RH: LANDERER ET AL.— predicting amino acid functionality

Predicting amino acid functionality from sequence data in a phylogenetic framework.

Abstract

CEDRIC LANDERER[1,2,*], BRIAN C. OMEARA[1,2], AND MICHAEL A. GILCHRIST[1,2]

[1]Department of Ecology & Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610

[2]National Institute for Mathematical and Biological Synthesis, Knoxville, TN 37996-3410

*Corresponding author. E-mail: cedric.landerer@gmail.com

Version dated: April 11, 2018

Outline

Introduction

- Classic phylogenetic approaches are designed to emulate the mutation of sequences and infere the time it takes to mutate from one to another as well as the relationship of sequences to oneanother.
- However, these models still often ignore the effects mutations have on a protein sequence and how this effect in turn affects the fixation probability due to differences in fitness.
- Incorporating selection into phylogenetic analysis is therefore important and many attempts have been made to do so (Yang & Nilesen, Halpern & Bruno, ...)
- These codon models do not include information on site specific amino acid preference, all amino acids are equally prefered at each site (staionary distribution is uniform).
- ????How to go from here without copying SelAC introduction????
- Novel approaches attempt to incorporate information about site specific amino acid fitness obtained from deep mutation scanning (DMS) experiments.
- These models allow the incorporation of amino acid preference and estimate how preferences in natural sequences deviate from experimentally obtained preferences.
- While there is great value in DMS experiments and it has been shown that DMS informed models are superior to classical codon models (Bloom) these experiments are limited to singele proteins that can be put under artificial selection in the lab.
- SelAC in turn estiamtes the functionality of each amino acid at each site based on a mechanistical model of protein selection.
- SelAC has the advaantage that:

    - in contrast to other codon models does not assume a uniform amino acid preference at each site.
    - it estimates the amino acid preference at each site from the sequence data and therefore is not limited to single proteins, or proteins that can be experimentally modified.
    - is not limited to fast growing lab cultivated organisms.
    - does not requiere experimental overhead (mutation libray, artificial selection, ...)
    - no artificial stress factors (selection pressure), but sequences that have evolved naturally.
- In this work, we check compare the SelAC estimated amino acid preference with amino acid prefernces obtained by DMS approaches and a simple majority estimate under two models, SelAC itself, and phydms (Hilton 2017)
    - phydms was specifically developed to work with data from DMS experiments, using experimental amino acid preferences.
    - SelAC estimates amino acid preferences based on physico-chemical (PC) properties and distances between amino acids and an infered prefered amino acid
    - We converted SelAC and majority rule amino acid preferences into a format that works with phydms and checked which sequence performes best under each model.

Results

- Compare DMS from Firnberg and Stiffler to SelAC and majority under SelAC and phydms
    - Comparison of Frinberg under SelAC for TEM and SHV (three sequences: DMS, Majority, SelAC)

  - Comparison of Frinberg under phydms for TEM and SHV (three sequences: DMS, Majority, SelAC)
  - Comparison of Stiffler under SelAC for TEM and SHV (three sequences: DMS, Majority, SelAC)
  - Comparison of Stiffler under phydms for TEM and SHV (three sequences: DMS, Majority, SelAC)

- Comparison of perfered sequence
  - Simulations of sequences under each prefered sequence.
  - Only majority rule (duh) and SelAC agree with observed sequences.

- SelAC is dependent on choice of PC propertie to produce amino acid rankorder and assumes stabilizing selection.
  - Rankorder of certain sites can not be produced by any of the PC checked (no combination checked)

Discussion

- SelAC sequence outperformes DMS experiments, reflecting evolution better than DMS sequences under artificial selection pressure.

- SelAC only uses prefered state as input, no information about 2nd or third prefered amino acid.

- The reduction of a DMS experiment to this state might be considered an unfair comparison, however, we tested the sequences under phydms (no reduction of information), with the same result.

- This also means that SelAC produces the same information a DMS experiment would, but for naturally evolving sequences and can be applied to any sequence.

- TEM/SHV have not evolved to combate specific human developed antibiotics, but as means of ”warfare” between bacteria (need more reading here).
- This could be the cause for the great difference between DMS and observed sequences.
- SelAC, however can not provide any information about antibiotic resistency, making DMS very valuable, but not for phylogenetics.
- but additional tip information could be combined with SelAC to get at this information (out of scope? future directions?).

Introduction

Materials & Methods

Results

Discussion

Supplemental Material