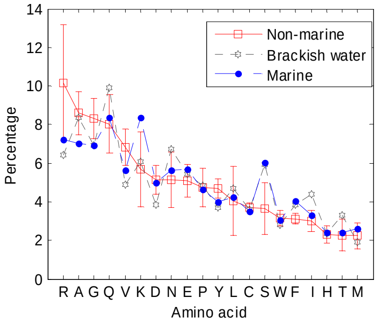
3 Methods and applications

We consider the amino acid sequence compositions, and substitution patterns, to determine whether there are underlying trends that explain the observed variation. More than 80 different physicochemical properties of the amino acid may also be applied in order to reduce the sequence alphabet to measurements. Each situation is analysed by appropriate statistical methods:

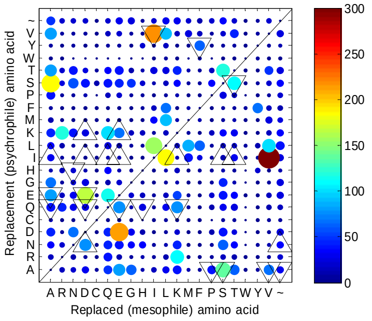
Composition:
Linear regression
Substitutions:
Fishers exact test, Chi-square tests, Mantel-Haenszel test
Properties:
Wilcoxon paired test,
Non-parametric regression based on Mann-Kendall statistics
Multiple test correction:
False discovery rate (FDR).

If the secondary and/or 3D structure is known, or may be predicted, the analyses can be performed in each of these regions.

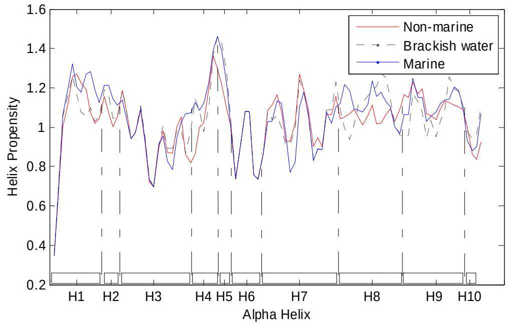
4 Some results



Composition of amino acids in ortholog sequences of bacterial Endonuclease I, ranked according to frequencies in the Non-marine group.



Visualization of the mumber of pairwise substitutions observed in a comparison of 298 ortholog proteins between two groups. The favored substitutions (P-value < 0.05) in the Mesophile→Psychrophile direction are marked with upward-pointing triangles, and the non-favoured substitutions (P-value < 0.05) are marked with downward-pointing triangles..



Peptide plot of a physicochemical property (helix propensity) for three groups of Endonuclease I from different environments. A box filter of size m x 3 was used as smoothing technique to recover the underlying structure in the data, where m is the number of sequences in the group. There is a significantly higher helix propensity in the salt adapted group (P-value=0.00001)

5 References

DeltaProt is a Matlab© companion Toolbox that can be used freely for academic, non-profit purposes.
Available from http://www.math.uit.no/bi/deltaprot/

S. Thorvaldsen, T. Flå and N. P. Willassen: Extracting molecular diversity between populations through sequence alignments. *Lecture Notes in Bioinformatics*, Vol. 3745, Springer-Verlag 2005, pp. 317-328.

S. Thorvaldsen, E. Ytterstad and T. Flå: Property-dependent analysis of aligned proteins from two or more populations. *Proceedings of the 4th Asia-Pacific Bioinformatics Conference* (Eds.: T. Jiang et al.). Imperial College Press 2006, pp. 169-178.