Genome 540 discussion

February 6th, 2025 Joe Min





Homework 4 questions?

Sizing the protein sequence space

The evolution of Rubisco

Homework 4 questions?

Sizing the protein sequence space

Protein sequences

We know sequences aligning is significant, but why?

Consider:

Alphabet size: 20 canonical amino acids (AAs)

Typical sequence length: 300aa

For even just one (1) short protein of length 100aa

• 20^100 possibilities, roughly 1.27 * 10^130

Shaping the protein sequence space

Imagine flattening the space into a 2D plane, and adding a third dimension for fitness

 This landscape is very jagged because adjacent sequences can have very different fitnesses (e.g., point mutations)



Evolution traverses this landscape



Evolution parallelizes explorations

Through parallel explorations of the space (different lineages), we can end up with a diverse set of fit sequences that perform the same function



Evolution is sometimes not enough

Evolution has only explored about 4 * 10^21 sequences

- Bounded by allowed mutations (i.e., in local maxima)
- Sequences are generally similar (can be found by sequence alignment!)



Dryden et al. J R Soc Interface (2008) Huang et al. Nature (2016)

Evolution is sometimes not enough

Scientists can also explore this space, using methods like directed evolution and protein design

• Protein design is a claw machine, can reach down anywhere in this space to pick a sequence to test



Dryden et al. J R Soc Interface (2008) Huang et al. Nature (2016)

Evolution of Rubisco

Rubisco review



https://openstax.org/books/biology-2e/pages/8-1-overview-of-photosynthesis

Rubisco evolved into distinct forms





Different forms are 30-90% similar in sequence

Bouvier et al. bioRxiv (2022) Tabita et al. J Exp. Bot. (2008)

Rubisco evolved into *entrenched* forms





Bouvier et al. bioRxiv (2022) Mueller-Cajar. Front Mol Biosci (2017)

Nature provided countermeasures

Small subunit helps protect the active site from oxygen

Rubisco activase (Rca) kicks off inhibitory non-substrate ligands



Sequence similarity scales with function similarity

Both Form IA and Form IB have within-group similarity of >90%

IA and IB are ~85% identity to each other



Sequence similarity scales with function similarity

- Form IV is the most diverged in sequence
- Does not catalyze the same reactions as the other three





Reminder:

Homework 4 is due Sunday, February 9th at 11:59pm!