Advancing the Discovery of Phage-Host Interactions and Disease Classification from Metagenomic Profiles Using Deep Learning

Yiyan Yang[1], Tong Wang[1], Dan Huang[1], Xu-Wen Wang[1], Scott T Weiss[1], Joshua Korzenik[2], Yang-Yu Liu[1]

[1] *Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA*

[2] *Division of Gastroenterology, Hepatology and Endoscopy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA*

ABSTRACT

Microbiomes, the communities of microorganisms, are essential for sustaining ecosystem functions in diverse environments, including the human gut. Bacteriophages (phages) interact dynamically with their prokaryotic hosts and play a crucial role in shaping the structure and function of microbial communities. Previous approaches for inferring phage-host interactions from metagenomic data (e.g., assembly-based methods) are constrained by high computational demands, limited sensitivity, and the inability to accurately capture ecological relationships. To address these issues, we developed phiNODE (phage-host interaction predictor using Neural Ordinary Differential Equations), a deep learning method for predicting phage-host interactions directly from metagenomic profiles. We first validated phiNODE using synthetic datasets generated by ecological models and found that it outperformed both alternative deep-learning and co-abundance-based methods in inferring phage-host interactions. We then applied phiNODE to a large-scale metagenomic dataset comprising 7,016 stool samples from healthy individuals and identified 90% more genus-level phage-host interactions than traditional assembly-based methods. Finally, we demonstrated that the latent representations learned by phiNODE served as more powerful features than taxonomic abundance-based features for disease classifications. In summary, phiNODE offers a novel framework for inferring phage-host interactions from shotgun metagenomic data, representing an innovative approach for advancing microbiome research and clinical applications.

INTRODUCTION

Microbiomes exist in virtually every environment on Earth, from soil and oceans to the surfaces and interiors of plants and animals[1–4]. These communities typically comprise a diverse set of microorganisms, including archaea, bacteria, viruses (notably phages), and fungi, that collectively sustain ecosystem functions or influence host health[5–7]. In the human microbiomes, while numerous studies have documented shifts in prokaryotic taxa between healthy individuals and those with disease[8–13], the role of phages has become increasingly recognized[14,15]. Emerging evidence indicates that the interactions between phages and their bacterial hosts, referred to as phage-host interactions hereafter, are critical determinants in shaping the structure and dynamics of the human microbiomes[16–18], with profound implications for the development and progression of various diseases[19–24].

In the absence of comprehensive viral genome catalogs, virome profiling has traditionally relied on *de novo* assembly of shotgun metagenomic reads into contigs for viral identification[25–29]. To infer phage–host interactions, viral contigs are assembled and phage markers (e.g., clustered regularly interspaced short palindromic repeats (CRISPR) spacers and prophage signatures) are mapped to metagenome-assembled genomes (MAGs)[30,31]. However, this pipeline presents several drawbacks. First, low-abundance phages are under-represented, leading to poor genome coverage and unreliable assemblies[32–34]. Second, phage-host interactions are inferred exclusively from characterized genomic features, restricting predictions to known mechanisms and overlooking novel or uncharacterized relationships. Finally, the high computational demands of assembly-based workflows prolong processing times and increase resource requirements, limiting scalability for large-scale studies.

Recent establishment of human gut virome reference databases[30,35–37] has driven the development of read-based, reference-dependent virome profilers[32,38]. These tools map viral reads directly to curated catalogs, overcoming the sensitivity limitations of assembly-based methods and detecting approximately four times more viral species[32]. As a result, they deliver more comprehensive cross-domain profiles of the gut microbiome, accurately representing both prokaryotes and viruses. Moreover, read-based profiling enables unprecedented exploration of phage-host interactions in the human gut. Although these associations can be inferred from co-abundance correlations across samples[39,40], they remain susceptible to spurious associations driven by coincidental co-occurrence rather than genuine biological relationships[41].

To address these challenges, we investigated deep learning as an alternative approach to uncover phage-host interactions from metagenomic profiles. Deep learning approaches have shown significant potential for elucidating intrinsic, biologically relevant interactions by modeling complex, high-dimensional nonlinear relationships[42–45]. For example, a deep learning model trained on paired

microbiome and metabolome datasets accurately predicted metabolomic outputs from microbial compositions and inferred microbe-metabolite interactions[44]. However, no existing deep learning framework has been applied to directly infer phage-host interactions from metagenomic profiles.

In this study, we propose a framework of using deep learning models to predict prokaryotic abundance profiles from phage abundance profiles (**Fig. 1a**), infer candidate phage-host interactions by interpreting the models (**Fig. 1b**), and leverage the latent representations learned from the models as features for downstream disease classification (**Fig. 1c**). To achieve this, we introduce phiNODE (phage-host interaction predictor using Neural Ordinary Differential Equations), a NODE-based deep learning model that infers phage-host interactions directly from metagenomic taxonomic abundance profiles. On synthetic data generated from ecological models, phiNODE accurately predicted the hosts for 90-97% of phages and identified more true interactions than both traditional co-abundance methods and alternative deep learning approaches. Applied to a large-scale dataset of 7,016 healthy-donor stool metagenomic samples, phiNODE uncovered 90% more phage-host interactions than assembly-based pipelines. Furthermore, the latent embeddings learned by phiNODE provide more informative features than features derived from taxonomic abundance profiles for distinguishing cases from controls across multiple diseases.

RESULTS

Overview of phiNODE

For a given set of shotgun metagenomic samples, phiNODE takes paired phage and prokaryotic abundance profiles derived from taxonomic profiling as inputs (**Fig. 2**). In an illustrative dataset of twelve samples, eight were allocated to training and four to testing (**Fig. 2a**). In practice, we randomly split the data into training, validation, and test sets at an 8:1:1 ratio. The validation set serves both for early stopping to prevent overfitting and for hyperparameter tuning to maximize predictive performance. After training, model accuracy is assessed on the test set by computing the Pearson correlation coefficient (PCC) between predicted and experimentally measured relative abundances for each prokaryotic taxon (**Fig. 2b**). The overall predictive power of phiNODE is reported as the mean PCC averaged over all taxa.

Previous studies[32,46] and our analysis of real-world data indicate that phage taxa typically outnumber prokaryotic taxa by a factor of 7~24 (**Supplementary Table 1**). The high diversity and individual specificity of phages motivated us to utilize phage profiles to predict prokaryotic profiles instead of the other way around. Considering the large disparity in the numbers of prokaryotic and phage taxa, it is not possible to directly apply the original NODE architecture[47] as it requires input and output dimensions to be equal. Inspired by the architecture of mNODE[44], we incorporated two fully connected hidden layers with the NODE in between to transform the data dimensions (**Fig. 2c**). The NODE module in the middle of

the architecture computes the time evolution of ODEs whose first-order time derivatives are approximated by two fully connected layers. We used the NODE to reduce memory and training time as NODE shares connection weights across continuous layers and utilizes fewer parameters (see **Methods**).

To evaluate phiNODE's ability to infer phage-host interactions, we conducted a sensitivity analysis (**Supplementary Figure 1a**). We perturbed the relative abundance of phage *i*, denoted as x_i, by a small amount Δx_i, and used the trained phiNODE to predict the relative abundance of prokaryote *j*, denoted as y_j, and measure the deviation from its original unperturbed abundance, denoted as Δy_j. We define the sensitivity of prokaryote *j* to phage *i* as $S_{ji} = \frac{\Delta y_j}{\Delta x_i}$. Putative hosts for phage *i* are then identified by one of two criteria: (1) single-host prediction: select the prokaryote *j* with the highest S_{ji} across samples or (2) multi-host prediction: include any prokaryote *j* for which S_{ji} exceeds a predefined threshold (**Supplementary Figure 1b**). More technical details on the computation of S_{ji} are provided in Methods.

phiNODE can predict prokaryotic profiles from phage profiles on synthetic data

To investigate if prokaryotic profiles can be predicted from phage profiles and if phage-host interactions can be inferred from the trained model, we first validated phiNODE using synthetic data with known ground truth. The data was generated by two ecological models: the Generalized Lotka-Volterra (GLV) model[48,49] and the dynamic population (DP) model[50]. Each independent synthetic community ('sample') was generated by sampling a set of phages and prokaryotes from a pool of taxa to initiate the community assembly and then collecting the steady-state or dynamically balanced phage and prokaryotic relative abundances as phage and prokaryotic profiles, respectively. Specifically, we assumed a metapopulation pool of N_{taxa}=1,100 (100 prokaryotic taxa and 1000 phage taxa) for both models, with a prokaryote and phage ratio of 1:10 reflecting the taxa ratio we observed in real-world data (**Supplementary Table 1**) and within the estimated ranges from 1:1 to 1:100 reported in previous studies[32,46,51–54].

The two models differ in complexity and ability to simulate the phage-host interactions. The GLV model is relatively simple and ignores the molecular details of phage infection and only account for phage-host and inter-host interactions (**Fig. 3a**). We also assumed extremely narrow-host range for phages during the simulation – each phage taxon can only infect one host while a bacterial host could be potentially infected by ten phage taxa. Unlike the GLV model, the DP model is more sophisticated and simulates a variety of phage-host processes like infection, lysis/lysogen, induction, reversion and phage death, which is one of the state-of-art phage-host dynamic models[55]. In the DP model, each phage is allowed to infect multiple hosts (**Fig. 3b**).

We first assessed the predictive performance of phiNODE and the other two deep learning models (ResNet and Transformer) using synthetic datasets generated by the GLV and DP models. The mean PCCs for all deep learning models increased with the ratio of training samples to the total number of phage and prokaryotic taxa (N_{train}/N_{taxa}). When trained on all available samples with the largest N_{train}/N_{taxa}, all deep learning models accurately predicted prokaryotic abundances across samples, achieving mean PCC values of greater than 0.89 on the GLV-generated data (phiNODE: 0.958, ResNet: 0.890, and Transformer: 0.933) and lower values on the DP-generated data (phiNODE: 0.783, ResNet: 0.764, and Transformer: 0.786) (**Fig. 3c-d** and **Supplementary Table 2**). The relatively low mean Bray-Curtis dissimilarities between predicted and observed community profiles further indicate that these models generally capture prokaryotic composition (**Supplementary Figure 2**).

We also investigated whether prokaryotic profiles can be used to predict phage profiles by swapping the input and output in the phiNODE model (denoted as "prokaryote-to-phage" in **Fig. 3c-d**). When trained with all available samples, the mean PCCs for prokaryote-to-phage are 0.136 for the GLV model and 0.401 for the DP model, which are substantially lower than phage-to-prokaryote predictions. This underperformance likely reflects the challenge of predicting a higher-dimensional output (phage profile) from a relatively lower-dimensional input (prokaryotic profile).

To further confirm the effectiveness of the deep learning models, on one hand, we implemented a null model with the same architecture as phiNODE but trained on datasets with randomly shuffled sample labels (denoted as "Null" in **Fig. 3c-d**). Across all sample sizes from both GLV and DP models, the null model consistently yielded mean PCCs below 0.035, showing no improvement as N_{train}/N_{taxa} increased. On the other hand, we tested whether similar phage profiles could predict similar prokaryotic profiles by training k-nearest neighbors (KNN) regressors. For each query phage profile, the model selects its K closest phage profile neighbors and predicts the prokaryotic profile as the distance-weighted average of those neighbors' prokaryotic profiles. Across varying N_{train}/N_{taxa}, the KNN models yield mean PCCs of only 0.08-0.15 for GLV and 0.28-0.59 for DP, indicating that simple inter-profile similarity alone cannot reliably predict prokaryotic profiles (denoted "KNN" in **Fig. 3c-d**). In contrast, the deep learning models consistently achieved higher mean PCCs, demonstrating their effectiveness to predict prokaryotic profiles from phage profiles, especially when sufficient training samples are available.

Finally, we evaluated whether explicitly modeling inter-prokaryotic interactions could improve prokaryotic profile prediction. Because only the GLV model incorporates inter-prokaryotic interactions, we tested this hypothesis on GLV-generated synthetic data. We appended a graph convolutional network layer to the output of phiNODE, using a known prokaryotic interaction graph to smooth raw predictions before generating the final profiles (see **Methods**). Incorporating these inter-prokaryotic relationships yielded a

154 minor but statistically significant increase in mean PCC (P-value < 0.05; **Supplementary Figure 3**).
155 However, due to the marginal improvement observed in the synthetic data and the absence of high
156 confidence inter-prokaryotic interaction networks for empirical datasets, we did not apply this
157 interaction-aware approach to real-world data, although our results demonstrate its potential utility.

158 **phiNODE enables phage-host interaction predictions on synthetic data**

159 Next, we investigated whether phage-host interactions could be inferred from well-trained deep learning
160 models using sensitivity analysis (see **Methods**). Notably, we observed that the F1-score, precisions, and
161 recalls of all models increase as the ratio $N_{\text{train}}/N_{\text{taxa}}$ increases. With the largest $N_{\text{train}}/N_{\text{taxa}}$, the models
162 accurately reconstruct phage-host interactions, achieving F1-scores no less than 0.94 on the GLV-generated
163 data (phiNODE: 0.966, ResNet: 0.943, and Transformer: 0.949). Since the DP model allows phages to
164 infect multiple hosts, we conducted multi-host predictions for all deep learning models by selecting the
165 thresholds for S_{ji} that maximize the F1-scores across different $N_{\text{train}}/N_{\text{taxa}}$ ratios. Compared to the
166 performances on the GLV model, lower F1-scores were observed on the DP-generated data (phiNODE:
167 0.837, ResNet: 0.837, and Transformer: 0.825), reflecting the complexity of the DP model (**Fig. 3e-f**). With
168 the largest $N_{\text{train}}/N_{\text{taxa}}$, setting an optimal S_{ji} threshold of approximately 3.74 for phiNODE multi-host
169 prediction significantly improved its performance compared to single-host prediction, with precision
170 increasing from 0.924 to 0.955, recall from 0.670 to 0.745, and the F1-score from 0.777 to 0.837
171 (**Supplementary Figure 4 and Supplementary Table 3**). Again, as expected, the null model achieved F1-
172 score below 0.06, which was the lowest among all methods.

173 We also applied the "prokaryote-to-phage" method with different top phage numbers and optimal
174 thresholds. Although the precisions were moderate, the F1-scores were notably lower (0.61 for the GLV
175 model; 0.66 for the DP model in **Supplementary Figure 5**). This result demonstrates that predicting phage
176 profiles from prokaryotic profiles is less effective than the phage-to-prokaryote predictions.

177 In addition to the null model and "prokaryote-to-phage" model, deep learning models were
178 compared with a co-abundance-based model, FastSpar[40]. While FastSpar with a threshold of ±0.5
179 demonstrated very high precision across different $N_{\text{train}}/N_{\text{taxa}}$ ratios, it generally exhibited extremely low
180 recall on the GLV-generated data (**Supplementary Figure 6a**), tending to reconstruct phage-host
181 interactions more accurately for the prevalent prokaryotic species (**Supplementary Figure 7**). When the
182 threshold was adjusted to ±0.1, FastSpar's recall improved with small sample sizes but subsequently
183 decreased to below 0.25 as N_{train} exceeded 1,200 (**Supplementary Figure 6a**). However, when evaluated
184 on the DP-generated data, FastSpar with a threshold of ±0.5 exhibited a substantial reduction in precision,
185 with an average of 0.51 (**Supplementary Figure 6b**). In contrast, FastSpar with a threshold of ±0.1 yielded
186 higher recall but at the expense of precision, averaging only 0.16 and suffered from high false positive rates

(**Supplementary Figure 6b**). It is evident that FastSpar was highly sensitive to the chosen thresholds. Therefore, we selected the thresholds that maximized the F1-scores across different sample sizes for FastSpar. This resulted in a modest and slight improvements in F1-scores on GLV and DP model-generated data, respectively (**Fig. 3e-f** and **Supplementary Figure 6a-b**). It is worth mentioning that the optimal thresholds determined for the largest N_{train}/N_{taxa} were ±0.037 for the GLV model and ±0.048 for the DP model with the largest sample size. These thresholds are substantially lower than FastSpar's default of 0.1, potentially leading to high false positive rates, particularly in DP model-based synthetic data (**Supplementary Figure 6a-b**, the last panels). These findings highlight the limitations of co-abundance-based methods and underscore the superior predictive performance of deep learning-based models to infer phage-host interactions.

Furthermore, phiNODE outperformed alternative methods in both area under the precision-recall curve (AUPRC) and area under the receiver operating characteristic curve (AUROC), identifying more ground-truth interactions and reaching its peak performance with fewer training samples (**Fig. 3e-f**, **Supplementary Figure 8**). On GLV-generated data, phiNODE achieved the best performance, correctly predicting the hosts of 925 out of 958 phages (96.6%). On DP-generated data, it identified hosts for 919 phages (92.4%) using single-host predictions and for 899 phages (90.4%) using multi-host predictions, corresponding to 919 and 1,022 of the 1,371 ground-truth interactions, respectively. Consequently, phiNODE was selected as the optimal deep-learning model for all subsequent analyses.

phiNODE uncovers 90% more phage-host interactions than the assembly-based approach in microbiome samples from healthy individuals

To evaluate phiNODE on real-world data, we downloaded a large-scale dataset comprising 7,016 metagenomic samples of healthy individuals obtained from curatedMetagenomicData[56] (**Supplementary Table 4**). Taxonomic profiles were generated at the genus level rather than the species level. This decision was informed by our synthetic data analysis, which showed that the F1-score plateaus when N_{train}/N_{taxa} exceeds 1.0 (**Fig. 3e-f**, **Supplementary Table 3**). Using genus-level taxonomies, we achieved a training sample size (N_{train} = 5,612), which is 2.14 times the number of phage and prokaryotic taxa (N_{taxa} = 2,620) (**Supplementary Figure 9a**), and hence is sufficient to ensure reliable interaction prediction. This reliability was further corroborated by a PCC of 0.793 obtained by phiNODE on the test dataset (**Supplementary Figure 9b**). We evaluated phiNODE's sensitivity scores across a range of thresholds and observed that the number of inferred interactions stabilized (relative change < 0.15% over five consecutive thresholds) at a cutoff of 5.0. At this threshold, we recovered 1,906 genus-level phage-host interactions (**Supplementary Figure 10, Supplementary Table 4**).

To compare with existing methods, we derived phage-host interactions using a conventional assembly-based approach[57]. Specifically, for each sample, contigs were assembled and classified as either viral or non-viral. Non-viral contigs were subsequently binned to generate MAGs of high to medium quality, while viral contigs of medium or higher quality were retained. Phage-host interactions were then inferred from genomic features using a strategy similar to that described by Nayfach *et al*. (see **Methods**), which relies on matching prokaryotic CRISPR spacers to viral contigs or performing local alignments between prokaryotic MAGs and viral contigs[57]. This approach yielded over 2,500 MAG-to-contig interactions. To ensure a fair comparison with phiNODE, both prokaryotic MAGs and viral contigs were annotated at the genus level, resulting in the identification of 221 phage-host genus-level interactions (**Supplementary Figure 11**).

In addition to the assembly-based method, we also applied FastSpar to infer phage-host interactions using two correlation thresholds: ±0.10 (the default setting) and ±0.09 (the threshold at which interaction counts stabilize) (**Supplementary Figure 10**). At ±0.10, FastSpar identified 8,111 interactions spanning 285 phage genera; at ±0.09, it detected 9,227 interactions across 334 phage genera, averaging 28 host genera per phage genus. When we compared the overlap of interactions by the assembly-based method, phiNODE, FastSpar (±0.10), and FastSpar (±0.09), we observed 48, 101, and 102 shared interactions, corresponding to positive predictive values (PPVs) of 2.52%, 1.25%, and 1.11%, respectively. Overlapping the phage-host interactions from the Metagenomic Gut Virus (MGV) catalogue[30] with those from phiNODE, FastSpar (±0.10), and FastSpar (±0.09) yielded 239, 363, and 396 common interactions, with PPVs of 12.54%, 4.48%, and 4.29%, respectively (**Fig. 4a**, **Supplementary Figure 12a**). Specialist-generalist analysis revealed that 71% of phages identified by phiNODE are specialists, whereas FastSpar predominantly classifies 73% of phages as generalists with more than three host taxa (**Fig. 4b**, **Supplementary Figure 12b**). The much higher proportion of generalist phages predicted by FastSpar may reflect a high false-positive rate, consistent with our benchmarking results using synthetic data (**Fig. 3e,f**). Notably, phiNODE identified 89.4% more interactions than the assembly-based method, potentially suggesting its enhanced sensitivity to dynamic, ecologically relevant phage-host relationships; however, further validation is warranted.

Based on phiNODE's predictions, we reconstructed a phage-host interaction network and identified 20 densely connected modules (**Fig. 4c**). Several of these modules exhibited clear biological relevance. For instance, module 18 is enriched in pathogenic genera such as *Klebsiella*, *Escherichia*, and *Shigella*. Module 13 comprises lactic acid-producing cocci, including *Streptococcus*, *Lactococcus*, and *Enterococcus*. Module 9 contains the age-associated Actinobacteria genera *Bifidobacterium* and *Collinsella* [58–60]. Module 7 includes *Sutterella* and *Oscillibacter*, both previously linked to obesity [61–64]. Additionally, modules 4, 5,

and 20 are dominated by *Ruminococcus*, *Bacteroides*, and *Prevotella*, respectively, which are genera that define the three gut enterotypes [65].

The network also revealed that a single prokaryotic host genus can be infected by phages from multiple families, suggesting diverse phages can interact with a relatively narrow host taxonomic range. Furthermore, the prokaryotic hosts in all predicted interactions were predominantly affiliated with the *Firmicutes* phylum, while phages primarily belonged to the *Caudovirales* order, which includes the tailed phage families *Myoviridae*, *Siphoviridae*, and *Podoviridae* (**Fig. 4d**).

Among the 239 interactions shared by phiNODE and the MGV catalogue, only 39 overlapped with those identified by the assembly-based method (**Fig. 4a**). The remaining 200 exhibited significantly lower, yet predominantly non-zero, abundances in both phage and host genera across samples (**Supplementary Figure 13**), suggesting that phiNODE is more sensitive in detecting low abundant phage-host interactions. Among these, 154 interactions were missed by FastSpar.

To examine whether genomic features could verify these interactions, we randomly selected seven of strong phiNODE sensitivity values (≥ 9.0). phiNODE ranked the *Pseudoruminococcus* genus as the top host for phage genus mgv_g_4001208 despite a weak FastSpar correlation of 0.03 and failed viral contig quality control. Consistently, this association was confirmed by seven *Pseudoruminococcus* genomes in the MGV catalogue and the prophage features found in the host contig (**Fig. 5a**). A similar pattern was observed for mgv_g_4001910-*Treponema* and mgv_g_4000151-*Bifidobacterium*, where phiNODE's high values (10.37 and 9.15, respectively) were supported by MGV records, despite poor assembly-based and FastSpar predictions. The viral and prokaryotic contigs, which were relatively short, aligned well to high-quality phage and host genomes, thereby substantiating these phage-host interactions (**Figs. 5b-c**). Interactions mgv_g_4000293-*Sutterella* and mgv_g_4000868-*Megamonas* were not detected by the assembly-based method due to poor prokaryotic binning but supported by high phiNODE sensitivity values and phage-host genome alignments (**Figs. 5d-e**). In addition, the MGV catalogue lists multiple host genera for these phages—*Sutterella* and *Duodenibacillus* for mgv_g_4000293, and *Megamonas* and *Mitsuokella* for mgv_g_4000868—yet phiNODE narrowed and prioritized specific host genera in the healthy human gut. Finally, both phiNODE and MGV catalogue uncover the interactions *Brussowvirus*-*Streptococcus* and mgv_g_4001791-*Megamonas*. However, no assembled phage contigs were detected in these cases, highly likely due to extremely low abundances of the phage genera across samples ($7.44 \times 10^{-6} \pm 1.81 \times 10^{-4}$ (mean±standard deviation) for *Brussowvirus* and $2.43 \times 10^{-6} \pm 6.52 \times 10^{-5}$ for mgv_g_4001791) (**Fig. 5f-g**).

Together, these findings demonstrate phiNODE's ability to detect genomically supported, low-abundance, and environment-specific phage-host interactions that both FastSpar and assembly-based methods overlook.

phiNODE latent representations informed by phage-host interactions can be utilized to predict human host phenotypes

Due to insufficient healthy individual stool samples from curatedMetagenomicData for reliably inferring species-level phage-host interactions, which would require over 30,000 samples, we explored alternative applications of phiNODE using species-level microbial profiles. Numerous studies have demonstrated the prediction of human host phenotypes using features derived solely from prokaryotic profiles[66,67]. To our knowledge, no prior study has attempted to predict human host phenotypes using phage-inclusive profiles or by incorporating phage-host interactions. phiNODE inherently learns phage-host interactions by predicting prokaryotic profiles from phage profiles through its internal layers (**Fig. 6a**). We hypothesize that these internal latent representations can serve as informative features for predicting human host phenotypes.

Given that human host gender and age metadata were available for nearly all stool samples (98.8% for gender and 100% for age in **Supplementary Figure 9e-f** and **Supplementary Table 5**), we applied the phiNODE-derived latent representations along with other microbial abundance features to three classifiers (Random Forest (RF); Logistic Regression (LR); Support Vector Machine (SVM)) to predict different gender and age groups. Notably, significant differences in Shannon diversities for both prokaryotic and phage compositions were observed between different gender and age groups (**Supplementary Figure 9g-h**), suggesting that phages may also contribute to distinguishing between these groups. Two representations were extracted from the trained phiNODE model: one from a hidden layer closer to the input (representation 1, repr1) and one from a hidden layer closer to the output (representation 2, repr2) (**Fig. 6a**).

Among all model-feature combinations, the SVM with phiNODE repr2 and repr1 achieved the highest average AUROC of 0.943 and the highest mean F1-score of 0.886, respectively, using five-fold cross validation (**Supplementary Figure 14a-b**). Regarding age prediction, where the data were categorized into five qualitative groups based on established criteria[66], the SVM with phiNODE repr1 and repr2 achieved the best average AUROC of 0.950 and the best mean F1-score of 0.777, respectively, across all strategies (**Supplementary Figure 14c-d**). Collectively, these results demonstrate that phiNODE-derived latent representations exhibit higher predictive power for human host phenotypes than other types of features.

313 **phiNODE latent representations serve as promising features for disease classification**

314 Building on the promising results of using phiNODE-derived latent representations for phenotype
315 prediction in healthy individuals, we next evaluated whether these representations could also enhance
316 disease classification. While numerous attempts have been made to differentiate cases from controls using
317 microbial features, most studies consider these features in isolation without incorporating cross-domain
318 data or interactions, specifically phage-host interactions[68,69]. To demonstrate phiNODE's potential in
319 clinical diagnosis, we collected ten shotgun metagenomic datasets representing twelve diseases, including
320 inflammatory bowel disease (IBD), urinary tract infection (UTI), liver cirrhosis (LC), irritable bowel
321 syndrome (IBS), COVID-19 infection, hypertension (HTN), colorectal cancer (CRC), multiple sclerosis
322 (MS), human immunodeficiency virus (HIV) infection and Parkinson's disease (PD). Cross-domain
323 taxonomic profiles at the species level for both prokaryotes and phages were generated for all datasets
324 (**Supplementary Figure 15a**), which were used to train phiNODE and extract two latent representations
325 (**Fig. 6a**).

326 We first assessed differences between cases and controls across diseases using permutational
327 multivariate analysis of variance (PERMANOVA). In PERMANOVA, the F statistic represents the ratio of
328 the variance between groups to the variance within groups, with higher values indicating greater inter-
329 groups variance than that of intra-groups. We found that the F values for phiNODE representations (either
330 repr1 or repr2) in 7 out of 10 datasets are higher than those obtained using prokaryotic features, phage
331 features, or combined features (**Supplementary Figure 16**). Similar to the previous analysis conducted for
332 the healthy individual dataset, we next used these features as input variables for three classifiers (RF, LR,
333 and SVM) to predict disease conditions. Compared with alternative features, either repr1 or repr2 achieved
334 the highest average AUROCs from five-fold cross-validation in 8 out of 12 diseases (**Fig. 6b**,
335 **Supplementary Figure 17**), and they obtained the highest average F1-scores in 7 diseases and
336 (**Supplementary Figure 18** and **Supplementary Table 6**). Among all models, LR most consistently
337 achieved the best AUROCs. We also noticed that the diseases for which phiNODE-derived representations
338 outperformed other features in PERMANOVA did not always align with those showing superior
339 performance in machine learning prediction, suggesting that phiNODE representations may better capture
340 non-linear and subtle patterns than traditional statistical tests.

341 We next investigated why phiNODE-derived representations underperformed in four diseases:
342 colorectal cancer (CRC), multiple sclerosis (MS), human immunodeficiency virus (HIV) infection, and
343 Parkinson's disease (PD). We examined potential factors and found that it could not be attributed to case-
344 control imbalance. Actually, except for HIV, COVID-19, and hypertension (HTN), most datasets
345 maintained balanced case-control ratios (**Supplementary Figure 15b**). Additionally, differences in model

346 fitting are unlikely to be the cause, as the mean PCCs for the four underperforming diseases were
347 comparable to those for the others; notably, MS had a mean PCC of 0.73, the second highest among all
348 diseases (**Supplementary Figure 15c**). We then examined whether alpha diversity differs between cases
349 and controls when using prokaryotic and phage profiles. For diseases showing significant differences in
350 both profiles, such as IBD-CD, IBD-UC, UTI, and LC, phiNODE-derived representations achieved the
351 highest mean AUROCs among all features (**Fig. 6c**, the first row). For diseases without significant
352 differences between cases and controls using any profiles, phiNODE representations still outperformed
353 other features in four diseases (IBS-C, IBS-D, COVID-19, and HTN) (**Fig. 6c**, the second row), whereas
354 prokaryotic abundance-based features performed the best in the remaining diseases, HIV and PD (**Fig. 6c**,
355 right two panels on the third row). Another scenario where phiNODE underperformed was when significant
356 differences were found only in prokaryotic profiles but not in phage profiles (**Fig. 6c**, left two panels on the
357 third row). In other words, phiNODE-derived representations consistently outperform other features in
358 disease classification when both prokaryotic and phage profile Shannon diversities differ significantly
359 between cases and controls.

360 Notably, phiNODE's latent representations reduce dimensionality by roughly two orders of
361 magnitude compared to phage or combined abundance-based features. They were the smallest in seven
362 datasets and the second smallest in the remaining three (**Supplementary Figure 15d**). This substantial
363 reduction in feature numbers, especially compared to combined abundance features, highlights a more
364 efficient approach of integrating cross-domain species data. Overall, our findings demonstrate that latent
365 representations learned from phiNODE outperform traditional species-based features in predicting most
366 disease states, underscoring their potential as robust, low-dimensional biomarkers for future research.

367 DISCUSSION

368 Existing assembly-based methods for inferring phage-host interactions from shotgun metagenomic data
369 tend to miss many interactions because assembled genomes poorly represent low-abundance phage taxa[32–
370 34]. Moreover, these methods rely heavily on known molecular mechanisms, such as infection processes or
371 host defense systems, which limits their ability to discover novel phage-host relationships. In this study, we
372 developed phiNODE, a deep learning framework that predicts prokaryotic profiles from paired phage
373 profiles. By interpreting the trained model, we infer phage-host interactions in a purely data-driven fashion
374 without relying on any predefined biological mechanisms, potentially capturing more phage-host ecological
375 dynamics. When tested on synthetic data, phiNODE outperformed alternative deep learning approaches as
376 well as co-abundance-based methods. When applied to a large, cross-cohort collection of >7,000 healthy
377 individual stool samples, phiNODE identified more plausible phage-host interactions and a higher

proportion of specialist phages than the co-abundance approach. More importantly, it recovered substantially more interactions than the traditional assembly-based pipeline, which enabled a biologically meaningful reconstruction and modulization of the healthy gut phage-host interaction network.

Although phiNODE currently infers interactions at the genus level—primarily due to sample size constraints of existing datasets—this limitation does not appear to hinder its utility for disease prediction. In our disease cohorts, microbial profiles were generated at the species level, and the lowest training sample-to-feature-ratio (N_{train}/N_{taxa}) was approximately 0.015, far below the conventional 1.0 threshold for reliable interaction inference. More importantly, phiNODE-derived representations consistently outperform other relative abundance-based features in disease prediction when both prokaryotic and phage profile-based Shannon diversities differ significantly between cases and controls and perform the best in most cases even when such differences are not significant. These results indicate that phiNODE's latent representations could capture subtle, predictive signals beyond simple taxonomic shifts and offer potential for non-invasive diagnosis across diverse diseases.

Despite its advantages, we acknowledge several limitations of the current framework. First, phiNODE does not distinguish temperate from virulent phages—a distinction that may require genomic context or auxiliary sequencing data beyond abundance information. Second, although synthetic benchmarks yielded higher Pearson correlations than real-world datasets, real samples can be influenced by inter-prokaryotic interactions —which we showed in synthetic data can be incorporated into phiNODE once a reliable interaction matrix becomes available—as well as nutrient gradients, host immune factors, and potential false-positive taxonomic assignments, all of which warrant further study. Third, the model is restricted to DNA phages and bulk metagenomic data, suggesting its extension to RNA phages or virus-like particle (VLP) metagenomes will depend on the availability of sufficient metatranscriptomic and VLP enrichment datasets. Finally, because our predictions operate at the genus level, larger and more targeted sample collections, especially from specific populations or disease states, will be required to achieve finer taxonomic resolution.

In summary, by leveraging the vast repositories of existing metagenomic data, phiNODE provides a scalable, mechanism-agnostic approach to phage-host interaction inference. With appropriate reference databases, this framework can be extended to microbiota from other human body sites or environments. phiNODE offers an efficient and robust alternative to traditional assembly- and co-abundance-based methods for reconstructing phage-host interactions and predicting host metadata. Its flexible architecture also holds promise for broader applications in other cross-domain interaction predictions.

METHODS

Synthetic data generated using ecological models

Generalized Lotka-Volterra model.

We first generated synthetic data for a pool containing N prokaryotes and M phages using generalized Lotka-Volterra (GLV) population dynamics. We defined the GLV model as below[48,49]:

$$\frac{dx_i(t)}{dt} = x_i(t)\left[r_i + \sum_{j=1}^{N+M} a_{ij}x_j(t)\right], i = 1, \dots, N+M,$$

where $x_i(t)$ denotes the abundance of the i-th prokaryote/phage at time $t \geq 0$. The GLV model has two major parameters: the interaction matrix $A = (a_{ij}) \in \mathbb{R}^{(N+M)\times(N+M)}$, and the intrinsic growth-rate vector $r = (r_i) \in \mathbb{R}^{(N+M)}$. The parameter a_{ij} denotes the interaction strength of microorganism i to the per-capita growth rate of microorganism j. The parameter r_i is the intrinsic growth rate of microorganism i. The prokaryote-prokaryote interaction strengths were randomly choosing from $a_{b_i b_j} \sim Bernoulli(C) \cdot Normal(0, \sigma)$ if $b_i \neq b_j$ and $a_{b_i b_i}$ =-1, where $C = 0.1$ and $\sigma = 0.05$. For phage-host interactions, we assume each prokaryotic host can be targeted by 10 phages and the interaction from prokaryotes to phages were selected from $a_{b_i p_j} \sim Normal(\mu, \sigma)$ where $\mu = 0.9$ and $\sigma = 0.1$. The interaction from phages to prokaryotes were selected from $a_{p_i b_j} \sim Normal(\mu, \sigma)$ where $\sigma = 0.1$, $\mu = -0.5$ for 20% of the phages, $\mu = 0.5$ for the remaining 80%. The GLV model is designed to reflect the narrow host ranges of phages. We assume extremely strong inter-phage competition, such that when sampling from the phage pool, only a single and randomly selected phage taxon can infect a given prokaryotic host. From a pool consisting of $N = 100$ prokaryotes and $M = 1{,}000$ phages, we generated 200 to 1,800 samples in steps of 100 using GLV model, with each pair of samples shared 80% of prokaryotes. After confirming that all microorganisms had reached steady state at a simulation time of T = 1000, we extracted paired phage and prokaryotic profiles from each sample.

To evaluate the impact of inter-prokaryotic interactions, we generated between 200 and 1,800 synthetic samples by setting the GLV connectivity parameter to $C = 0.9$ and the interaction-strength standard deviation to $\sigma = 0.25$, while holding all other parameters constant.

Dynamic population model.

Next, we generated synthetic data for a pool comprising N prokaryotes and M phages by employing a more sophisticated dynamic population (DP) model, adapted from the "more comprehensive model including sensitive prokaryotes (namely the S4 model)" proposed by Kimchi et al.[50]. Compared to the GLV model, our DP model resembles a Susceptible-Infected-Recovered (SIR) model in epidemiology and considers a variety of phage-host interaction processes including infection, lysis/lysogeny, induction, reversion and phage death. This model also permits a prokaryotic host that is already infected by a certain phage to acquire a second infection by another phage, yet it cannot be infected by a third one. In detail, the population densities of sensitive prokaryotes S_i, phage P_j and their associated lysogens L_{ij} change according to the following three equations, respectively:

$$\frac{dS_i}{dt} = \alpha S_i + r\sum_j L_{ij} - \beta S_i \sum_j P_j a_{ij} \quad (1)$$

$$\frac{dP_j}{dt} = b\gamma \sum_i L_{ij} + \beta\big(b(1-f_j)P_j - P_j\big)\left(\sum_i S_i\, a_{ij} + \sum_i \sum_{j' \neq j} L_{ij'}\, a_{ij'}\right) - \delta P_j \quad (2)$$

$$\frac{dL_{ij}}{dt} = \alpha L_{ij} + \beta f_j S_i P_j a_{ij} - (\gamma + r)L_{ij} - \beta L_{ij} \sum_{j' \neq j} (1 - f_{j'})P_{j'}\, a_{ij'} \quad (3)$$

Here, $A = (a_{ij}) \in \mathbb{R}^{(N+M)\times(N+M)}$ is a pre-defined binary matrix denoting if phage j is capable to infect prokaryote i, where we assume that a prokaryotic host can be infected by multiple phage predators and a phage can also target multiple prokaryotic hosts. α indicates the prokaryotic growth rate, while β indicates the phage infectivity. γ represents the prophage induction rate and δ represents the phage death rate. b denotes the burst size, the number of new virus particles released from a single infected host cell when it is lysed. r is defined as the reversion rate for a lysogen to lose a prophage. f is the fraction of phages' infections leading to lysogeny, where phage with $f = 0$ are obligate lytic and have no associated lysogens. f is randomly chosen for each phage from a list of [0.0, 0.2, 0.4, 0.6, 0.8], representing different phage lysogeny abilities. Details and literature supports for the parameter settings are provided in **Supplementary Table 7**. Equation (1) defines that the population density of i-th uninfected prokaryote is increased by self-growth and reversion from infected prokaryotes L_{ij}, while decreased by the infection of phage j if their interaction is set to non-zero in the matrix A. Equation (2) defines that the population density of j-th phage is positively affected by the new phage particles produced by the burst from prophage induction and from the direct lysis of sensitive prokaryotes and those infected by other phages. The density is negatively affected by the phage infection and integration into available prokaryotic genomes (including sensitive prokaryotes and prokaryotes infected by other phages) and its own death. Equation (3) defines the

463 population density of i-th prokaryote infected and integrated by j-th phage in its genome (also known as
464 lysogens). The density increases due to self-prokaryotic growth and phage j infecting sensitive prokaryote
465 i, while decreases due to phage induction, its conversion into sensitive prokaryote and lysis by other phages
466 j' (where $j' \neq j$). From a pool comprising $N = 100$ prokaryotes and $M = 1{,}000$ phages, we generated
467 between 200 and 1800 samples in increments of 100 using DP model, with each sample containing exactly
468 10 prokaryotes and each pair of samples shared 60% of prokaryotes. In each round of sampling, a random
469 number of phages was selected for a specific prokaryote. After simulating for a longer period (T=5,000
470 steps) than the GLV model, instead of reaching a steady state, the phage and prokaryotic abundances may
471 still fluctuate in a chaotic way but maintain a dynamic balance. In fact, this chaotic dynamic population
472 model was first proposed to explain the coexistence of lytic and temperate phages in nature[50].

473 ## Metagenomic sample collection and processing

474 ### Healthy individual dataset.

475 This datasets include 7,016 metagenomic stool samples of sequencing depth greater than 1 million reads
476 from healthy individuals ranging in age from 8-75 years old without antibiotics usage. The sample
477 accessions were collected from curatedMetagenomicData[56] and downloaded from Sequence Read Archive
478 (SRA) database. According to previously reported criterion, age data were grouped into five categories
479 including children adolescents ($8 \leqslant$ age $\leqslant 18$), young adult ($18 <$ age $\leqslant 35$), middle aged ($35 <$ age $\leqslant 50$),
480 senior ($50 <$ age $\leqslant 65$), and elderly (age > 65). Metadata can be found in **Supplementary Table 5**.

481 ### Disease datasets.

482 Human fecal raw metagenomic sequencing data and clinical metadata of the samples for (1) IBD and
483 controls were collected and downloaded from the Inflammatory Bowel Disease Multi'omics Database
484 (IBDMDB) (https://ibdmdb.org/results) including HMP2 and HMP2 Pilot studies[70,71], PRJEB15371[72],
485 PRJNA1086048[73], PRJNA400072[74] and PRJNA429990[75]. (2) CRC and controls were from PRJDB4176[76],
486 PRJEB10878[77], PRJEB12449[78], PRJEB27928[79], PRJEB6070[80], PRJEB7774[81], PRJNA389927[82],
487 PRJNA429097[83], PRJNA447983[84], PRJNA531273 (with PRJNA397112 as controls)[85], PRJNA731589[86],
488 PRJNA763023[87] and PRJNA936589 (downloaded from
489 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA936589/ since no paper was published). The metadata of
490 CRC were partially organized from Zun et al.[88] (3) Multiple sclerosis (MS) and controls were from
491 PRJEB51635 (patients) and PRJEB41786/PRJEB41787 (controls)[89]. (4) Parkinson's disease (PD) and
492 controls were from PRJEB53403 and PRJEB53401[90] and PRJNA834801[91]. (5) Liver cirrhosis (LC) and
493 controls were obtained from PRJEB6337[92], PRJNA373901[93] and from a paired set where patients are from

494 PRJNA912122 and controls from PRJNA838648[94]. (6) Urinary tract infection (UTI) and controls were
495 from PRJNA400628[95]. (7) Irritable bowel syndrome (IBS) and controls were collected from PRJEB37924[96].
496 (8) COVID19 and controls were obtained from PRJDB13214[97] and PRJNA650244[98]. (9) HIV and controls
497 were obtained from PRJNA307231[99], PRJNA692830[100], PRJNA391226[101], and PRJNA450025[102]. (10)
498 Hypertension (HTN) and controls were from PRJEB13870[103]. More details in metadata can be available in
499 **Supplementary Table 8**.

500 Metagenomic sequencing reads were quality controlled with Trimmomatic v.0.39[104] with options
501 "ILLUMINACLIP:<adapter file>:2:30:10:2:TRUE LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
502 MINLEN:30" and host contamination was removed using Kraken v.2.1.2[105]. Phanta v.1.1.0 was
503 subsequently executed with default parameters and the prophage-masked version of database provided by
504 Phanta v.1.1.0[32]. Specifically, to further filter false positive species, different viral coverage threshold and
505 prokaryotic coverage threshold have been selected based on the sample sequencing depths following the
506 suggestions at https://github.com/bhattlab/phanta/tree/main. Phage and prokaryotic profiles were derived
507 from the Phanta generated relative taxonomic profiles. To only include viruses infecting prokaryotic hosts
508 (namely phages), only viral taxa belonging to the orders *Caudovirales*, *Tubulavirales*, *Petitvirales*,
509 *Cremevirales*, and *Halopanivirales* were considered.

510 Deep learning model architectures and hyperparameter tuning

511 We evaluated and compared three representative deep learning models: Transformer[106], Residual Neural
512 Network (ResNet)[107] and Neural Ordinary Differential Equations (NODE)[47]. We included Transformer
513 because of its robust attention mechanism and its ability to capture long-range dependencies within the
514 data[106]. ResNet was selected due to its use of skip connections, which facilitate the training of deeper
515 architectures by ensuring direct gradient flow and mitigating the vanishing gradient problem[107]. NODE
516 extends its precursor, ResNet, by integrating skip connections that are analogous to Euler steps in ODE
517 solvers[47]. Each residual block in NODE represents a small time-change of variables, inspiring its
518 continuous-depth architecture. Unlike ResNet's discrete hidden layers, NODE parameterizes the derivative
519 of the hidden state using a neural network and computes outputs through an ODE solver. This methodology
520 allows NODE to maintain constant memory, adapt its evaluation strategy for each input, and balance
521 numerical precision with speed.

522 phiNODE.

523 phiNODE is a method that integrates an implicit NODE module within a deep learning architecture. Prior
524 to modeling, the centered log-ratio transformation is applied to both the phage-excluded and phage
525 taxonomic abundance profiles. The architecture comprises three sequential modules: (1) A fully connected

layer maps the input phage compositions to a hidden representation of dimension N_h, followed by the hyperbolic tangent (tanh) activation function, (2) a neural ODE module approximates the first-order time derivative using a two-layer multilayer perceptron (MLP) (each layer with dimension N_h) to implement the NODE[47], with each layer followed by the tanh activation function, and (3) a fully connected layer transforms the hidden representation to yield the phage-excluded output profiles. The Adam optimizer[108] is used for the gradient descent. To prevent overfitting, we used (1) the L2 regularization with the weight decay parameter λ in the Adam optimizer, (2) early stopping in the training. Training stops if the validation mean squared error (MSE) loss does not decrease by at least 1×10^{-4} or by at least 1% relative to the best observed loss over 20 consecutive epochs. Three hyperparameters are tunned based on the validation loss: the dimension of the hidden layer N_h, the L2 regularization parameter λ and the learning rate l. Specifically, N_h is selected from the three candidate values closest to the number of output features within the set [4, 8, 16, 32, 64, 128, 256, 320, 512, 640, 1024, 2048, 2560, 4096], λ is chosen from the range of 10^{-6} to 10^{-2} and l is selected from 10^{-8} to 10^{-5}. The optimal hyperparameter combination was determined using the Optuna v.4.2.1 framework in 10 trials[109].

To enable model explicitly to consider inter-prokaryotic interactions, we appended a graph convolutional network (GCN) layer defined by

$$Y = \left(D'^{-\frac{1}{2}}(A+I)D'^{-\frac{1}{2}}\right)XW + b$$

where $X \in \mathbb{R}^{B\times N}$ is the batch of phiNODE originally predicted prokaryotic abundances (B: batch size, N: taxa), $A \in \mathbb{R}^{N\times N}$ is the normalized Laplacian of the inter-prokaryotic interaction network, and I is the identity matrix that adds self-loops. The diagonal matrix D' has entries $D'_{ii} = \sum_j A'_{ij}$, so that $\hat{A} = D'^{-\frac{1}{2}}(A+I)D'^{-\frac{1}{2}}$ symmetrically normalizes the graph. The smoothed feature matrix $\hat{A}X$ is mapped back to the output space by the learnable weight $W \in \mathbb{R}^{N\times N}$ and bias $b \in \mathbb{R}^N$, yielding $Y \in \mathbb{R}^{B\times N}$, which is the final inter-prokaryotic interaction-aware profile predictions. This operation ensures that each taxon's final prediction is informed by both its own abundance and those of its interacting partners.

Residual Neural Network.

The Residual Neural Network (ResNet) is implemented as a residual neural network to learn complex representations while mitigating the vanishing gradient problem. The architecture is organized into three sequential modules: (1) An input layer maps the input phage compositions to a hidden representation of dimension N_h, (2) a series of hidden layers, each implemented as a fully connected layer, is arranged with residual connections: the output of each hidden layer is added to its input, and the sum is passed through a

ReLU activation function, and (3) an output layer projects the final hidden representation to the output dimensionality. Hyperparameter tunning (the dimension of the hidden layer N_h, the L2 regularization parameter λ and the learning rate l and early stopping were implemented using the same strategy as that employed for phiNODE.

Transformer.

The transformer model is designed to capture contextual dependencies in input data through self-attention mechanisms. Its architecture is composed of the following modules: (1) a linear layer projects the input features from their original dimension to a higher-dimensional space d_{model}, where the embedded input is then reshaped (by adding an extra dimension) for compatibility with subsequent layers, (2) a multi-layer transformer encoder (layer numbers determined by N_{layer}), constructed from identical encoder layers, that applies multi-head self-attention and feedforward operations (with dropout regularization $r_{dropout}$) to extract rich representations from the embedded inputs, (3) a corresponding multi-layer transformer decoder (layer numbers determined by N_{layer}) that takes a target tensor (initialized to zeros) along with the encoder's output (memory) to generate context-aware features with each decoder layer uses multi-head attention ($N_{head} = \max(2,\ \mathrm{int}(d_{model}/8))$) and feedforward networks ($d_{feedforward} = 2 \cdot d_{model}$), and (4) a linear layer transforms the decoder's output into the final output with the desired dimensionality. To reinforce the original signal and stabilize training, a residual connection is implemented by linearly mapping the original input directly to the output and summing it with the decoder's result. A total of five hyperparameter are tuned based on the validation loss: for the model dimension d_{model}, the number of encoder/decoder layers N_{layer}, the dropout rate $r_{dropout}$, the L2 regularization parameter λ and the learning rate l. Specifically, d_{model} is selected from the three candidate values closest to the number of output features within the set [4, 8, 16, 32, 64, 128, 256, 320, 512, 640, 1024, 2048, 2560, 4096], N_{layer} is chosen from [2, 4, 6, 8], $r_{dropout}$ is selected from [0.3, 0.4, 0.5], while other hyperparameters were tuned following a strategy analogous to that used for phiNODE. Model construction, training and validation are implemented using PyTorch v2.4.1[110].

Model training, validation and testing

For synthetic and real-world datasets, deep learning models including phiNODE, ResNet and transformer were trained on 80% of samples (as training set). Their hyperparameters were heuristically determined on the 10% of the samples (as validation set) and the mean PCCs of all prokaryotic taxa across samples on the remaining 10% samples (as test set) were calculated to evaluate the models' abilities to predict prokaryotic compositions.

Inferring phage-host interactions via sensitivity analysis

The well-trained phiNODE takes the relative abundance of phage i (x_i) as inputs and generate predictions for the relative abundance of prokaryote j (y_j). For the sample m in the training set, $x_i^{(m)}$ for all i is provided as the input vector to the trained phiNODE and phiNODE can predict the relative abundances $y_j^{(m)}$ for all prokaryotes. To investigate the influence of phage i on prokaryote j for a particular sample m, we perturb $x_i^{(m)}$ by setting it as the mean of relative abundances for phage i across samples (denoted as $\bar{x}_i$) while keeping values of $x_{j\neq i}^{(m)}$ intact. Thus, the perturbation amount for phage i is $\Delta x_i^{(m)} = \bar{x}_i - x_i^{(m)}$. This perturbed vector for phage relative abundances is provided to the trained phiNODE to regenerate predictions for prokaryotic relative abundance $y_j^{(m)} + \Delta y_j^{(m)}$ where $\Delta y_j^{(m)}$ is the deviation of newly predicted relative abundance for prokaryote j when phiNODE uses the perturbed input vector from that when it uses the unperturbed input vector. For the sample m, the sensitivity of prokaryote j to phage i can be defined as $S_{ji}^{(m)} = \frac{\Delta y_j^{(m)}}{\Delta x_i^{(m)}}$. Sensitivity value is averaged across samples and z-score normalized for each phage. The overall sensitivity of prokaryote j to phage i is $\tilde{S}_{ji} = \frac{S_{ij} - \bar{S}_i}{\sigma_i}$, where $S_{ji} = \frac{\sum_m S_{ji}^{(m)}}{N_{train}}$ and N_{train} is the number of samples in the training set. $\bar{S}_i$ and σ_i are the mean and standard deviation, respectively, of S_{ji} calculated for phage j.

Network analysis and visualization

The phage-host interaction network was visualized using Cytoscape v.3.10.3[111]. Modules were identified using the community clustering algorithm GLay[112], which is available as an app within the Cytoscape.

Phage-host interactions at the genus level in the Metagenomic Gut Virus catalogue

The phage-host associations in the Metagenomic Gut Virus (MGV) catalogue were sourced from https://portal.nersc.gov/MGV/MGV_v1.0_2021_07_08/mgv_host_assignments.tsv (accessed on January 24th, 2025). In this dataset, host genomes from the Unified Human Gastrointestinal Genome (UHGG) database[113] were linked to viruses via a combination of CRISPR spacer matches and near-identical 1 kb BLAST hits between phage and prokaryotic genomes[36]. As the taxonomy for prokaryotic hosts is based on the Genome Taxonomy Database (GTDB r89), we converted GTDB taxonomy to NCBI taxonomy at the genus level using gtdb_to_taxdump (https://github.com/nick-youngblut/gtdb_to_taxdump)[114].

Predict phage-host interactions using co-abundance-based method

614 FastSpar v.1.0.0 was used to calculate the co-abundance correlations between the prokaryotic and phage
615 taxa[40]. P-values were generated by performing 1000 random permutations. Statistically significant ($P < 0.05$)
616 correlations with magnitude exceeding certain thresholds were considered as potential phage-host
617 interactions.

618 Infer phage-host interactions using the assembly-based method

619 For the 7,016 metagenomic stool samples from healthy individuals, we performed sample-wise assembly
620 using MEGAHIT v.1.2.9[115], after excluding contigs less than 2,000 bp, resulting in a total of 64,888,908
621 assembled contigs. Viral and non-viral sequences were differentiated using geNomad v.1.7.6[27], which
622 excluded provirus and plasmid sequences, yielding 62,053,011 non-viral contigs and 2,835,897 viral
623 contigs. Non-viral contigs were binned using MetaBAT v.2.15[116], producing 206,411 bins, which were
624 further assessed for quality using CheckM v.1.2.3[117], retaining those with ⩾50% completeness and <10%
625 contamination, leading to 90,447 metagenome-assembled genomes (MAGs). Taxonomic classification of
626 prokaryotic bins was performed using CAT pack v.6.0.1[118] with HumGut as the reference database[119],
627 identifying 45 prokaryotic genera. CRISPR spacers were detected in the MAGs using CasCRISPRTyper
628 v.1.8.0[120], yielding 56,461 CRISPR spacers. Viral contigs were quality-filtered using CheckV v1.0.3[121],
629 retaining 96,599 high- to medium-quality viral contigs, which were then assigned taxa uhgv-tools
630 (https://github.com/snayfach/UHGV/blob/main/CLASSIFY.md) with MGV as the reference database[30],
631 identifying 1,947 viral genera. To predict host-phage interactions, we followed the strategy described by
632 Nayfach et al.[57] and employed CRISPR spacer matching via blastn-short (allowing ⩽1 mismatch) and
633 sequence similarity-based matching via BLASTN v.2.15.0+[122], considering hits with >90% identity
634 and >500 bp overlap. These approaches identified 2,589 host MAG-viral contig interactions, resolving into
635 221 host-virus interactions at the genus level.

636 Machine learning classifiers

637 The machine learning models were trained and validated using a nested cross-validation framework to
638 ensure robust model selection and performance evaluation. Input features were standardized using
639 StandardScaler in the scikit-learn module[123]. Three classification models were evaluated: Support Vector
640 Machine (SVM), Logistic Regression (LR) and Random Forest (RF). Model selection and performance
641 assessment were conducted using a 5-fold cross-validation, with folds generated via StratifiedKFold from
642 the scikit-learn library[123]. One of the five folds was reserved as a held-out test set. This approach was
643 employed to address class imbalance in the data, ensuring that each fold preserved the original dataset's
644 class distribution and thereby providing a more robust evaluation of model performance. The remaining
645 data were subjected to hyperparameter optimization using GridSearchCV from the scikit-learn library with

an inner 5-fold cross-validation[123]. For SVM model, the hyperparameter grid included regularization values (C) of 0.1, 1, and 10 and kernel types of linear and radial basis function (rbf). For LR model, the hyperparameter grid explored different penalty types ("none", L2, and ElasticNet) and regularization strength (C) values of 0.1, 1, and 10. The hyperparameter grid for the Random Forest included the number of trees (n_estimators: 100, 300, 500) and the fraction of samples used for building each tree (max_samples: 0.2, 0.5, 0.8). Model selection was based on the highest mean area under the curve (AUC) score obtained from the inner cross-validation, with the best hyperparameters applied to the outer test fold for evaluation. Model performance was assessed using AUC, precision, recall, and F1-score, and predictions for each fold. For age data, which was treated as a multi-class classification problem, we computed one-vs-all AUC as well as macro-averaged precision, recall, and F1-score.

Data and code availability

All code for running phiNODE is available at https://github.com/YiyanYang0728/phiNODE. The scripts for data processing and figure generation can be found at https://doi.org/10.5281/zenodo.15500282.

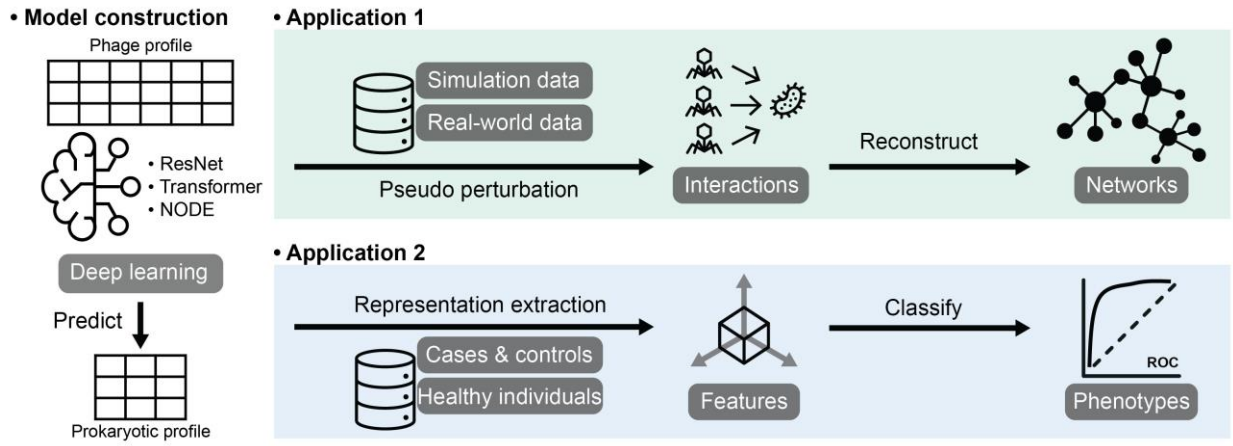


Figure 1. The framework of applying deep learning-based models to discover phage-host interactions and disease classification from metagenomic data. Different deep learning models can be trained on paired phage and prokaryotic relative taxonomic abundance profiles from the same samples, which can be utilized in two practical applications. The first is to interpret phage-host interactions by conducting sensitivity analysis that perturb the abundances of input features and predict alterations in the output. In this way, it help reveal possible interactions between the input (i.e., phage taxon) and output (i.e., prokaryotic taxon) and corresponding interaction networks. The second application involves extracting latent representations from trained models to encode informative phage-host interactions, which serve as novel features to predict human host phenotypes such as gender in healthy individuals and disease conditions between patients and healthy controls. ResNet: Residual neural network; NODE: Neural ordinary differential equations.

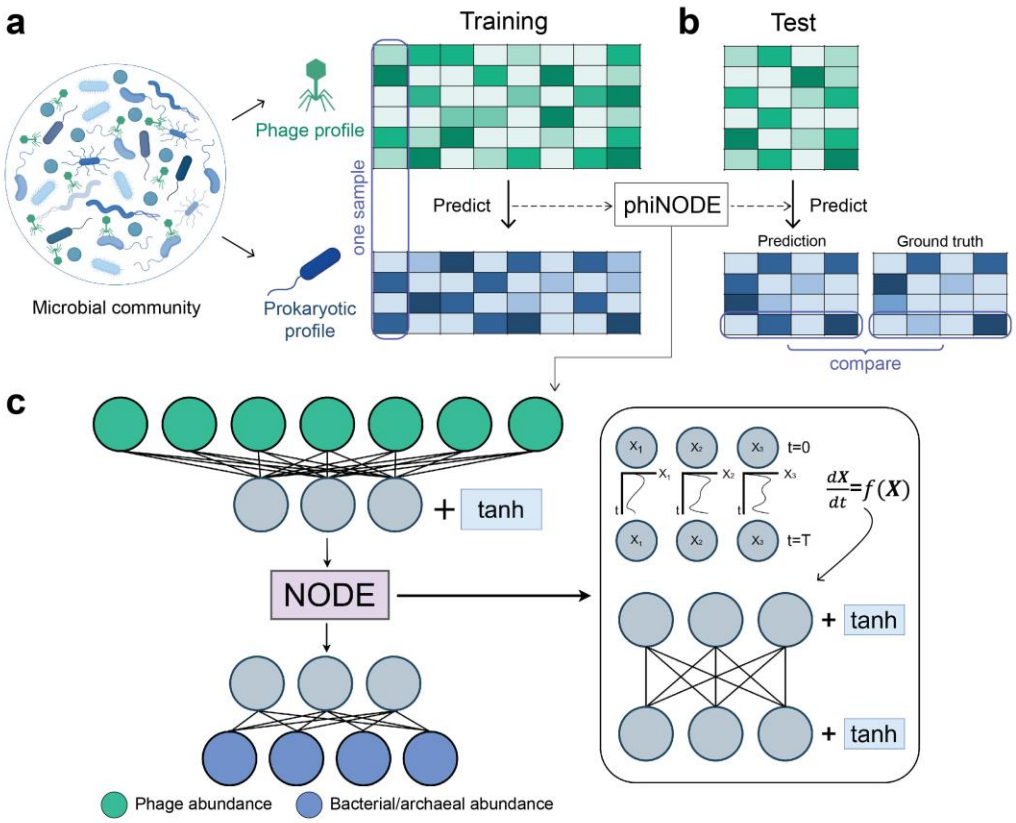


Figure 2. The model architecture and applications of phiNODE. Across all panels, green- and blue-colored matrices represent the phage profiles and prokaryotic profiles, respectively. None of the cells in the matrices has missing data and white cells mean small values. Twelve hypothetical samples with six phage taxa and four prokaryotic taxa are used to illustrate the idea. The samples are divided into training and test sets with a 2:1 ratio. **a**, Eight samples in the training set are used for training deep learning models. **b**, Fully trained phiNODE can generate predictions for prokaryotic microbial profiles in the test set. For each prokaryotic taxon, the Pearson correlation coefficient (PCC) between its predicted and true profiles across samples in the test set is computed to quantify models' performance. **c**, The architecture of phiNODE. The NODE is a module in the middle of the architecture t computes the time evolution of ODEs whose first-order time derivatives are approximated by two fully connected neural network layers. Grey nodes represent neurons in hidden layers of phiNODE.

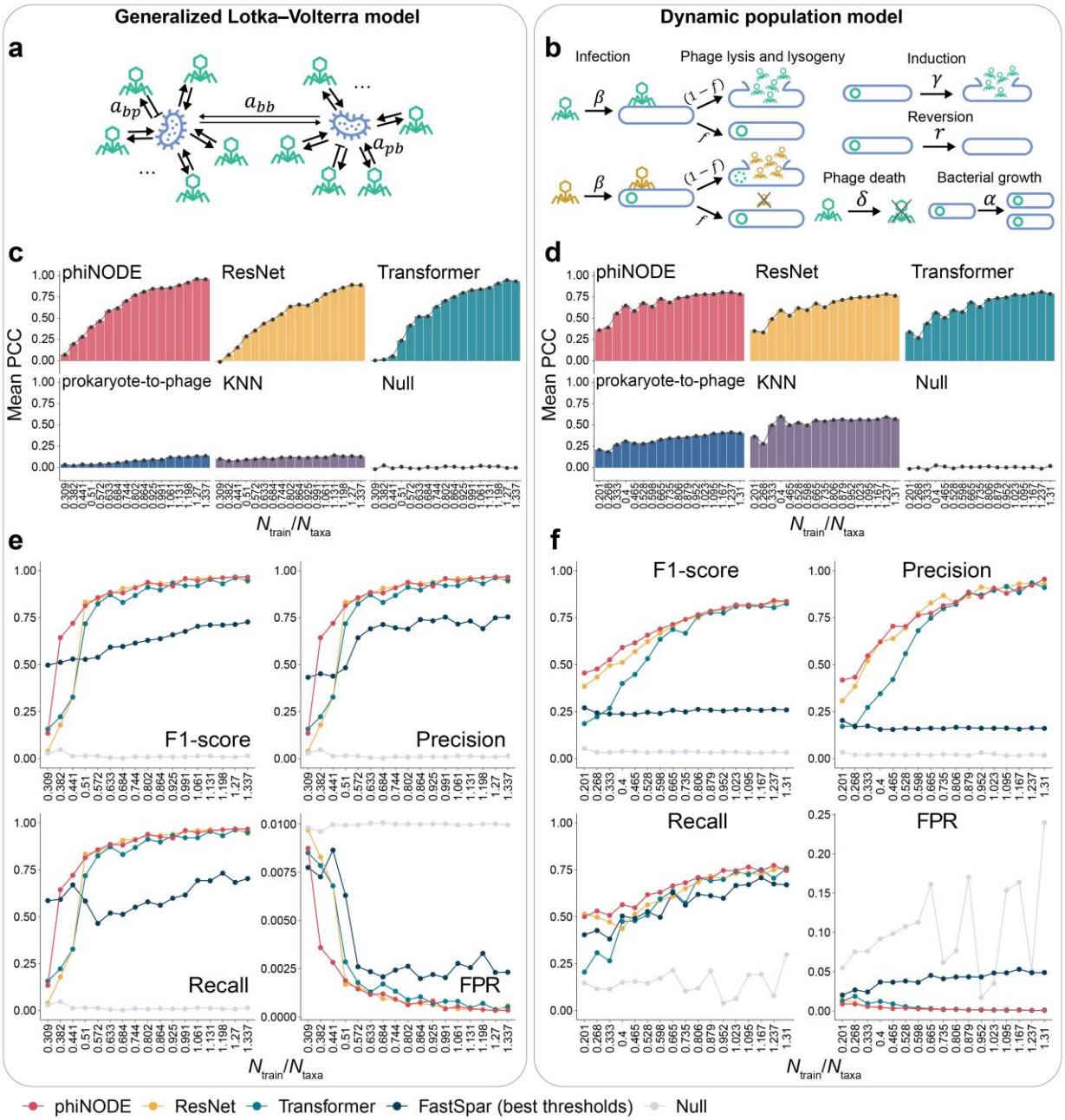
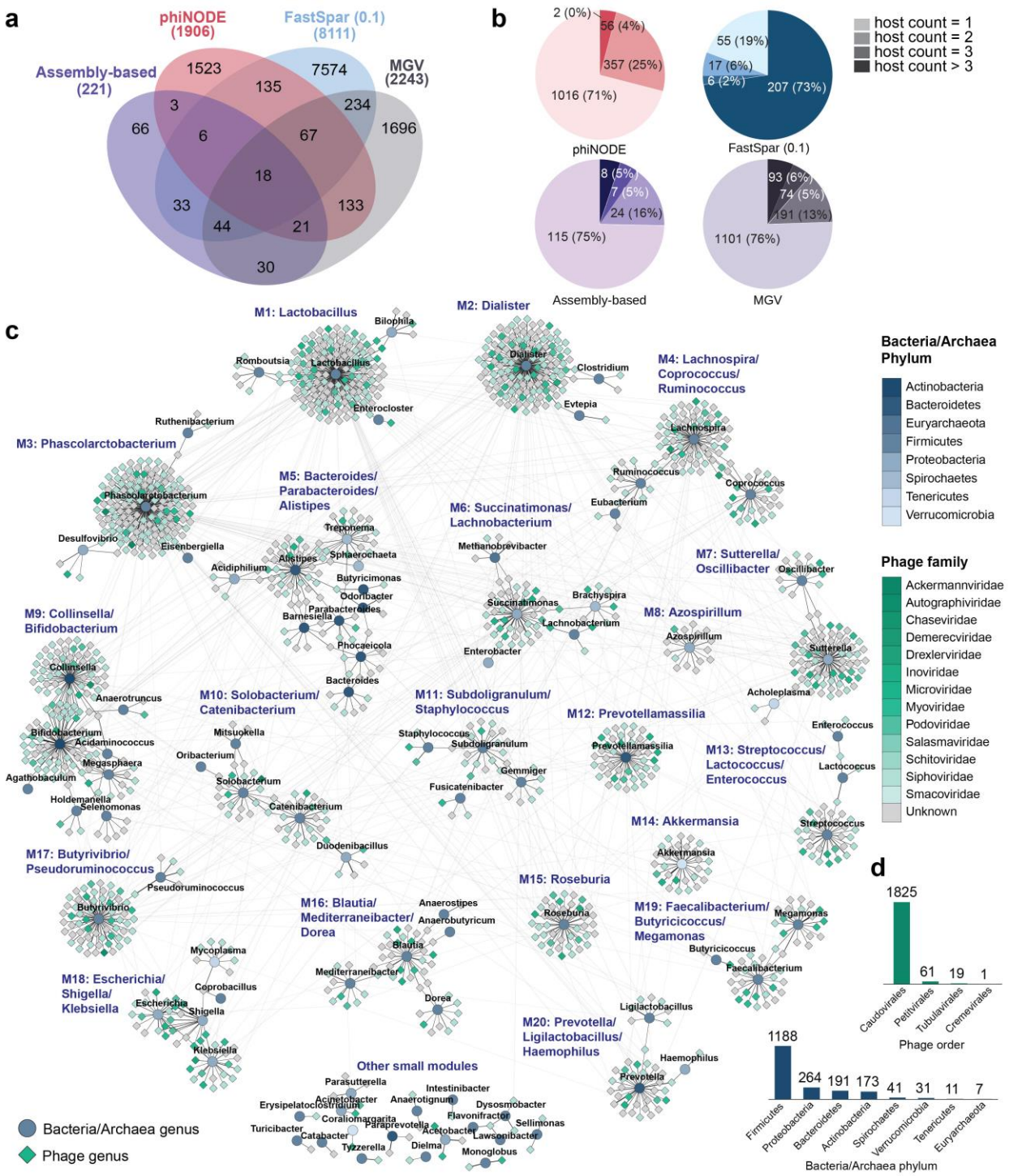


Figure 3. Validation of phiNODE using synthetic data. a, Schematic of data generation using generalized Lotka-Volterra model (GLV). a_{pb} and a_{bp} represent the phage-prokaryote interactions, while a_{bb} presents the inter-prokaryotic interactions. Each prokaryote can be infected by 10 phages and we set strict host specificity for phages, where each phage has only one host. The GLV model was simulated until all phage and prokaryotic abundances reach steady states. **b**, Schematic of data generation using dynamic population model (DP). Although each prokaryote is set to be targeted by 10 phages, phages are allowed to infect multiple hosts. The DP model was simulated until all phage and prokaryotic abundances reach dynamic

balances. Each phage has a lysogen fraction (f), indicating its infections leading to lysogeny. Phage with f=0 are obligate lytic. More details can be found in Methods section. **c-d**, Performance of phiNODE, other deep learning methods and a k-nearest neighbors (KNN) model on the test sets in simulation data from (**c**) GLV and (**d**) DP models with different ratios of training sample size/phage and prokaryote taxa number, N_{train}/N_{taxa}. The mean PCCs for all prokaryotic abundances across samples are used to reflect the models' prediction power on test sets. **e-f**, Performance of phiNODE, other deep learning methods and a co-abundance-based method (i.e., FastSpar) to predict phage-host interactions using synthetic data generated by (**e**) GLV and (**f**) DP models with varying N_{train}/N_{taxa} ratios. F1-score, precision, recall and false positive rate (FPR) were used to evaluate the model performance. Note that all deep learning models performed single-host prediction on data generated by the GLV model and multi-host prediction on data generated by the DP model. For single-host prediction on DP model data, please refer to Supplementary Figure 4.



705

706 **Figure 4. Reconstruction of phage-host interactions in healthy individuals using phiNODE. a**, Venn
707 diagram illustrating the overlaps among phage-host interactions predicted by FastSpar (using a ±0.1
708 threshold), the assembly-based method, phiNODE, and the interactions recorded in the MGV
709 (Metagenomic Gut Virus) catalogue. **b**, Pie charts illustrating the distribution of phage specificities toward

710 the hosts using different approaches. **c**, Phage-host interaction networks inferred by phiNODE, where
711 circles represent bacteria/archaea genera and diamonds represent phage genera. Phage families and
712 bacteria/archaea phyla are displayed in different colors, and the thickness of the edges indicates the
713 interaction strengths. Nodes that belong to the same modules are clustered together. Edges within modules
714 are shown in black, while those between modules are displayed in light gray. **d**, Distributions of phage
715 orders (top) and bacteria/archaea phyla (bottom) among the interactions identified by phiNODE.

716

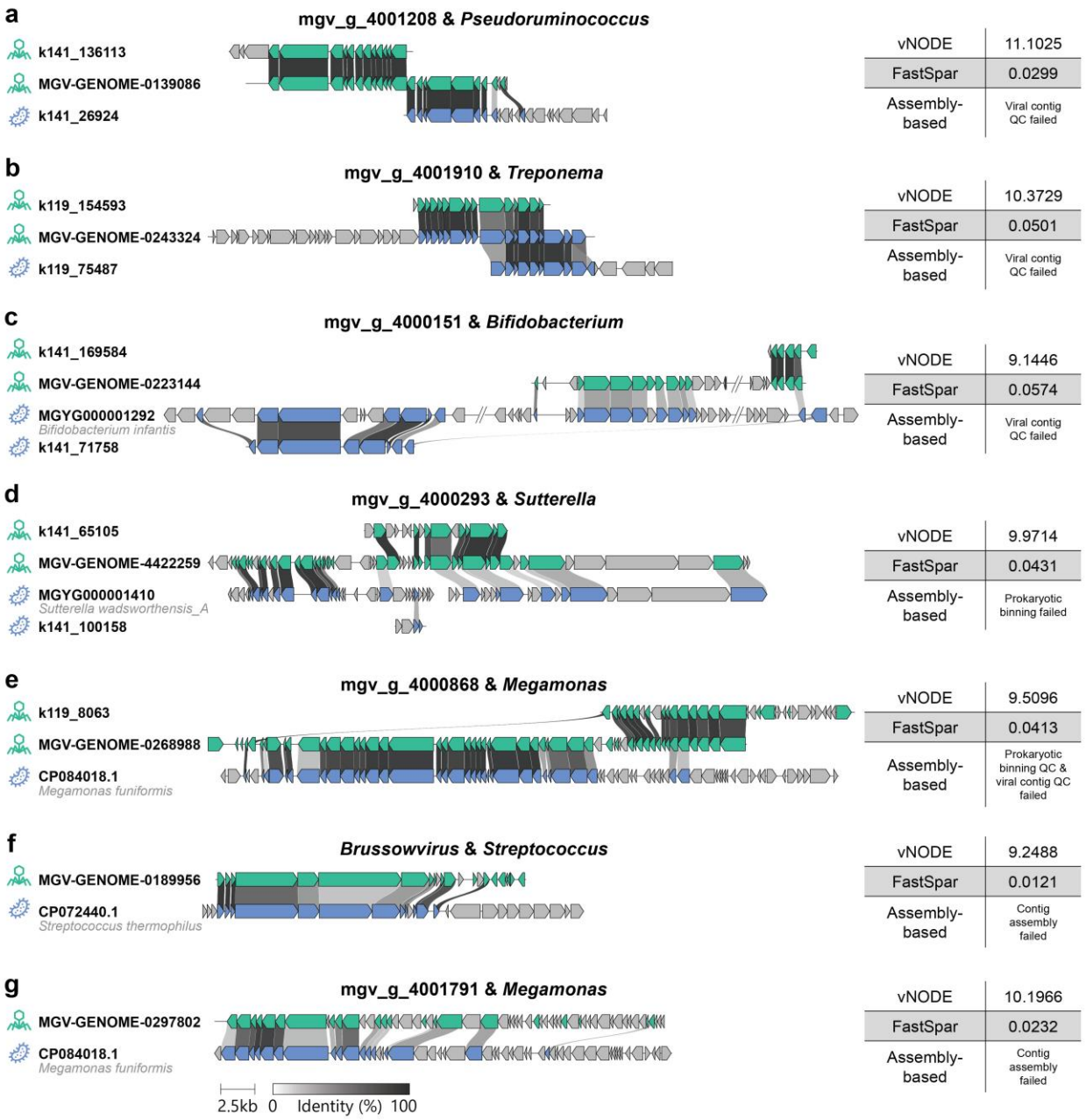


Figure 5. Examples illustrating phiNODE's higher sensitivity in reconstructing interactions supported by genomic features. a-c, Examples show interactions that were not supported by FastSpar or the assembly-based method yet are supported by the Metagenomic Gut Virus (MGV) catalogue. **d-e**, Examples illustrate the interactions were not identified by the assembly-based method due to poor prokaryotic binning. The MGV catalogue reported multiple prokaryotic host genera for the phage genera, while phiNODE prioritized host genera for these phages in the gut of healthy individuals. **f-g**, Examples demonstrates an interaction detected by phiNODE in samples that lacked any assembled phage contigs. Although neither FastSpar nor the assembly-based method identified this interaction, it is supported by the

726 MGV catalogue. The superkingdom icons, genome or contig accessions are shown to the left of each
727 sequence.

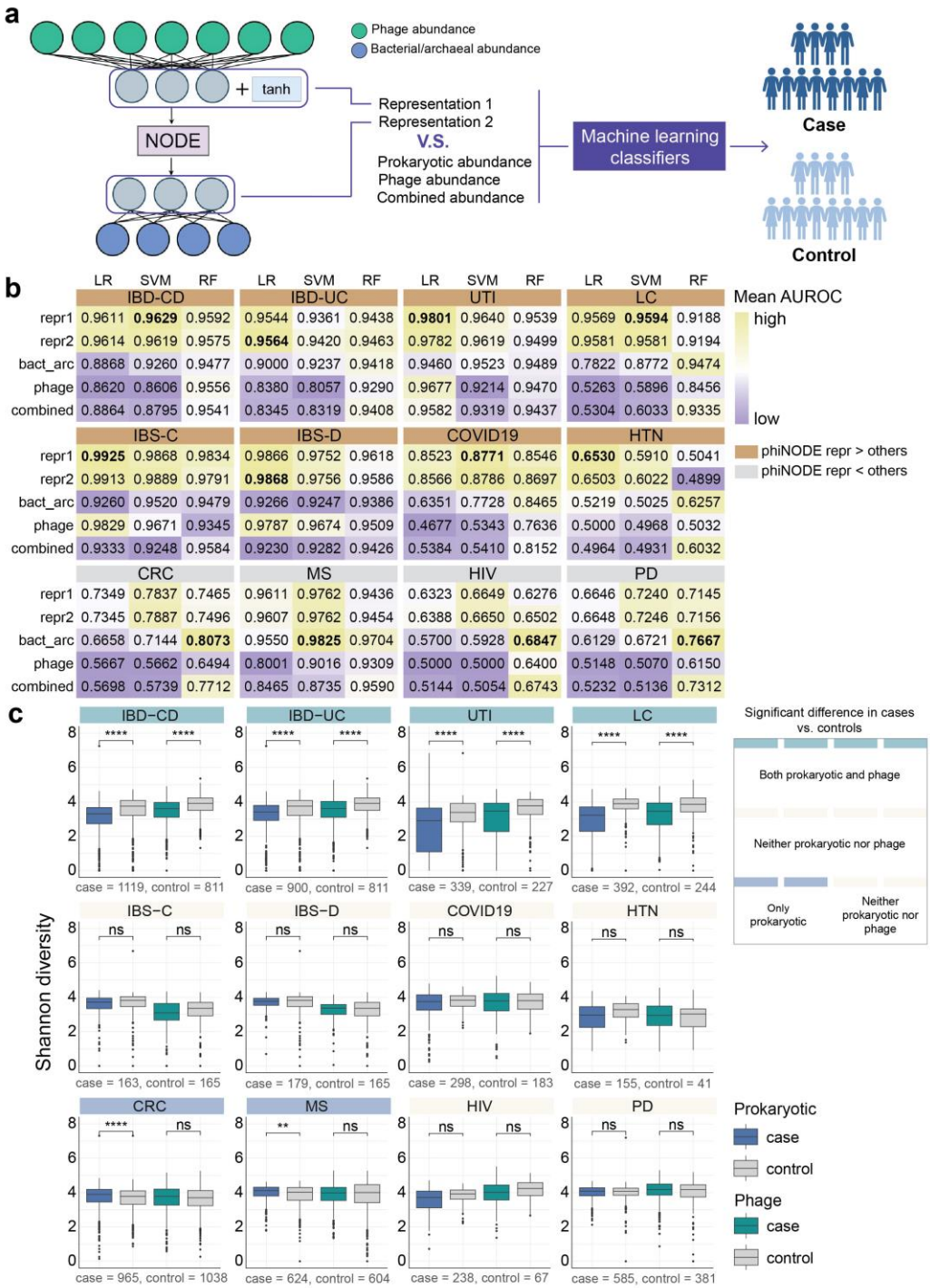


Figure 6. Application of phiNODE's latent representations for case-control classification across diseases. a, For a trained phiNODE to predict disease-related prokaryotic profiles from phage profiles, its hidden layers implicitly reflecting phage-host interactions can be regarded as latent representations, which can be fed into machine learning-based classifiers as features to distinguish healthy individuals and diseased individuals. Representation 1 (repr1) is the representation close to the input layer (phage profile) while

representation 2 (repr2) is close to the output layer (prokaryotic profile). **b**, Heatmap illustrating the mean areas under the receiver operating characteristic curve (AUROCs) from five-fold cross-validation of three machine learning classifiers using phiNODE-derived representations, prokaryotic profiles, phage profiles, and combined cross-domain profiles to classify disease conditions across 12 diseases. Higher AUROC values are shown in yellow and lower values in purple, with color scales determined individually for each disease. Green headers denote that phiNODE-derived representations achieve the best performance, whereas grey headers indicate that other features perform the best. Rows represent five feature types across different diseases, and columns represent Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) classifiers. **c**, Boxplots illustrating Shannon diversity indices based on prokaryotic (blue and gray boxes) and phage (green and gray boxes) profiles, comparing cases and controls across various diseases. Sample sizes for cases and controls for each disease are indicated at the bottom of each panel. We employed two-sided Wilcoxon rank-sum tests to assess differences between groups. Significance levels are denoted as follows: ns ($p > 0.01$), ** ($p \leqslant 0.01$), *** ($p \leqslant 0.001$), and **** ($p \leqslant 0.0001$). IBD-CD: Crohn's disease; IBD-UC: Ulcerative colitis; UTI: Urinary tract infection; LC: Liver cirrhosis; IBS-C: Irritable bowel syndrome with constipation; IBS-D: Irritable bowel syndrome with diarrhea; HTN: Hypertension; CRC: Colorectal cancer; MS: Multiple sclerosis; PD: Parkinson's disease.

751 # REFERENCES

752 1. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 103
753 (2020).
754 2. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*
755 **551**, 457–463 (2017).
756 3. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev*
757 *Microbiol* **15**, 579–590 (2017).
758 4. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359
759 (2015).
760 5. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's
761 Biogeochemical Cycles. *Science* **320**, 1034–1039 (2008).
762 6. Shreiner, A. B., Kao, J. Y. & Young, V. B. The gut microbiome in health and in disease: *Current*
763 *Opinion in Gastroenterology* **31**, 69–75 (2015).
764 7. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and
765 resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
766 8. Wu, G. *et al.* A core microbiome signature as an indicator of health. *Cell* **187**, 6550-6565.e11 (2024).
767 9. Pittayanon, R. *et al.* Differences in Gut Microbiota in Patients With vs Without Inflammatory Bowel
768 Diseases: A Systematic Review. *Gastroenterology* **158**, 930-946.e1 (2020).
769 10. Pittayanon, R. *et al.* Gut Microbiota in Patients With Irritable Bowel Syndrome—A Systematic
770 Review. *Gastroenterology* **157**, 97–108 (2019).
771 11. Chanda, D. & De, D. Meta-analysis reveals obesity associated gut microbial alteration patterns and
772 reproducible contributors of functional shift. *Gut Microbes* **16**, 2304900 (2024).
773 12. Zheng, J. *et al.* Noninvasive, microbiome-based diagnosis of inflammatory bowel disease. *Nat Med*
774 **30**, 3555–3567 (2024).

775 13. Liu, N.-N. *et al.* Multi-kingdom microbiota analyses identifybacterial–fungal interactions and
776 biomarkers of colorectal cancer acrosscohorts. *Nat Microbiol* **7**, 238–250 (2022).
777 14. Shkoporov, A. N. *et al.* The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific.
778 *Cell Host & Microbe* **26**, 527-541.e5 (2019).
779 15. Cao, Z. *et al.* The gut virome: A new microbiome component in health and disease. *eBioMedicine* **81**,
780 104113 (2022).
781 16. Scanlan, P. D. Bacteria–Bacteriophage Coevolution in the Human Gut: Implications for Microbial
782 Diversity and Functionality. *Trends in Microbiology* **25**, 614–623 (2017).
783 17. Popescu, M., Van Belleghem, J. D., Khosravi, A. & Bollyky, P. L. Bacteriophages and the Immune
784 System. *Annu. Rev. Virol.* **8**, 415–435 (2021).
785 18. Rasmussen, T. S. *et al.* Bacteriophage-mediated manipulation of the gut microbiome – promises and
786 presents limitations. *FEMS Microbiology Reviews* **44**, 507–521 (2020).
787 19. Norman, J. M. *et al.* Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel
788 Disease. *Cell* **160**, 447–460 (2015).
789 20. Nakatsu, G. *et al.* Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival
790 Outcomes. *Gastroenterology* **155**, 529-541.e5 (2018).
791 21. Gogokhia, L. *et al.* Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation
792 and Colitis. *Cell Host & Microbe* **25**, 285-299.e8 (2019).
793 22. Leal Rodríguez, C. *et al.* The infant gut virome is associated with preschool asthma risk
794 independently of bacteria. *Nat Med* **30**, 138–148 (2024).
795 23. Ritz, N. L. *et al.* The gut virome is associated with stress-induced changes in behaviour and immune
796 responses in mice. *Nat Microbiol* **9**, 359–376 (2024).
797 24. Fan, G. *et al.* Alterations in the gut virome are associated with type 2 diabetes and diabetic
798 nephropathy. *Gut Microbes* **15**, 2226925 (2023).
799 25. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA
800 viruses. *Microbiome* **9**, 37 (2021).

801 26. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of
802 microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*
803 **8**, 90 (2020).
804 27. Camargo, A. P. *et al.* Identification of mobile genetic elements with geNomad. *Nat Biotechnol* **42**,
805 1303–1312 (2024).
806 28. Johansen, J. *et al.* Genome binning of viral entities from bulk metagenomics data. *Nat Commun* **13**,
807 965 (2022).
808 29. Chen, L. & Banfield, J. F. COBRA improves the completeness and contiguity of viral genomes
809 assembled from metagenomes. *Nat Microbiol* **9**, 737–750 (2024).
810 30. Nayfach, S. *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut
811 microbiome. *Nat Microbiol* **6**, 960–970 (2021).
812 31. Roux, S. *et al.* iPHoP: An integrated machine learning framework to maximize host prediction for
813 metagenome-derived viruses of archaea and bacteria. *PLoS Biol* **21**, e3002083 (2023).
814 32. Pinto, Y., Chakraborty, M., Jain, N. & Bhatt, A. S. Phage-inclusive profiling of human gut
815 microbiomes with Phanta. *Nat Biotechnol* **42**, 651–662 (2024).
816 33. Bharti, R. & Grimm, D. G. Current challenges and best-practice protocols for microbiome analysis.
817 *Briefings in Bioinformatics* **22**, 178–193 (2021).
818 34. Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L. & Prosperi, M. Challenges in the analysis
819 of viral metagenomes. *Virus Evol* **2**, vew022 (2016).
820 35. Gregory, A. C. *et al.* The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity
821 in the Human Gut. *Cell Host & Microbe* **28**, 724-740.e8 (2020).
822 36. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive
823 expansion of human gut bacteriophage diversity. *Cell* **184**, 1098-1109.e9 (2021).
824 37. Camargo, A. P. *et al.* IMG/VR v4: an expanded database of uncultivated virus genomes within a
825 framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Research* **51**,
826 D733–D743 (2023).

827 38. Shaw, J. & Yu, Y. W. Rapid species-level metagenome profiling and containment estimation with
828 sylph. *Nat Biotechnol* (2024) doi:10.1038/s41587-024-02412-y.
829 39. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput*
830 *Biol* **8**, e1002687 (2012).
831 40. Watts, S. C., Ritchie, S. C., Inouye, M. & Holt, K. E. FastSpar: rapid and scalable correlation
832 estimation for compositional data. *Bioinformatics* **35**, 1064–1066 (2019).
833 41. Edwards, R. A., McNair, K., Faust, K., Raes, J. & Dutilh, B. E. Computational approaches to predict
834 bacteriophage–host relationships. *FEMS Microbiology Reviews* **40**, 258–272 (2016).
835 42. Michel-Mata, S., Wang, X., Liu, Y. & Angulo, M. T. Predicting microbiome compositions from
836 species assemblages through deep learning. *iMeta* **1**, e3 (2022).
837 43. Wang, X.-W. *et al.* Identifying keystone species in microbial communities using deep learning. *Nat*
838 *Ecol Evol* **8**, 22–31 (2023).
839 44. Wang, T. *et al.* Predicting metabolomic profiles from microbial composition through neural ordinary
840 differential equations. *Nat Mach Intell* **5**, 284–293 (2023).
841 45. Shtossel, O., Koren, O., Shai, I., Rinott, E. & Louzoun, Y. Gut microbiome-metabolome interactions
842 predict host condition. *Microbiome* **12**, 24 (2024).
843 46. Tisza, M. J. *et al.* Longitudinal phage–bacteria dynamics in the early life gut microbiome. *Nat*
844 *Microbiol* **10**, 420–430 (2025).
845 47. Chen, R. T., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential equations.
846 *Advances in neural information processing systems* **31**, (2018).
847 48. Ram, G. & Solomon, S. Generalized Lotka-Volterra (glv) Models. in (1997).
848 49. An Illustrated Guide to Theoretical Ecology. *Journal of Mammalogy* **82**, 247–248 (2001).
849 50. Kimchi, O., Meir, Y. & Wingreen, N. S. Lytic and temperate phage naturally coexist in a dynamic
850 population model. *The ISME Journal* **18**, wrae093 (2024).
851 51. Dikareva, E. *et al.* An extended catalog of integrated prophages in the infant and adult fecal
852 microbiome shows high prevalence of lysogeny. *Front. Microbiol.* **14**, 1254535 (2023).

853 52. Kim, M.-S., Park, E.-J., Roh, S. W. & Bae, J.-W. Diversity and Abundance of Single-Stranded DNA
854 Viruses in Human Feces. *Appl Environ Microbiol* **77**, 8062–8070 (2011).
855 53. Sausset, R., Petit, M. A., Gaboriau-Routhiau, V. & De Paepe, M. New insights into intestinal phages.
856 *Mucosal Immunology* **13**, 205–215 (2020).
857 54. Hoyles, L. *et al.* Characterization of virus-like particles associated with the human faecal and caecal
858 microbiota. *Research in Microbiology* **165**, 803–812 (2014).
859 55. Styles, K. M., Brown, A. T. & Sagona, A. P. A Review of Using Mathematical Modeling to Improve
860 Our Understanding of Bacteriophage, Bacteria, and Eukaryotic Interactions. *Front. Microbiol.* **12**,
861 724767 (2021).
862 56. Pasolli, E. *et al.* Accessible, curated metagenomic data through ExperimentHub. *Nat Methods* **14**,
863 1023–1024 (2017).
864 57. Nayfach, S. *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol* **39**, 499–509 (2021).
865 58. Arboleya, S., Watkins, C., Stanton, C. & Ross, R. P. Gut Bifidobacteria Populations in Human Health
866 and Aging. *Front. Microbiol.* **7**, (2016).
867 59. Tzemah-Shahar, R. *et al.* Signs of aging in midlife: physical function and sex differences in
868 microbiota. *GeroScience* **46**, 1477–1488 (2023).
869 60. Torma, F. *et al.* Alterations of the gut microbiome are associated with epigenetic age acceleration and
870 physical fitness. *Aging Cell* **23**, e14101 (2024).
871 61. Squillario, M. *et al.* Gut-microbiota in children and adolescents with obesity: inferred functional
872 analysis and machine-learning algorithms to classify microorganisms. *Sci Rep* **13**, 11294 (2023).
873 62. Chen, X. *et al.* Characterization of the gut microbiota in Chinese children with overweight and
874 obesity using 16S rRNA gene sequencing. *PeerJ* **9**, e11439 (2021).
875 63. Bischoff, S. C. *et al.* Gut Microbiota Patterns Predicting Long-Term Weight Loss Success in
876 Individuals with Obesity Undergoing Nonsurgical Therapy. *Nutrients* **14**, 3182 (2022).
877 64. Hu, H.-J. *et al.* Obesity Alters the Microbial Community Profile in Korean Adolescents. *PLoS ONE*
878 **10**, e0134333 (2015).

879 65. MetaHIT Consortium (additional members) *et al.* Enterotypes of the human gut microbiome. *Nature*
880 **473**, 174–180 (2011).
881 66. Monshizadeh, M., Hong, Y. & Ye, Y. Multitask knowledge-primed neural network for predicting
882 missing metadata and host phenotype based on human microbiome. *Bioinformatics Advances* **5**,
883 vbae203 (2024).
884 67. Shen, J., Zhang, D. & Liang, B. Prediction of host age and sex classification through gut microbes
885 based on machine learning. *Biochemical Engineering Journal* **178**, 108280 (2022).
886 68. Akiyama, S. *et al.* Multi-biome analysis identifies distinct gut microbial signatures and their crosstalk
887 in ulcerative colitis and Crohn's disease. *Nat Commun* **15**, 10291 (2024).
888 69. Clooney, A. G. *et al.* Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory
889 Bowel Disease. *Cell Host & Microbe* **26**, 764-778.e5 (2019).
890 70. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases.
891 *Nature* **569**, 655–662 (2019).
892 71. Zhang, Y. *et al.* Discovery of bioactive microbial gene products in inflammatory bowel disease.
893 *Nature* **606**, 754–760 (2022).
894 72. Mitchell, A. L. *et al.* EBI Metagenomics in 2017: enriching the analysis of microbial communities,
895 from sequence reads to assemblies. *Nucleic Acids Research* **46**, D726–D735 (2018).
896 73. Zheng, J. *et al.* Noninvasive, microbiome-based diagnosis of inflammatory bowel disease. *Nat Med*
897 **30**, 3555–3567 (2024).
898 74. Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel
899 disease. *Nat Microbiol* **4**, 293–305 (2018).
900 75. Weng, Y. J. *et al.* Correlation of diet, microbiota and metabolite networks in inflammatory bowel
901 disease. *J of Digest Diseases* **20**, 447–459 (2019).
902 76. Yachida, S. *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes
903 of the gut microbiota in colorectal cancer. *Nat Med* **25**, 968–976 (2019).

904 77. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive
905 biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).
906 78. Vogtmann, E. *et al.* Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-
907 Genome Shotgun Sequencing. *PLoS ONE* **11**, e0155362 (2016).
908 79. Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are
909 specific for colorectal cancer. *Nat Med* **25**, 679–689 (2019).
910 80. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular*
911 *Systems Biology* **10**, 766 (2014).
912 81. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma–carcinoma sequence. *Nat*
913 *Commun* **6**, 6528 (2015).
914 82. Zackular, J. P., Rogers, M. A. M., Ruffin, M. T. & Schloss, P. D. The Human Gut Microbiome as a
915 Screening Tool for Colorectal Cancer. *Cancer Prevention Research* **7**, 1112–1121 (2014).
916 83. Yang, J. *et al.* Establishing high-accuracy biomarkers for colorectal cancer by comparing fecal
917 microbiomes in patients with healthy families. *Gut Microbes* **11**, 918–929 (2020).
918 84. Thomas, A. M. *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-cohort
919 microbial diagnostic signatures and a link with choline degradation. *Nat Med* **25**, 667–678 (2019).
920 85. Gupta, A. *et al.* Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut
921 Microbiome of Colorectal Cancer Patients in India. *mSystems* **4**, 10.1128/msystems.00438-19 (2019).
922 86. Liu, N.-N. *et al.* Multi-kingdom microbiota analyses identifybacterial–fungal interactions and
923 biomarkers of colorectal cancer acrosscohorts. *Nat Microbiol* **7**, 238–250 (2022).
924 87. Yang, Y. *et al.* Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun*
925 **12**, 6757 (2021).
926 88. Zuo, W., Michail, S. & Sun, F. Metagenomic Analyses of Multiple Gut Datasets Revealed the
927 Association of Phage Signatures in Colorectal Cancer. *Front. Cell. Infect. Microbiol.* **12**, 918010
928 (2022).

929 89. Thirion, F. *et al.* The gut microbiota in multiple sclerosis varies with disease activity. *Genome Med*
930 **15**, 1 (2023).
931 90. Boktor, J. C. *et al.* Integrated Multi-Cohort Analysis of the Parkinson's Disease Gut Metagenome.
932 *Movement Disorders* **38**, 399–409 (2023).
933 91. Wallen, Z. D. *et al.* Metagenomics of Parkinson's disease implicates the gut microbiome in multiple
934 disease mechanisms. *Nat Commun* **13**, 6958 (2022).
935 92. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64 (2014).
936 93. Loomba, R. *et al.* Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of
937 Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metabolism* **25**, 1054-1062.e5
938 (2017).
939 94. Odenwald, M. A. *et al.* Bifidobacteria metabolize lactulose to optimize gut metabolites and prevent
940 systemic infection in patients with liver disease. *Nat Microbiol* **8**, 2033–2049 (2023).
941 95. Worby, C. J. *et al.* Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent
942 urinary tract infections in women. *Nat Microbiol* **7**, 630–639 (2022).
943 96. Mars, R. A. T. *et al.* Longitudinal Multi-omics Reveals Subset-Specific Mechanisms Underlying
944 Irritable Bowel Syndrome. *Cell* **182**, 1460-1473.e17 (2020).
945 97. Nagata, N. *et al.* Human Gut Microbiota and Its Metabolites Impact Immune Responses in COVID-
946 19 and Its Complications. *Gastroenterology* **164**, 272–288 (2023).
947 98. Yeoh, Y. K. *et al.* Gut microbiota composition reflects disease severity and dysfunctional immune
948 responses in patients with COVID-19. *Gut* **70**, 698–706 (2021).
949 99. Noguera-Julian, M. *et al.* Gut Microbiota Linked to Sexual Preference and HIV Infection.
950 *EBioMedicine* **5**, 135–146 (2016).
951 100. Bai, X. *et al.* Whole-Genome Metagenomic Analysis of the Gut Microbiome in HIV-1-Infected
952 Individuals on Antiretroviral Therapy. *Front. Microbiol.* **12**, 667718 (2021).
953 101. Lu, W. *et al.* Association Between Gut Microbiota and CD4 Recovery in HIV-1 Infected Patients.
954 *Front. Microbiol.* **9**, 1451 (2018).

102. Rocafort, M. *et al.* Evolution of the gut microbiome following acute HIV-1 infection. *Microbiome* **7**, 73 (2019).

103. Li, J. *et al.* Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome* **5**, 14 (2017).

104. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

105. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257 (2019).

106. Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30**, (2017).

107. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, Las Vegas, NV, USA, 2016). doi:10.1109/CVPR.2016.90.

108. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. Preprint at https://doi.org/10.48550/ARXIV.1412.6980 (2014).

109. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2623–2631 (ACM, Anchorage AK USA, 2019). doi:10.1145/3292500.3330701.

110. Paszke, A. *et al.* Automatic differentiation in PyTorch. (2017).

111. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).

112. Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135–3137 (2010).

113. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* **39**, 105–114 (2021).

114. Youngblut, N. & Shen, W. nick-youngblut/gtdb_to_taxdump: Zenodo release. Zenodo https://doi.org/10.5281/ZENODO.3696964 (2020).

115. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* **31**, 1674–1676 (2015).

116. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).

117. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

118. Von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol* **20**, 217 (2019).

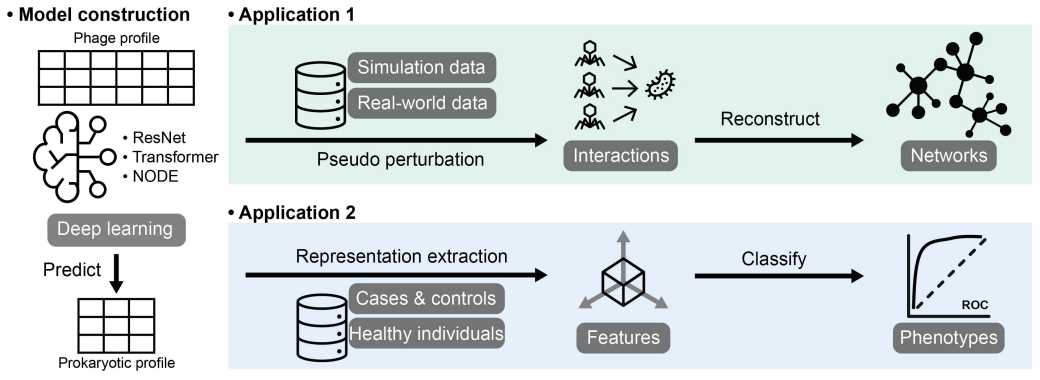
119. Hiseni, P., Rudi, K., Wilson, R. C., Hegge, F. T. & Snipen, L. HumGut: a comprehensive human gut prokaryotic genomes collection filtered by metagenome data. *Microbiome* **9**, 165 (2021).

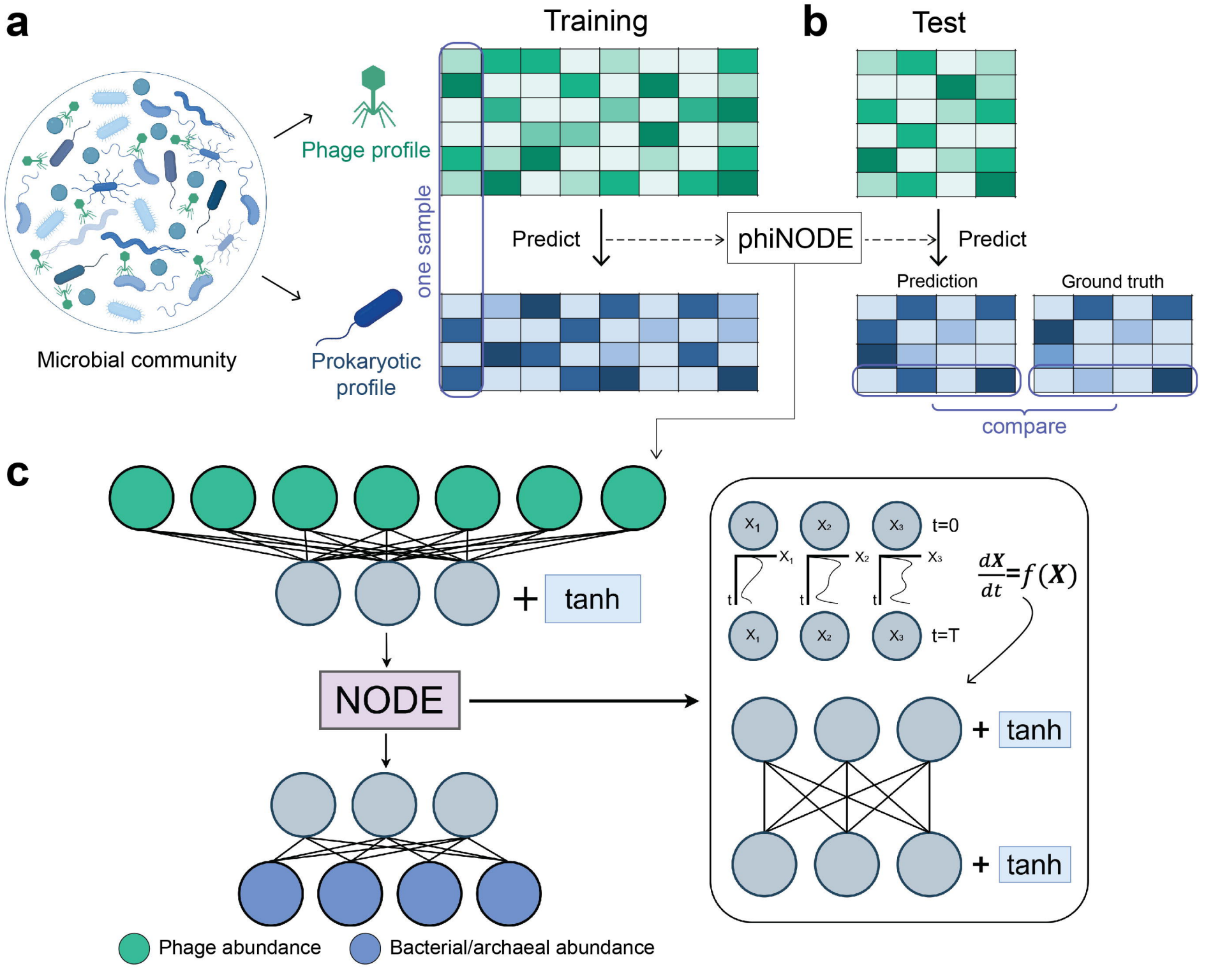
120. Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A. & Sørensen, S. J. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. *The CRISPR Journal* **3**, 462–469 (2020).

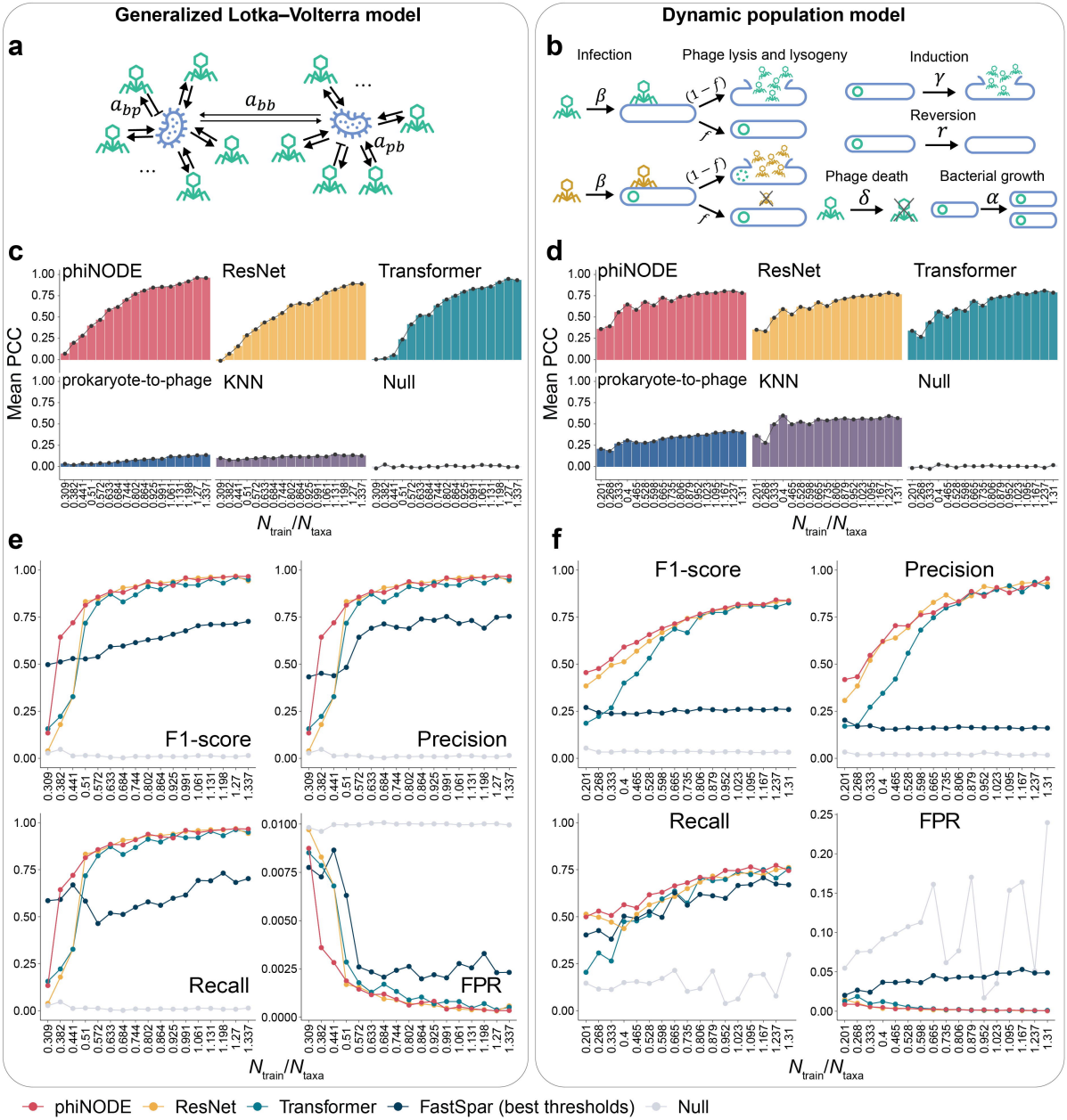
121. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* **39**, 578–585 (2021).

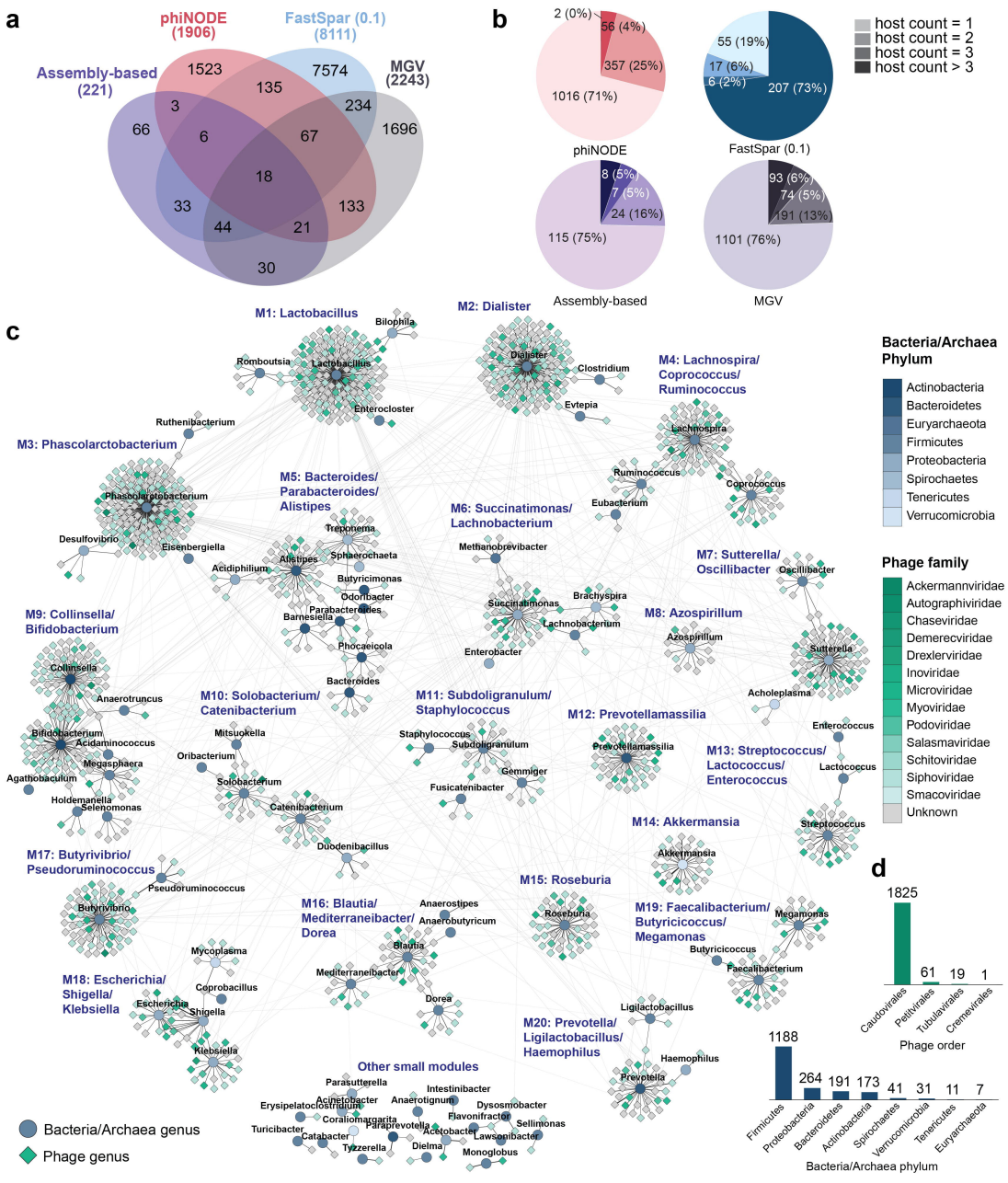
122. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

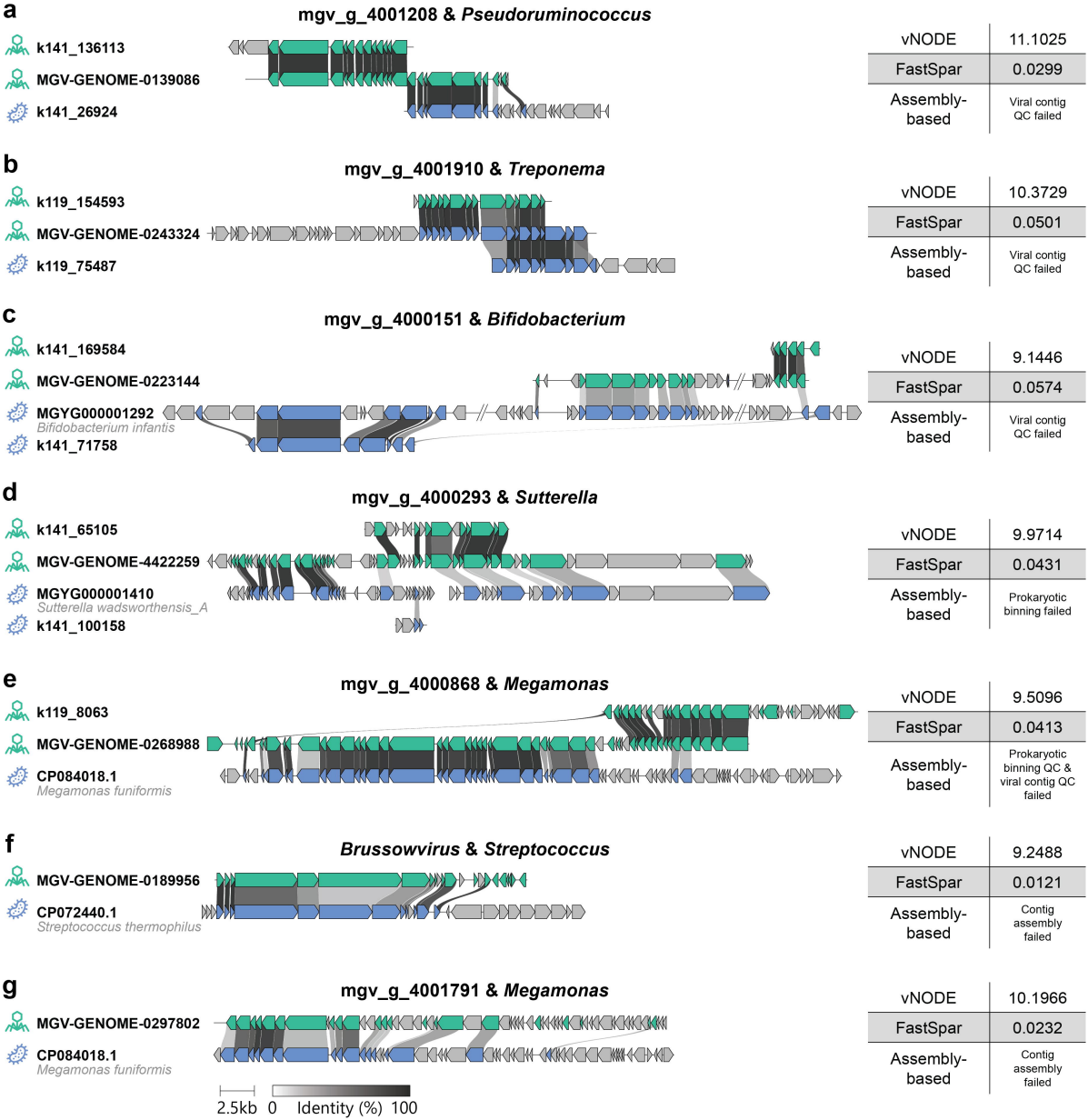
123. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

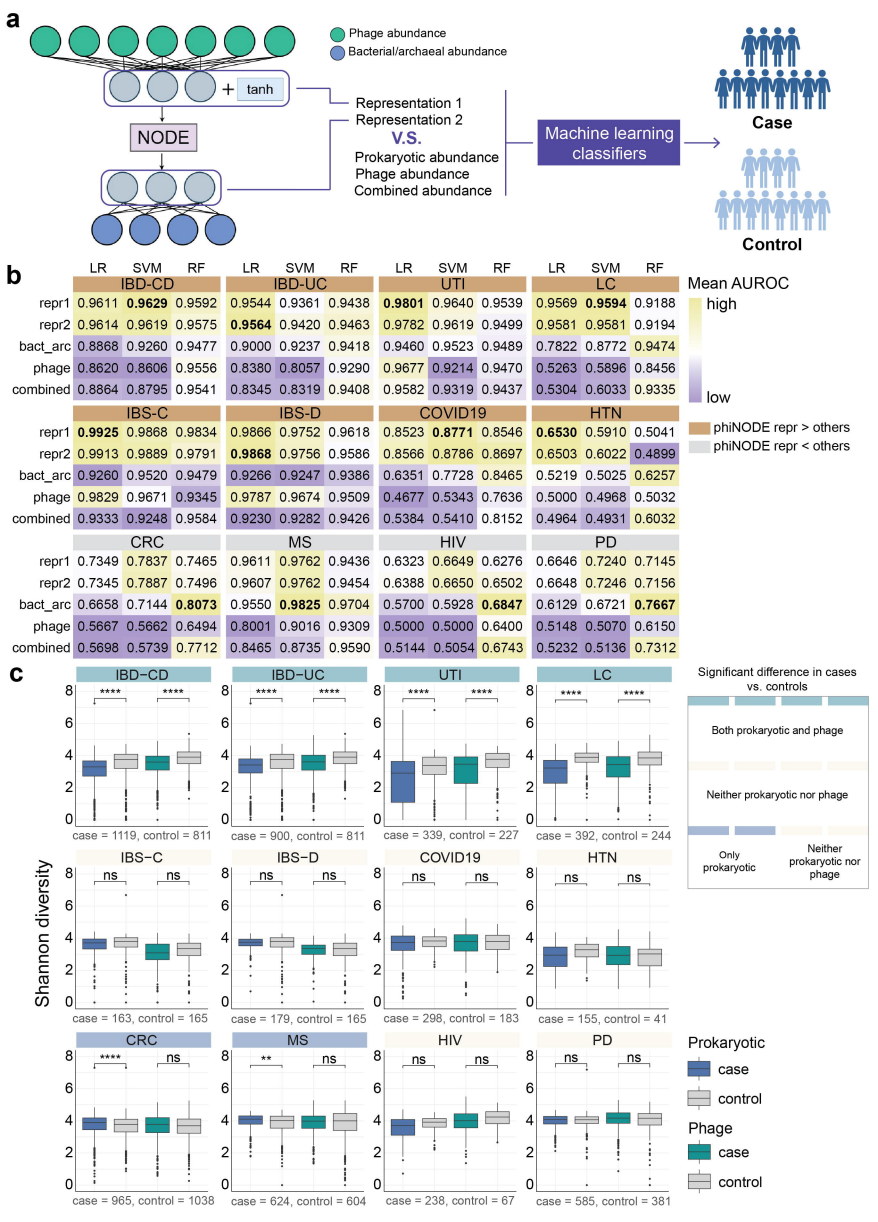

• Model construction
Phage profile
• ResNet
• Transformer
• NODE
Deep learning
Predict
Prokaryotic profile
• Application 1
Simulation data
Real-world data
Pseudo perturbation
Interactions
Reconstruct
Networks
• Application 2
Representation extraction
Cases & controls
Healthy individuals
Features
Classify
ROC
Phenotypes


a
Training
b
Test
Phage profile
Microbial community
Prokaryotic profile
one sample
Predict
phiNODE
Predict
Prediction
Ground truth
compare
c
tanh
NODE
t=0
t=T
$\frac{dX}{dt}=f(X)$
tanh
tanh
Phage abundance
Bacterial/archaeal abundance


Generalized Lotka–Volterra model
Dynamic population model
a
a_{bp}
a_{bb}
a_{pb}
b
Infection
Phage lysis and lysogeny
Induction
Reversion
Phage death
Bacterial growth
β
$(1-f)$
f
γ
r
δ
α
c
d
phiNODE
ResNet
Transformer
prokaryote-to-phage
KNN
Null
Mean PCC
N_{train}/N_{taxa}
e
f
F1-score
Precision
Recall
FPR
phiNODE
ResNet
Transformer
FastSpar (best thresholds)
Null


a
phiNODE (1906)
FastSpar (0.1) (8111)
Assembly-based (221)
MGV (2243)
1523
7574
135
3
234
66
6
67
1696
18
33
133
44
21
30
b
2 (0%)
56 (4%)
357 (25%)
1016 (71%)
phiNODE
55 (19%)
17 (6%)
6 (2%)
207 (73%)
FastSpar (0.1)
host count = 1
host count = 2
host count = 3
host count > 3
8 (5%)
7 (5%)
24 (16%)
115 (75%)
Assembly-based
93 (6%)
74 (5%)
191 (13%)
1101 (76%)
MGV
c
M1: Lactobacillus
Bilophila
Romboutsia
Lactobacillus
Enterocloster
Ruthenibacterium
M2: Dialister
Dialister
Clostridium
Evtepia
M3: Phascolarctobacterium
Phascolarctobacterium
Desulfovibrio
Eisenbergiella
M4: Lachnospira/ Coprococcus/ Ruminococcus
Lachnospira
Ruminococcus
Coprococcus
Eubacterium
M5: Bacteroides/ Parabacteroides/ Alistipes
Treponema
Alistipes
Sphaerochaeta
Acidiphilium
Butyricimonas
Odoribacter
Parabacteroides
Barnesiella
Phocaeicola
Bacteroides
M6: Succinatimonas/ Lachnobacterium
Methanobrevibacter
Succinatimonas
Brachyspira
Lachnobacterium
Enterobacter
M7: Sutterella/ Oscillibacter
Oscillibacter
Sutterella
Acholeplasma
M8: Azospirillum
Azospirillum
M9: Collinsella/ Bifidobacterium
Collinsella
Anaerotruncus
Bifidobacterium
Acidaminococcus
Megasphaera
Agathobaculum
Holdemanella
Selenomonas
M10: Solobacterium/ Catenibacterium
Mitsuokella
Oribacterium
Solobacterium
Catenibacterium
Duodenibacillus
M11: Subdoligranulum/ Staphylococcus
Staphylococcus
Subdoligranulum
Gemmiger
Fusicatenibacter
M12: Prevotellamassilia
Prevotellamassilia
M13: Streptococcus/ Lactococcus/ Enterococcus
Enterococcus
Lactococcus
Streptococcus
M14: Akkermansia
Akkermansia
M15: Roseburia
Roseburia
M16: Blautia/ Mediterraneibacter/ Dorea
Anaerostipes
Anaerobutyricum
Blautia
Mediterraneibacter
Dorea
M17: Butyrivibrio/ Pseudoruminococcus
Pseudoruminococcus
Butyrivibrio
M18: Escherichia/ Shigella/ Klebsiella
Mycoplasma
Coprobacillus
Escherichia
Shigella
Klebsiella
M19: Faecalibacterium/ Butyricicoccus/ Megamonas
Megamonas
Butyricicoccus
Faecalibacterium
M20: Prevotella/ Ligilactobacillus/ Haemophilus
Ligilactobacillus
Haemophilus
Prevotella
Other small modules
Parasutterella
Intestinibacter
Acinetobacter
Anaerotignum
Dysosmobacter
Erysipelatoclostridium
Flavonifractor
Coraliomargarita
Paraprevotella
Sellimonas
Turicibacter
Acetobacter
Catabacter
Lawsonibacter
Tyzzerella
Dielma
Monoglobus
Bacteria/Archaea genus
Phage genus
Bacteria/Archaea Phylum
Actinobacteria
Bacteroidetes
Euryarchaeota
Firmicutes
Proteobacteria
Spirochaetes
Tenericutes
Verrucomicrobia
Phage family
Ackermannviridae
Autographiviridae
Chaseviridae
Demerecviridae
Drexlerviridae
Inoviridae
Microviridae
Myoviridae
Podoviridae
Salasmaviridae
Schitoviridae
Siphoviridae
Smacoviridae
Unknown
d
1825
61
19
1
Caudovirales
Petitvirales
Tubulavirales
Cremevirales
Phage order
1188
264
191
173
41
31
11
7
Firmicutes
Proteobacteria
Bacteroidetes
Actinobacteria
Spirochaetes
Verrucomicrobia
Tenericutes
Euryarchaeota
Bacteria/Archaea phylum


a
mgv_g_4001208 & Pseudoruminococcus
k141_136113
MGV-GENOME-0139086
k141_26924
vNODE 11.1025
FastSpar 0.0299
Assembly-based Viral contig QC failed
b
mgv_g_4001910 & Treponema
k119_154593
MGV-GENOME-0243324
k119_75487
vNODE 10.3729
FastSpar 0.0501
Assembly-based Viral contig QC failed
c
mgv_g_4000151 & Bifidobacterium
k141_169584
MGV-GENOME-0223144
MGYG000001292
Bifidobacterium infantis
k141_71758
vNODE 9.1446
FastSpar 0.0574
Assembly-based Viral contig QC failed
d
mgv_g_4000293 & Sutterella
k141_65105
MGV-GENOME-4422259
MGYG000001410
Sutterella wadsworthensis_A
k141_100158
vNODE 9.9714
FastSpar 0.0431
Assembly-based Prokaryotic binning failed
e
mgv_g_4000868 & Megamonas
k119_8063
MGV-GENOME-0268988
CP084018.1
Megamonas funiformis
vNODE 9.5096
FastSpar 0.0413
Assembly-based Prokaryotic binning QC & viral contig QC failed
f
Brussowvirus & Streptococcus
MGV-GENOME-0189956
CP072440.1
Streptococcus thermophilus
vNODE 9.2488
FastSpar 0.0121
Assembly-based Contig assembly failed
g
mgv_g_4001791 & Megamonas
MGV-GENOME-0297802
CP084018.1
Megamonas funiformis
vNODE 10.1966
FastSpar 0.0232
Assembly-based Contig assembly failed
2.5kb
0
Identity (%)
100


a
Phage abundance
Bacterial/archaeal abundance
tanh
NODE
Representation 1
Representation 2
V.S.
Prokaryotic abundance
Phage abundance
Combined abundance
Machine learning classifiers
Case
Control
b
LR SVM RF
IBD-CD
repr1 0.9611 0.9629 0.9592
repr2 0.9614 0.9619 0.9575
bact_arc 0.8868 0.9260 0.9477
phage 0.8620 0.8606 0.9556
combined 0.8864 0.8795 0.9541
IBD-UC
repr1 0.9544 0.9361 0.9438
repr2 0.9564 0.9420 0.9463
bact_arc 0.9000 0.9237 0.9418
phage 0.8380 0.8057 0.9290
combined 0.8345 0.8319 0.9408
UTI
repr1 0.9801 0.9640 0.9539
repr2 0.9782 0.9619 0.9499
bact_arc 0.9460 0.9523 0.9489
phage 0.9677 0.9214 0.9470
combined 0.9582 0.9319 0.9437
LC
repr1 0.9569 0.9594 0.9188
repr2 0.9581 0.9581 0.9194
bact_arc 0.7822 0.8772 0.9474
phage 0.5263 0.5896 0.8456
combined 0.5304 0.6033 0.9335
IBS-C
repr1 0.9925 0.9868 0.9834
repr2 0.9913 0.9889 0.9791
bact_arc 0.9260 0.9520 0.9479
phage 0.9829 0.9671 0.9345
combined 0.9333 0.9248 0.9584
IBS-D
repr1 0.9866 0.9752 0.9618
repr2 0.9868 0.9756 0.9586
bact_arc 0.9266 0.9247 0.9386
phage 0.9787 0.9674 0.9509
combined 0.9230 0.9282 0.9426
COVID19
repr1 0.8523 0.8771 0.8546
repr2 0.8566 0.8786 0.8697
bact_arc 0.6351 0.7728 0.8465
phage 0.4677 0.5343 0.7636
combined 0.5384 0.5410 0.8152
HTN
repr1 0.6530 0.5910 0.5041
repr2 0.6503 0.6022 0.4899
bact_arc 0.5219 0.5025 0.6257
phage 0.5000 0.4968 0.5032
combined 0.4964 0.4931 0.6032
CRC
repr1 0.7349 0.7837 0.7465
repr2 0.7345 0.7887 0.7496
bact_arc 0.6658 0.7144 0.8073
phage 0.5667 0.5662 0.6494
combined 0.5698 0.5739 0.7712
MS
repr1 0.9611 0.9762 0.9436
repr2 0.9607 0.9762 0.9454
bact_arc 0.9550 0.9825 0.9704
phage 0.8001 0.9016 0.9309
combined 0.8465 0.8735 0.9590
HIV
repr1 0.6323 0.6649 0.6276
repr2 0.6388 0.6650 0.6502
bact_arc 0.5700 0.5928 0.6847
phage 0.5000 0.5000 0.6400
combined 0.5144 0.5054 0.6743
PD
repr1 0.6646 0.7240 0.7145
repr2 0.6648 0.7246 0.7156
bact_arc 0.6129 0.6721 0.7667
phage 0.5148 0.5070 0.6150
combined 0.5232 0.5136 0.7312
Mean AUROC
high
low
phiNODE repr > others
phiNODE repr < others
c
Shannon diversity
IBD-CD
case = 1119, control = 811
IBD-UC
case = 900, control = 811
UTI
case = 339, control = 227
LC
case = 392, control = 244
IBS-C
case = 163, control = 165
IBS-D
case = 179, control = 165
COVID19
case = 298, control = 183
HTN
case = 155, control = 41
CRC
case = 965, control = 1038
MS
case = 624, control = 604
HIV
case = 238, control = 67
PD
case = 585, control = 381
Significant difference in cases vs. controls
Both prokaryotic and phage
Neither prokaryotic nor phage
Only prokaryotic
Neither prokaryotic nor phage
Prokaryotic
case
control
Phage
case
control