Contribution ID: **18** Type: **Poster only**

Poster #18: Decoding the phage unknome with protein language models

Monday, 5 October 2026 20:39 (1 minute)

Bacteriophage genomes encode large numbers of proteins without functional annotation, collectively termed the phage "unknome."This limits both mechanistic understanding of phage biology and the discovery of proteins with potential applications in biotechnology and therapy. Current annotation approaches rely primarily on sequence or structural homology and provide indispensable first evidence for newly sequenced phages. However, they often fail when similarity to characterized proteins is weak, leaving few experimentally testable hypotheses.

Here, we develop a machine-learning framework that predicts functional classes for phage proteins of unknown function. The model uses protein language model embeddings to assign proteins to one of 89 curated functional categories. It accurately recovers existing annotations in both a held-out test set and entirely unseen phages (F1-score > 90 %). Moreover, although genomic context is not included during training, predictions for proteins encoded within operons are functionally coherent, suggesting that the model captures biologically meaningful sequence features. Applied to the phage unknome, the framework produces confident functional-class predictions that increase the proportion of classifiable phage proteins to more than 60%. Analyses of the model phages T4 and phiKZ illustrate how these predictions can generate hypotheses for previously uncharacterized genes. A prototype web server integrates model outputs with sequence- and structure-based evidence and enables interactive visual inspection across the phage protein universe. Together, this framework expands functional annotation beyond detectable homology and supports targeted experimental characterization of the phage unknome.

Reference #1

Reference #2

Reference #3

Authors: HACKL, Simon (University of Tübingen); SCHEURENBRAND, Mona (University of Tübingen); NIESELT, Kay (University of Tübingen); Dr WOLFRAM-SCHAUERTE, Maik

Presenter: Dr WOLFRAM-SCHAUERTE, Maik

Session Classification: Poster-Session