Project	Model name	Model architecture	Protein source	Mutl-task training objective	Number of proteins	Training objective
ProSE[30]	MT-LSTM	LSTM	Unifref90	yes	76,215,872	Masked token + predicting contact between residues in protein structure + structural similarity of proteins by SCOP hierarchy
ProSE[30]	DLM-LSTM	LSTM	Unifref90	no	76,215,872	Masked token
ProteinBERT[29]	ProteinBERT	Transformer	Uniref90	yes	$\sim$106M	Masked token + GO term annotation
ProtTrans[28]	protbert_bfd	Transformer	BFD	no	2,122M	Masked token

Supplemental Table 1: PLMs evaluated for viral protein VPF functional classification. Four PLMs were used to embedd PHROG VPFs proteins for training and testing the viral function classifier. Model name is name from original study. PLMs were chosen to vary in the model architecture, protein source, and use of multiple training objectives.

Method	Recall (%)	Precision (%)	F1-score (%)
PLM-based classifier	90.32	96.88	93.48
*DeePVP	88.10	96.75	92.22
*PHANNs	91.68	76.11	83.17

Supplemental Table 3: PLM-based viral protein sequence embedding for phage virion protein (PVP) classification task performance. PVP classification task designed previously[36] with PHANNs dataset[37]. * is performance reported previously[36].

Category	Support	Precision	Recall	F1-score	FDR	Decision threshold
DNA, RNA and nucleotide metabolism	20,240	0.92	0.93	0.93	0.08	0.25
connector	2,412	090	0.63	0.74	0.10	0.74
head and packaging	15,601	0.90	0.80	0.85	0.10	0.41
integration and excision	484	0.99	0.91	0.95	0.01	0.46
lysis	2,237	0.90	0.65	0.76	0.10	0.53
moron, auxiliary metabolic gene and host takeover	4,414	0.90	0.72	0.80	0.10	0.64
other	7,810	0.90	0.62	0.74	0.10	0.50
tail	12,257	0.90	0.83	0.86	0.10	0.45
transcription regulation	682	0.94	0.87	0.90	0.06	0.29

Supplemental Table 4: Functional classifier evaluation for EFAM VPFs labeled with PHROGs annotation. PHROGs functional annotation was assigned to EFAM VPFs using HMM matching. Support indicates how many EFAM VPFs matched PHROG VPFs per category. False discovery rate (FDR) is reported for the category decision threshold.