

# MSA-based pLMs encode evolutionary distance but don't reliably exploit it

**We characterize how MSA Pairformer encodes phylogenetic relationships. Sequence weights correlate with evolutionary distance, with distinct layers specializing as phylogenetic filters. Yet uniform averaging often outperforms learned weights for contact prediction.**

Published Jan 30, 2026



Arcadia Science

DOI: 10.57844/arcadia-ppdz-asaz

## Purpose

Auditing the evolutionary information encoded by MSA Pairformer is a critical step in understanding the behavior of protein language models (pLMs) that leverage external context in the form of MSAs. In this study, we bridge MSA Pairformer with classical phylogenetics to quantify how the model's internal sequence weighting maps onto inferred phylogenetic trees. We show that while MSA Pairformer effectively encodes evolutionary relatedness, and does so in specific layers, this signal is a subtle optimization rather than a load-bearing pillar for contact prediction accuracy. This work is intended for biologists and model developers seeking to interpret how evolutionary history is represented within pLMs. It provides a framework for evaluating the "phylogenetic operating range" of MSA-based models and highlights the need for future architectures to selectively gauge input signal reliability.

## View the notebook

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

The **source code** to generate it is available in [this GitHub repo](#) (DOI: [10.5281/zenodo.18423956](#)).

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.

---

## Social

## Contributors (A–Z)

- **Rishi De-Kayne**: Critical feedback, Editing
- **James R. Golden**: Critical feedback, Supervision, Validation
- **Evan Kiefl**: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing
- **Austin H. Patton**: Critical feedback, Editing
- **Ryan York**: Critical feedback, Supervision