Check-ins, Presentations, and Review

Jan 17 – first project check-in. Data curation and calculation of basic alignment statistics. You have been given four alignments. For each alignment, please do the following:

- Convert the alignment from letters to numbers (one hot encoding)
- For each alignment, remove highly gapped (more than 50% gapped) positions and sequences
- For each alignment, report the number of sequences and number of positions following gap removal
- For each alignment, plot a histogram of the average pairwise sequence identity

Feb 14 – second project check-in. Analysis of covariance between amino acid positions.

- Decide how you will compute covariance. Between individual amino acid types, or classes? How will you compress covariance between amino acid types to obtain a positional measure?
- Compute co-variance between all position pairs within each protein.
- Concatenate the alignments by species. You will need to make one concatenated alignment per protein pair.
- Compute co-variance between all protein pairs.
- Plot the resulting covariance matrix as a heatmap.
- Plot histograms of covariance within and between individual domains.

Feb 21 – third project check-in. Analysis of mutual information between amino acid positions.

- Decide how you will compute mutual information (MI). Between individual amino acid types, or classes? How will you compress covariance between amino acid types to obtain a positional measure?
- Compute MI between all position pairs within each protein.
- Using the concatenated alignments, compute MI between all protein pairs.
- Plot the resulting MI matrix as a heatmap.
- Plot histograms of MI within and between individual domains.

The final project presentations will be on February 28 and March 4. Final presentations will be 20 minutes long with 10 minutes for questions. The goal is now to examine if covariance and/or mutual information can be used to predict protein interactions. Choose an appropriate statistical testing strategy to decide whether the co-evolution between two proteins is significant. Consider carefully your null model.

The presentations (Feb 28):

Each group will have 30 minutes to present (including questions)

Please present your work in its entirety, starting with alignment processing and statistics

Make sure the rationale behind any strategy decisions is described (and justified)

All plots should be well-labeled and easy to read

Make a statistical argument for your predictions of interacting pairs

The review process (March 4):

To facilitate critical thinking, and invite constructive feedback, we will assess all projects through an in-class review on *March 4*. Each project will be assigned three peer reviewers. These reviewers will assign scores in the following categories:

- Technical soundness and execution. Are the methods used appropriately? Did the authors complete the work they set out to do? Are the results interesting?
- Innovation. Is the work creative? Does it make use of new ideas/concepts/approaches?
- Presentation. Was the work clearly presented? Was the use of plots/graphics appropriate and well-described?

Scores will be assigned on an NIH-like scale of 1-10. A score of 1 corresponds to "outstanding", 4-5 indicates "very good", and 10 indicates an area that needs substantial work or revision. The three reviewers will present their scores in class on March 4, along with their rationale – these presentations should be intellectually rigorous, provide specific feedback, and above all constructive. After each reviewer presents their scores and rationale, all students in the class will submit a final score to the instructors. The instructors will use these peer scoring sheets alongside their own opinions in assigning a final grade. The final project presentation and your participation in the review process and check-ins will be worth 20% of your grade.

(You will submit your written reviews after class on March 4)

Review assignments:

Group 1: Devin, Rongqing, Yifei

Group 2: Junru, Yao, Yuchen

Group 3: Asharkumar, Elliot, Ruheng

Group 4: Sriya, Joey, Karolina

Reviewers 1: Junru, Asharkumar, Sriya

Reviewers 2: Devin, Elliot, Joey

Reviewers 3: Rongqing, Yao, Karolina

Reviewers 4: Yifei, Yuchen, Ruheng