

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.

Published Jul 29, 2025

 Arcadia Science

DOI: 10.57844/arcadia-5cwu-spn8

The **full pub** is available [here](#).

The **source code** to generate it is available in [this GitHub repo](#) (DOI: [10.5281/zenodo.16620544](https://doi.org/10.5281/zenodo.16620544)).

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