Do protein language models understand evolution? Mixed evidence from ancestral sequences and ESM2

Protein language models are trained on evolutionarily related sequences, yet the extent to which they capture the underlying evolutionary relationships remains unclear. We explore this question using reconstructed ancestral protein sequences and the ESM2 protein language model.

Purpose

We wondered how protein language models trained on extant sequences would interpret plausible ancestral sequences. To explore this, we used ESM2 to evaluate maximum likelihood ancestral sequence reconstructions for two example gene families. We found that ESM2 often finds these ancestral sequences more plausible than extant descendants, and can distinguish between crude consensus ancestral sequences and more sophisticated maximum likelihood reconstructions. However, these patterns are context- and model-dependent, suggesting that further investigation is needed to

determine which evolutionary relationships are truly captured by large protein language models like ESM2.

We are sharing these results to encourage further exploration of how large foundation models interpret the evolutionary relationships embedded in their training data. We think this could be useful to researchers interested in interrogating the implicit learning in large protein language models, like ESM2, and we propose that ancestral sequences offer a useful tool for this purpose.

View the notebook

In the future, we hope to host notebook pubs directly on our publishing platform. Until that's possible, we'll create stubs like this with key metadata like the DOI, author roles, citation information, and an external link to the pub itself.

Contributors (A–Z)

- **Evan Kiefl**: Validation

- **Isabel Nocedal**: Conceptualization, Formal Analysis, Software, Visualization, Writing

- **Ryan York**: Editing, Supervision