Deep Learning Transforms Phage-Host Interaction Discovery from Metagenomic Data

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*

** Correspondence: yyl@channing.harvard.edu*

Abstract

Microbial communities are essential for sustaining ecosystem functions in diverse environments, including the human gut. 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 (PHIs) from metagenomic data are constrained by low sensitivity and the inability to accurately capture ecological relationships. To overcome these limitations, we developed PHILM (Phage-Host Interaction Learning from Metagenomic profiles), a deep learning framework that predicts PHIs directly from the taxonomic profiles of metagenomic data. We validated PHILM on both synthetic datasets generated by ecological models and real-world data, finding that it consistently outperformed the co-abundance-based approach for inferring PHIs. When applied to a large-scale metagenomic dataset comprising 7,016 stool samples from healthy individuals, PHILM identified 90% more genus-level PHIs than the traditional assembly-based approach. In a longitudinal dataset tracking PHI dynamics, PHILM's latent representations recapitulated microbial succession patterns originally described using taxonomic abundances. Furthermore, we demonstrated that PHILM's latent representations served as more discriminative features than taxonomic abundance-based features for disease classifications. In summary, PHILM represents a novel computational framework for predicting phage-host interactions from metagenomic data, offering valuable insights for both microbiome science and translational medicine.

Main

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 prokaryotic hosts, referred to as phage-host interactions (PHIs) 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 PHIs, viral contigs are assembled and then phage markers (e.g., clustered regularly interspaced short palindromic repeats (CRISPR) spacers and prophage signatures) are mapped to prokaryotic metagenome-assembled genomes (MAGs)[30,31]. This assembly-based approach of PHI inference has several drawbacks. First, low-abundance phages are under-represented, leading to poor genome coverage and unreliable assemblies[32–34]. Second, PHIs are inferred exclusively from characterized genomic features, restricting predictions to known mechanisms and overlooking novel or uncharacterized relationships. Finally, the high computational demands of the pipelines 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 approaches 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 PHIs in the human gut through the co-abundance analysis[39,40]. Yet, this co-abundance approach of PHI inference suffers from spurious correlations driven by coincidental co-occurrence rather than genuine biological relationships[41].

To overcome the limitations of the existing approaches, we investigated deep learning as an alternative approach of inferring PHIs 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 developed to directly infer PHIs from metagenomic profiles.

In this study, we proposed PHILM (Phage-Host Interaction Learning from Metagenomic profiles), a deep learning framework that (1) predicts prokaryotic abundance profiles from phage abundance profiles, (2) infers PHIs by performing a sensitivity analysis of the trained deep learning model, and (3) leverages the latent representations learned from the deep learning model as features for downstream tasks (e.g., disease classification and microbial succession pattern tracking) (**Fig. 1**). We tested three deep learning model architectures: NODE (Neural Ordinary Differential Equations), ResNet (Residual Neural Network), and Transformer, and benchmarked them on synthetic data generated from ecological models. We found that all three model architectures outperformed the co-abundance approach in inferring PHIs directly from metagenomic taxonomic profiles. Among them, NODE accurately predicted hosts for 90-97% of phages and recovered more true interactions than both the co-abundance approach and other deep learning model architectures. Therefore, we chose NODE as the representative deep learning model for PHILM in real data analysis. When applied to a dataset over 7,000 healthy-donor stool metagenomes, PHILM uncovered 90% more genus-level PHIs than the assembly-based approach. When applied to a longitudinal dataset tracking PHI dynamics, we further demonstrated that PHILM's latent representations reproduced the succession patterns originally observed in taxonomic abundances, demonstrating its ability to identify temporal shifts in PHIs. In both real-world datasets, PHILM consistently outperformed the co-abundance approach in recovering known interactions. Finally, PHILM's latent representations also proved more informative than taxonomic abundance features for distinguishing cases from controls across multiple diseases.

Results

Overview of PHILM

For a given set of shotgun metagenomic samples, PHILM takes paired phage and prokaryotic abundance profiles derived from taxonomic profiling as inputs for a deep learning model. In an illustrative dataset of twelve samples, eight were allocated to training and four to testing (**Fig. 1a**). 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. 1b**). The overall predictive power of deep learning models used within PHILM 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.

To evaluate PHILM's ability to infer PHIs, we conducted a sensitivity analysis (**Fig. 1c**, **Supplementary Fig. 1a**). We perturbed the relative abundance of phage i, denoted as x_i, by a small amount Δx_i, and used the trained deep learning model 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 Fig. 1b**). More technical details on the computation of S_{ji} are provided in **Methods**. Furthermore, the latent representations extracted from PHILM can be used to track the microbial succession pattern and serve as more discriminative features than taxonomic abundance-based features for disease classifications (**Fig. 1d**).

PHILM can predict prokaryotic profiles from phage profiles on synthetic data

To investigate if prokaryotic profiles can be predicted from phage profiles and if PHIs can be inferred from the trained model, we first validated PHILM using synthetic data with known ground truth. The data was generated by two ecological models: the Generalized Lotka-Volterra (GLV) model[47,48] and the dynamic population (DP) model[49]. 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 $N_{\text{taxa}} = 1{,}100$ (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[32,46,50–53].

The two models differ in complexity and ability to simulate PHIs. The GLV model is relatively simple and ignores the molecular details of phage infection and only accounts for phage-host and inter-host interactions (**Fig. 2a**). We also assumed an 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-the-art phage-host dynamic models[54]. In the DP model, each phage is allowed to infect multiple hosts (**Fig. 2b**).

We first evaluated the predictive performance of PHILM using three different deep learning architectures (NODE, ResNet, and Transformer) on synthetic datasets generated by the GLV and DP ecological models. The respective models phiNODE, phiResNet, and phiTransformer are named to reflect their underlying model architectures (**Supplementary Fig. 2** and see **Methods**). The mean PCCs for all deep learning models increased with the ratio of training samples to the total number of phage and prokaryotic taxa ($N_{\text{train}}/N_{\text{taxa}}$). When trained on all available samples with the largest $N_{\text{train}}/N_{\text{taxa}}$, 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, phiResNet: 0.890, and phiTransformer: 0.933) and lower values on the DP-generated data (phiNODE: 0.783, phiResNet: 0.764, and phiTransformer: 0.786) (**Fig. 2c-d, 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 Fig. 3**).

We also investigated whether prokaryotic profiles can be used to predict phage profiles by swapping the input and output in the deep learning model (based on NODE and denoted as "prokaryote-to-phage" in **Fig. 2c-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, we implemented a null model using the NODE model architecture but trained on datasets with randomly shuffled sample labels (denoted as "Null" in **Fig. 2**). Across all sample sizes from both GLV and DP models, the null model consistently yielded mean PCCs below 0.035, showing no improvement as $N_{\text{train}}/N_{\text{taxa}}$ increased. In addition, 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 $N_{\text{train}}/N_{\text{taxa}}$, 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. 2**). In contrast, the deep learning models consistently achieved higher mean PCCs, demonstrating their effectiveness in predicting 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 statistically significant but minor increase in mean PCC (P-value < 0.05; **Supplementary Fig. 4**). However, due to the marginal improvement observed in the synthetic data and the absence of high-confidence inter-prokaryotic interaction networks for empirical datasets, we did not apply this interaction-aware approach to real-world data, although our results demonstrate its potential utility.

PHILM enables phage-host interaction predictions on synthetic data

Next, we assessed whether PHIs could be inferred from a well-trained PHILM model (e.g., phiNODE, phiResNet, or phiTransformer) by performing a sensitivity analysis (see **Methods**). Notably, we observed that the F1-score, precisions, and 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 accurately reconstruct PHIs, achieving F1-scores no less than 0.94 on the GLV-generated data (phiNODE: 0.966, phiResNet: 0.943, and phiTransformer: 0.949) (**Fig. 2e**). Since the DP model allows phages to infect multiple hosts, we conducted multi-host predictions for all deep learning models by selecting the

thresholds for S_{ji} that maximize the F1-scores across different $N_{\text{train}}/N_{\text{taxa}}$ ratios. Compared to the performances on the GLV model, lower F1-scores were observed on the DP-generated data (phiNODE: 0.837, phiResNet: 0.837, and phiTransformer: 0.825), reflecting the complexity of the DP model (**Fig. 2f**). With the largest $N_{\text{train}}/N_{\text{taxa}}$, setting an optimal S_{ji} threshold of approximately for deep learning model multi-host prediction significantly improved their performance compared to single-host prediction. For example, phiNODE increases its precision from 0.924 to 0.955, recall from 0.670 to 0.745, and F1-score from 0.777 to 0.837 (**Supplementary Fig. 5, Supplementary Table 3**). Again, as expected, the null model achieved an F1-score below 0.06 (**Fig. 2e-f**), which was the lowest among all methods.

We also applied the "prokaryote-to-phage" method with different top phage numbers and optimal thresholds. Although the precisions were moderate, the F1-scores were notably lower (0.61 for the GLV model; 0.66 for the DP model in **Fig. 2e-f** and **Supplementary Fig. 6**). This result demonstrates that predicting phage profiles from prokaryotic profiles is less effective than the phage-to-prokaryote predictions.

In addition to the null model and "prokaryote-to-phage" model, deep learning models were compared with a co-abundance-based model, FastSpar[40]. While FastSpar with a threshold of ±0.5 demonstrated very high precision across different $N_{\text{train}}/N_{\text{taxa}}$ ratios, it generally exhibited extremely low recall on the GLV-generated data (**Supplementary Fig. 7a**), tending to reconstruct PHIs more accurately for the prevalent prokaryotic species (**Supplementary Fig. 8**). When the threshold was adjusted to ±0.1, FastSpar's recall improved with small sample sizes but subsequently decreased to below 0.25 as N_{train} exceeded 1,200 (**Supplementary Fig. 7a**). However, when evaluated on the DP-generated data, FastSpar with a threshold of ±0.5 exhibited a substantial reduction in precision, with an average of 0.51 (**Supplementary Fig. 7b**). In contrast, FastSpar with a threshold of ±0.1 yielded higher recall but at the expense of precision, averaging only 0.16, and suffered from high false positive rates (**Supplementary Fig. 7b**). 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 modest and slight improvements in F1-scores on GLV and DP model-generated data, respectively (**Fig. 2e-f**, **Supplementary Fig. 7a-b**). It is worth mentioning that the optimal FastSpar thresholds determined for the largest $N_{\text{train}}/N_{\text{taxa}}$ 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 cutoff of 0.1, potentially leading to high false positive rates, particularly in DP model-based synthetic data (**Supplementary Fig. 7a-b**, the last panels). These findings highlight the limitations

of co-abundance-based methods and underscore the superior predictive performance of deep learning models to infer PHIs.

Furthermore, although all the tree PHILM models performed comparably, phiNODE demonstrated superior performance relative to alternative methods in the 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. 2e-f, Supplementary Fig. 9**). 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 chosen as the representative PHILM model and used for all subsequent analyses.

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

To evaluate PHILM on real-world data, we downloaded a large-scale dataset comprising 7,016 metagenomic samples of healthy individuals obtained from curatedMetagenomicData[55] (**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 $N_{\text{train}}/N_{\text{taxa}}$ exceeds 1.0 (**Fig. 2e-f, Supplementary Table 3**). Using genus-level taxonomies, we achieved a training sample size ($N_{\text{train}} = 5{,}612$), which is 2.14 times the number of phage and prokaryotic taxa ($N_{\text{taxa}} = 2{,}620$) (**Supplementary Fig. 10a**), and hence is sufficient to ensure reliable interaction prediction. This reliability was further corroborated by a PCC of 0.793 obtained by PHILM on the test dataset (**Supplementary Fig. 10b**). We evaluated PHILM'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 (**Supplementary Fig. 11**). At this threshold, we recovered 1,906 genus-level PHIs (**Supplementary Table 4**).

To compare with existing methods, we derived PHIs using a conventional assembly-based approach[56]. 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. PHIs 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

253 alignments between prokaryotic MAGs and viral contigs[56]. This approach yielded over 2,500
254 MAG-to-contig interactions. To ensure a fair comparison with PHILM, both prokaryotic MAGs and
255 viral contigs were annotated at the genus level, resulting in the identification of 221 genus-level
256 PHIs (**Supplementary Fig. 12**).

257 In addition to the assembly-based approach, we also applied the co-abundance approach by
258 using FastSpar to infer PHIs with two correlation thresholds: ±0.10 (the default setting) and ±0.09
259 (the threshold at which interaction counts stabilize) (**Supplementary Fig. 11**). At ±0.10, FastSpar
260 identified 8,111 interactions spanning 285 phage genera; at ±0.09, it detected 9,227 interactions
261 across 334 phage genera, averaging 28 host genera per phage genus. When we compared the
262 overlap of interactions by the assembly-based approach, PHILM, FastSpar (±0.10), and FastSpar
263 (±0.09), we observed 48, 101, and 102 shared interactions, corresponding to positive predictive
264 values (PPVs) of 2.52%, 1.25%, and 1.11%, respectively. Overlapping the PHIs from the
265 Metagenomic Gut Virus (MGV) catalogue[30] with those from PHILM, FastSpar (±0.10), and
266 FastSpar (±0.09) yielded 239, 363, and 396 common interactions, with PPVs of 12.54%, 4.48%,
267 and 4.29%, respectively (**Fig. 3a**, **Supplementary Fig. 13a**). To compare PHILM and FastSpar
268 without relying on fixed thresholds, we evaluated both methods across a series of top-N cutoffs.
269 Although they achieved similar overlap with the assembly-based approach, perhaps due to the
270 limited number of interactions detected by this method, PHILM recovered significantly more
271 interactions in common with the MGV catalogue than FastSpar (**Supplementary Fig. 13b-c**).
272 Additionally, specialist-generalist analysis revealed that 71% of phages identified by PHILM are
273 specialists, whereas FastSpar predominantly classifies 73% of phages as generalists with more
274 than three host taxa (**Fig. 3b**, **Supplementary Fig. 13d**). The much higher proportion of generalist
275 phages predicted by FastSpar may reflect a high false-positive rate, consistent with our
276 benchmarking results using synthetic data (**Fig. 2e-f**). Notably, PHILM identified 89.4% more
277 interactions than the assembly-based approach, potentially suggesting its enhanced sensitivity to
278 dynamic, ecologically relevant phage-host relationships.

279 Based on PHILM's predictions, we reconstructed a PHI network and identified 20 densely
280 connected modules (**Fig. 3c**). Several of these modules exhibited clear biological relevance. For
281 instance, module 18 is enriched in pathogenic genera such as *Klebsiella*, *Escherichia*, and
282 *Shigella*. Module 13 comprises lactic acid-producing cocci, including *Streptococcus*, *Lactococcus*,
283 and *Enterococcus*. Module 9 contains the age-associated Actinobacteria genera *Bifidobacterium*
284 and *Collinsella* [57–59]. Module 7 includes *Sutterella* and *Oscillibacter*, both previously linked to

obesity[60–63]. Additionally, modules 4, 5, and 20 are dominated by *Ruminococcus*, *Bacteroides*, and *Prevotella*, respectively, which are genera that define the three gut enterotypes[64].

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. 3d**).

Among the 239 interactions shared by PHILM and the MGV catalogue, only 39 overlapped with those identified by the assembly-based approach (**Fig. 3a**). The remaining 200 exhibited significantly lower, yet predominantly non-zero, abundances in both phage and host genera across samples (**Supplementary Fig. 14**), suggesting that PHILM is more sensitive in detecting low-abundant PHIs. Among these, 154 interactions were missed by FastSpar.

To examine whether genomic features could verify these interactions, we randomly selected seven of the strong PHILM sensitivity values (≥ 9.0). PHILM 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. 4a**). A similar pattern was observed for mgv_g_4001910-*Treponema* and mgv_g_4000151-*Bifidobacterium*, where PHILM'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 interactions (**Fig. 4b-c**). Interactions mgv_g_4000293-*Sutterella* and mgv_g_4000868-*Megamonas* were not detected by the assembly-based approach due to poor prokaryotic binning but were supported by high PHILM sensitivity values and phage-host genome alignments (**Fig. 4d-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 PHILM narrowed and prioritized specific host genera in the healthy human gut. Finally, both PHILM 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. 4f-g**).

317 Together, these findings demonstrate PHILM's ability to detect genomically supported, low-
318 abundance, and environment-specific PHIs that both FastSpar and assembly-based approaches
319 overlook.

320 PHILM's latent representations informed by phage-host interactions can be utilized to 321 predict human host phenotypes

322 Due to insufficient healthy individual stool samples from curatedMetagenomicData for reliably
323 inferring species-level PHIs, which would require over 30,000 samples, we explored alternative
324 applications of PHILM using species-level microbial profiles. Numerous studies have
325 demonstrated the prediction of human host phenotypes using features derived solely from
326 prokaryotic profiles[65,66]. To our knowledge, no prior study has attempted to predict human host
327 phenotypes using phage-inclusive profiles or by incorporating PHIs. PHILM inherently learns PHIs
328 by predicting prokaryotic profiles from phage profiles through the internal layers within its deep
329 learning model (**Fig. 5a**). We hypothesize that these internal latent representations can serve as
330 informative features for predicting human host phenotypes.

331 Given that human host gender and age metadata were available for nearly all stool samples (98.8%
332 for gender and 100% for age in **Supplementary Fig. 10e-f**, **Supplementary Table 5**), we applied
333 the PHILM-derived latent representations along with other microbial abundance features to three
334 classifiers (Random Forest (RF); Logistic Regression (LR); Support Vector Machine (SVM)) to
335 predict different gender and age groups. Notably, significant differences in Shannon diversities for
336 both prokaryotic and phage compositions were observed between different gender and age
337 groups (**Supplementary Fig. 10g-h**), suggesting that phages may also contribute to
338 distinguishing between these groups. Two representations were extracted from the PHILM's
339 trained phiNODE model: one from a hidden layer closer to the input (representation 1, repr1) and
340 one from a hidden layer closer to the output (representation 2, repr2) (**Fig. 5a**).

341 Among all model-feature combinations, the SVM with PHILM repr2 and repr1 achieved the
342 highest average AUROC of 0.943 and the highest mean F1-score of 0.886, respectively, using
343 five-fold cross validation (**Supplementary Fig. 15a-b**). Regarding age prediction, where the data
344 were categorized into five qualitative groups based on established criteria[65], the SVM with PHILM
345 repr1 and repr2 achieved the best average AUROC of 0.950 and the best mean F1-score of 0.777,
346 respectively, across all strategies (**Supplementary Fig. 15c-d**). Collectively, these results
347 demonstrate that PHILM-derived latent representations exhibit higher predictive power for human
348 host phenotypes than other types of features.

PHILM recapitulates phage-host dynamics in the early-life gut microbiome

We further evaluated PHILM on a dataset of 12,262 longitudinal stool samples collected from 887 children during early life[46]. In the original study, genus-level PHIs were predicted by iPHoP[31] by pairing shotgun read-mapped viral and bacterial species-level genome bins from a reference database. Because applying an assembly-based approach to such a large dataset is computationally prohibitive, we compared PHILM only with the co-abundance-based approach (e.g., FastSpar). For both PHILM and FastSpar, we ranked their inferred interaction by strength and measured their overlap with the iPHoP predicted interactions across a series of top-N cutoffs. At every threshold tested, PHILM recovered significantly more iPHoP-validated interactions than FastSpar (**Supplementary Fig. 16**), demonstrating stronger concordance with genome-based inference.

Next, we examined the latent representations learned by PHILM, trained on species-level taxonomic profiles and presumed to encode underlying phage-host interplays, for their ability to capture temporal dynamics. Remarkably, when clustering these embeddings along the day of life, we observed eight distinct subsets whose trajectories (**Fig. 6b-c**) closely aligned with microbial succession patterns reported in the original study (**Fig. 6a**). These findings highlight that PHILM not only infers static PHI networks but also learns representations that capture the temporal dynamics of gut microbiome development.

PHILM's latent representations serve as promising features for disease classification

Building on the promising results of using PHILM-derived latent representations for phenotype prediction in healthy individuals, we next evaluated whether these representations could also enhance disease classification. While numerous attempts have been made to differentiate cases from controls using microbial features, most studies consider these features in isolation without incorporating cross-domain data or interactions, specifically phage-host relationships[67,68]. To demonstrate PHILM's potential in clinical diagnosis, we collected ten shotgun metagenomic datasets representing twelve diseases, including inflammatory bowel disease (IBD), urinary tract infection (UTI), liver cirrhosis (LC), irritable bowel syndrome (IBS), COVID-19 infection, hypertension (HTN), colorectal cancer (CRC), multiple sclerosis (MS), human immunodeficiency virus (HIV) infection and Parkinson's disease (PD). Cross-domain taxonomic profiles at the species level for both prokaryotes and phages were generated for all datasets (**Supplementary Fig. 17a**), which were used to train PHILM and extract two latent representations (**Fig. 5a**).

We first assessed differences between cases and controls across diseases using permutational multivariate analysis of variance (PERMANOVA). In PERMANOVA, the F statistic represents the

ratio of the variance between groups to the variance within groups, with higher values indicating greater inter-group variance than that of intra-groups. We found that the F values for PHILM representations (either repr1 or repr2) in 7 out of 10 datasets are higher than those obtained using prokaryotic features, phage features, or combined features (**Supplementary Fig. 18**). Similar to the previous analysis conducted for the healthy individual dataset, we next used these features as input variables for three classifiers (RF, LR, and SVM) to predict disease conditions. Compared with alternative features, either repr1 or repr2 achieved the highest average AUROCs from five-fold cross-validation in 8 out of 12 diseases (**Fig. 5b**, **Supplementary Fig. 19**), and they obtained the highest average F1-scores in 7 diseases (**Supplementary Fig. 20, Supplementary Table 6**). Among all models, LR most consistently achieved the best AUROCs. We also noticed that the diseases for which PHILM-derived representations outperformed other features in PERMANOVA did not always align with those showing superior performance in machine learning prediction, suggesting that PHILM representations may better capture non-linear and subtle patterns than traditional statistical tests.

We next investigated why PHILM-derived representations underperformed in four diseases: CRC, MS,HIV infection, and PD. We examined potential factors and found that it could not be attributed to case-control imbalance. Actually, except for HIV, COVID-19, and HTN, most datasets maintained balanced case-control ratios (**Supplementary Fig. 17b**). Additionally, differences in model fitting are unlikely to be the cause, as the mean PCCs for the four underperforming diseases were comparable to those for the others; notably, MS had a mean PCC of 0.73, the second highest among all diseases (**Supplementary Fig. 17c**). We then examined whether alpha diversity differs between cases and controls when using prokaryotic and phage profiles. For diseases showing significant differences in both profiles, such as IBD-CD, IBD-UC, UTI, and LC, PHILM-derived representations achieved the highest mean AUROCs among all features (**Fig. 5c**, the first row). For diseases without significant differences between cases and controls using any profiles, PHILM representations still outperformed other features in four diseases (IBS-C, IBS-D, COVID-19, and HTN) (**Fig. 5c**, the second row), whereas prokaryotic abundance-based features performed the best in the remaining diseases, HIV and PD (**Fig. 5c**, right two panels on the third row). Another scenario where PHILM underperformed was when significant differences were found only in prokaryotic profiles but not in phage profiles (**Fig. 5c**, left two panels on the third row). In other words, PHILM-derived representations consistently outperform other features in disease classification when both prokaryotic and phage profile Shannon diversities differ significantly between cases and controls.

415 Notably, PHILM's latent representations reduce dimensionality by roughly two orders of
416 magnitude compared to phage or combined abundance-based features. They were the smallest
417 in seven datasets and the second smallest in the remaining three (**Supplementary Fig. 17d**).
418 This substantial reduction in feature numbers, especially compared to combined abundance
419 features, highlights a more efficient approach to integrating cross-domain species data. Overall,
420 our findings demonstrate that latent representations learned from PHILM outperform traditional
421 species-based features in predicting most disease states, underscoring their potential as robust,
422 low-dimensional biomarkers for future research.

423 Discussion

424 Existing assembly-based approaches for inferring PHIs from shotgun metagenomic data tend to
425 miss many interactions because assembled genomes poorly represent low-abundance phage
426 taxa[32–34]. Moreover, these methods rely heavily on known molecular mechanisms, such as
427 infection processes or host defense systems, which limit their ability to discover novel PHIs. In
428 this study, we developed PHILM, a deep learning framework that predicts prokaryotic profiles from
429 paired phage profiles. By interpreting the trained model, we infer PHIs in a purely data-driven
430 fashion without relying on any predefined biological mechanisms, potentially capturing more
431 phage-host ecological dynamics. When tested on synthetic data, PHILM outperformed co-
432 abundance-based methods, while applied to a large, cross-cohort collection of >7,000 healthy
433 individual stool samples, PHILM identified more plausible PHIs and a higher proportion of
434 specialist phages than the co-abundance approach. More importantly, it recovered substantially
435 more interactions than the traditional assembly-based pipeline, which enabled a biologically
436 meaningful reconstruction and modulization of the healthy gut PHI network.

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

Despite its advantages, we acknowledge several limitations of the current framework. First, PHILM 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 PHILM 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, PHILM provides a scalable, mechanism-agnostic approach to PHI inference. With appropriate reference databases, this framework can be extended to microbiota from other human body sites or environments. PHILM offers an efficient and robust alternative to traditional assembly- and co-abundance-based methods for reconstructing PHIs 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[47,48]:

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

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 $\sigma = 0.05$. 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 the GLV model, with each pair of samples sharing 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.[49]. 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:

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

505 $$\frac{dP_j}{dt} = b\gamma \sum_i L_{ij} + \beta\big(b(1-f_j)P_j - P_j\big)\left(\sum_i S_i\, a_{ij} + \sum_i \sum_{j' \neq j} L_{ij'} a_{ij'}\right) - \delta P_j \quad (2)$$

506 $$\frac{dL_{ij}}{dt} = \alpha L_{ij} + \beta f_j S_i P_j a_{ij} - (\gamma + r)L_{ij} - \beta L_{ij} \sum_{j' \neq j} (1 - f_{j'}) P_{j'}\, a_{ij'} \quad (3)$$

507 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
508 infect prokaryote i, where we assume that a prokaryotic host can be infected by multiple phage
509 predators and a phage can also target multiple prokaryotic hosts. α indicates the prokaryotic
510 growth rate, while β indicates the phage infectivity. γ represents the prophage induction rate and
511 δ represents the phage death rate. b denotes the burst size, the number of new virus particles
512 released from a single infected host cell when it is lysed. r is defined as the reversion rate for a
513 lysogen to lose a prophage. f is the fraction of phages' infections leading to lysogeny, where
514 phages with f = 0 are obligate lytic and have no associated lysogens. f is randomly chosen for
515 each phage from a list of [0.0, 0.2, 0.4, 0.6, 0.8], representing different phage lysogeny abilities.
516 Details and literature support for the parameter settings are provided in **Supplementary Table 7**.
517 Equation (1) defines that the population density of i-th uninfected prokaryote is increased by self-
518 growth and reversion from infected prokaryotes L_{ij}, while decreased by the infection of phage j if
519 their interaction is set to non-zero in the matrix A. Equation (2) defines that the population density
520 of j-th phage is positively affected by the new phage particles produced by the burst from
521 prophage induction and from the direct lysis of sensitive prokaryotes and those infected by other
522 phages. The density is negatively affected by the phage infection and integration into available
523 prokaryotic genomes (including sensitive prokaryotes and prokaryotes infected by other phages)
524 and its death. Equation (3) defines the population density of the i-th prokaryote infected and
525 integrated by the j-th phage in its genome (also known as lysogens). The density increases due
526 to self-prokaryotic growth and phage j infecting sensitive prokaryote i, while it decreases due to
527 phage induction, its conversion into a sensitive prokaryote and lysis by other phages j' (where
528 $j' \neq j$). From a pool comprising $N = 100$ prokaryotes and $M = 1{,}000$ phages, we generated
529 between 200 and 1800 samples in increments of 100 using the DP model, with each sample
530 containing exactly 10 prokaryotes, and each pair of samples shared 60% of prokaryotes. In each
531 round of sampling, a random number of phages was selected for a specific prokaryote. After
532 simulating for a longer period (T=5,000 steps) than the GLV model, instead of reaching a steady
533 state, the phage and prokaryotic abundances may still fluctuate in a chaotic way but maintain a

534 dynamic balance. In fact, this chaotic dynamic population model was first proposed to explain the
535 coexistence of lytic and temperate phages in nature[49].

536 Metagenomic sample collection and processing

537 Healthy individual dataset

538 These datasets include 7,016 metagenomic stool samples of sequencing depth greater than 1
539 million reads from healthy individuals ranging in age from 8-75 years old without antibiotics usage.
540 The sample accessions were collected from curatedMetagenomicData[55] and downloaded from
541 the Sequence Read Archive (SRA) database. According to previously reported criterion, age data
542 were grouped into five categories including children adolescents (8 ≤ age ≤ 18), young adult
543 (18 < age ≤ 35), middle aged (35 < age ≤ 50), senior (50 < age ≤ 65), and elderly (age > 65).
544 Metadata can be found in **Supplementary Table 5**.

545 Disease datasets

546 Human fecal raw metagenomic sequencing data and clinical metadata of the samples for (1) IBD
547 and controls were collected and downloaded from the Inflammatory Bowel Disease Multi'omics
548 Database (IBDMDB) (https://ibdmdb.org/results), including HMP2 and HMP2 Pilot studies[69,70],
549 PRJEB15371[71], PRJNA1086048[72], PRJNA400072[73] and PRJNA429990[74]. (2) CRC and controls
550 were from PRJDB4176[75], PRJEB10878[76], PRJEB12449[77], PRJEB27928[78], PRJEB6070[79],
551 PRJEB7774[80], PRJNA389927[81], PRJNA429097[82], PRJNA447983[83], PRJNA531273 (with
552 PRJNA397112 as controls)[84], PRJNA731589[85], PRJNA763023[86] and PRJNA936589
553 (downloaded from https://www.ncbi.nlm.nih.gov/bioproject/PRJNA936589/ since no paper was
554 published). The metadata of CRC were partially organized from Zun et al.[87] (3) Multiple sclerosis
555 (MS) and controls were from PRJEB51635 (patients) and PRJEB41786/PRJEB41787 (controls)[88].
556 (4) Parkinson's disease (PD) and controls were from PRJEB53403 and PRJEB53401[89] and
557 PRJNA834801[90]. (5) Liver cirrhosis (LC) and controls were obtained from PRJEB6337[91],
558 PRJNA373901[92] and from a paired set where patients are from PRJNA912122 and controls from
559 PRJNA838648[93]. (6) Urinary tract infection (UTI) and controls were from PRJNA400628[94]. (7)
560 Irritable bowel syndrome (IBS) and controls were collected from PRJEB37924[95]. (8) COVID19
561 and controls were obtained from PRJDB13214[96] and PRJNA650244[97]. (9) HIV and controls were
562 obtained from PRJNA307231[98], PRJNA692830[99], PRJNA391226[100], and PRJNA450025[101]. (10)
563 Hypertension (HTN) and controls were from PRJEB13870[102]. More details in the metadata can
564 be available in **Supplementary Table 8**.

565 Quality control and host contamination removal were done with KneadData v.0.12.0
566 (https://github.com/biobakery/kneaddata) using Trimmomatic v.0.39[103] with options
567 "ILLUMINACLIP:<adapter file>:2:30:10:2:TRUE LEADING:3 TRAILING:3
568 SLIDINGWINDOW:4:15 MINLEN:30" and Bowtie2 v.2.4.5[104], respectively. Phanta v.1.1.0 was
569 subsequently using the following parameters: "confidence_threshold: 0.1; class_mem_mb: 60000;
570 class_threads: 28" and the prophage-masked version of database[32]. Specifically, to further filter
571 false-positive species, different viral coverage thresholds and prokaryotic coverage thresholds
572 have been selected based on the sample sequencing depths following the suggestions at
573 https://github.com/bhattlab/phanta/tree/main. Phage and prokaryotic profiles were derived from
574 the Phanta-generated relative taxonomic profiles. To only include viruses infecting prokaryotic
575 hosts (namely phages), only viral taxa belonging to the orders *Caudovirales*, *Tubulavirales*,
576 *Petitvirales*, *Cremevirales*, and *Halopanivirales* were considered.

577 ## Deep learning model architectures and hyperparameter tuning

578 We evaluated and compared three deep learning models: Transformer[105], Residual Neural
579 Network (ResNet)[106] and Neural Ordinary Differential Equations (NODE)[107]. We included the
580 transformer because of its robust attention mechanism and its ability to capture long-range
581 dependencies within the data[105]. ResNet was selected due to its use of skip connections, which
582 facilitate the training of deeper architectures by ensuring direct gradient flow and mitigating the
583 vanishing gradient problem[106]. NODE extends its precursor, ResNet, by integrating skip
584 connections that are analogous to Euler steps in ODE solvers[107]. Each residual block in NODE
585 represents a small time change of variables, inspiring its continuous-depth architecture. Unlike
586 ResNet's discrete hidden layers, NODE parameterizes the derivative of the hidden state using a
587 neural network and computes outputs through an ODE solver. This methodology allows NODE to
588 maintain constant memory, adapt its evaluation strategy for each input, and balance numerical
589 precision with speed.

590 ## phiNODE

591 phiNODE is a method that integrates an implicit NODE module within a deep learning architecture.
592 Prior to modeling, the centered log-ratio transformation is applied to both the phage-excluded and
593 phage taxonomic abundance profiles. Considering the large disparity in the numbers of
594 prokaryotic and phage taxa, it is not possible to directly apply the original NODE architecture[107]
595 as it requires input and output dimensions to be equal. Inspired by the architecture of mNODE[44],
596 we incorporated two fully connected hidden layers with the NODE in between to transform the

data dimensions. 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.

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[107], 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 tuned 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 the model to explicitly 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.

phiResNet

phiResNet adopts the residual neural network architecture that learns 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 tuning (the dimension of the hidden layer N_h, the L2 regularization parameter λ and the learning rate l and early stopping was implemented using the same strategy as that employed for phiNODE.

phiTransformer

phiTransformer, based on the transformer architecture, 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 (keys and values) 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

659 hyperparameters were tuned following a strategy analogous to that used for phiNODE. Model
660 construction, training and validation are implemented using PyTorch v2.4.1[110].

661 Model training, validation and testing

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

667 Inferring phage-host interactions via sensitivity analysis

668 The well-trained deep learning model within PHILM takes the relative abundance of phage *i* (x_i)
669 as inputs and generates predictions for the relative abundance of prokaryote *j* (y_j). For the sample
670 *m* in the training set, $x_i^{(m)}$ for all *i* is provided as the input vector to the trained model and the
671 model can predict the relative abundances $y_j^{(m)}$ for all prokaryotes. To investigate the influence
672 of phage *i* on prokaryote *j* for a particular sample *m*, we perturb $x_i^{(m)}$ by setting it as the mean of
673 relative abundances for phage *i* across samples (denoted as $\bar{x}_i$) while keeping values of $x_{j\neq i}^{(m)}$
674 intact. Thus, the perturbation amount for phage *i* is $\Delta x_i^{(m)} = \bar{x}_i - x_i^{(m)}$. This perturbed vector for
675 phage relative abundances is provided to the trained deep learning model to regenerate
676 predictions for prokaryotic relative abundance $y_j^{(m)} + \Delta y_j^{(m)}$ where $\Delta y_j^{(m)}$ is the deviation of newly
677 predicted relative abundance for prokaryote *j* when the deep learning model uses the perturbed
678 input vector from that when it uses the unperturbed input vector. For the sample *m*, the sensitivity
679 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
680 samples and z-score normalized for each phage. The overall sensitivity of prokaryote *j* to phage
681 *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
682 σ_i are the mean and standard deviation, respectively, of S_{ji} calculated for phage *j*.

683 Network analysis and visualization

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

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

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

697 ## Predict phage-host interactions using a co-abundance-based method

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

702 ## Infer phage-host interactions using the assembly-based approach

703 For the 7,016 metagenomic stool samples from healthy individuals, we performed single-sample
704 assembly using MEGAHIT v.1.2.9[115], after excluding contigs less than 2,000 bp, resulting in a
705 total of 64,888,908 assembled contigs. Viral and non-viral sequences were differentiated using
706 geNomad v.1.7.6[27], which excluded provirus and plasmid sequences, yielding 62,053,011 non-
707 viral contigs and 2,835,897 viral contigs. Non-viral contigs were binned using MetaBAT v.2.15[116],
708 producing 206,411 bins, which were further assessed for quality using CheckM v.1.2.3[117],
709 retaining those with ≥50% completeness and <10% contamination, leading to 90,447
710 metagenome-assembled genomes (MAGs). Taxonomic classification of prokaryotic bins was
711 performed using CAT pack v.6.0.1[118] with HumGut as the reference database[119], identifying 45
712 prokaryotic genera. CRISPR spacers were detected in the MAGs using CasCRISPRTyper
713 v.1.8.0[120], yielding 56,461 CRISPR spacers. Viral contigs were quality-filtered using CheckV
714 v1.0.3[121], retaining 96,599 high- to medium-quality viral contigs, which were then assigned taxa
715 using uhgv-tools (https://github.com/snayfach/UHGV/blob/main/CLASSIFY.md) with MGV as the
716 reference database[30], identifying 1,947 viral genera. To predict host-phage interactions, we
717 followed the strategy described by Nayfach et al.[56] and employed CRISPR spacer matching via

718 blastn-short (allowing ≤1 mismatch) and sequence similarity-based matching via BLASTN
719 v.2.15.0+[122], considering hits with >90% identity and >500 bp overlap. These approaches
720 identified 2,589 host MAG-viral contig interactions, resolving into 221 host-virus interactions at
721 the genus level.

722 Machine learning classifiers

723 The machine learning models were trained and validated using a nested cross-validation
724 framework to ensure robust model selection and performance evaluation. Input features were
725 standardized using StandardScaler in the scikit-learn module[123]. Three classification models were
726 evaluated: Support Vector Machine (SVM), Logistic Regression (LR) and Random Forest (RF).
727 Model selection and performance assessment were conducted using a 5-fold cross-validation,
728 with folds generated via StratifiedKFold from the scikit-learn library[123]. One of the five folds was
729 reserved as a held-out test set. This approach was employed to address class imbalance in the
730 data, ensuring that each fold preserved the original dataset's class distribution and thereby
731 providing a more robust evaluation of model performance. The remaining data were subjected to
732 hyperparameter optimization using GridSearchCV from the scikit-learn library with an inner 5-fold
733 cross-validation[123]. For SVM model, the hyperparameter grid included regularization values (C)
734 of 0.1, 1, and 10 and kernel types of linear and radial basis function (rbf). For LR model, the
735 hyperparameter grid explored different penalty types ("none", L2, and ElasticNet) and
736 regularization strength (C) values of 0.1, 1, and 10. The hyperparameter grid for the Random
737 Forest included the number of trees (n_estimators: 100, 300, 500) and the fraction of samples
738 used for building each tree (max_samples: 0.2, 0.5, 0.8). Model selection was based on the
739 highest mean area under the curve (AUC) score obtained from the inner cross-validation, with the
740 best hyperparameters applied to the outer test fold for evaluation. Model performance was
741 assessed using AUC, precision, recall, and F1-score, and predictions for each fold. For age data,
742 which was treated as a multi-class classification problem, we computed one-vs-all AUC as well
743 as macro-averaged precision, recall, and F1-score.

744 Acknowledgements

745 Y.-Y.L. acknowledges funding support from the Resnek Family Center for PSC Research at
746 Brigham and Women's Hospital, the National Institutes of Health (R01AI141529, R01HD093761,
747 RF1AG067744, UH3OD023268, U19AI095219 and U01HL089856) as well as the Office of the
748 Assistant Secretary of Defense for Health Affairs, through the Traumatic Brain Injury and
749 Psychological Health Research Program (Focused Program Award) under award no. W81XWH-

22-S-TBIPH2, endorsed by the Department of Defense. X.-W.W. acknowledges funding support from the National Institutes of Health (K25HL166208).

Data and code availability

All code for running PHILM 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.

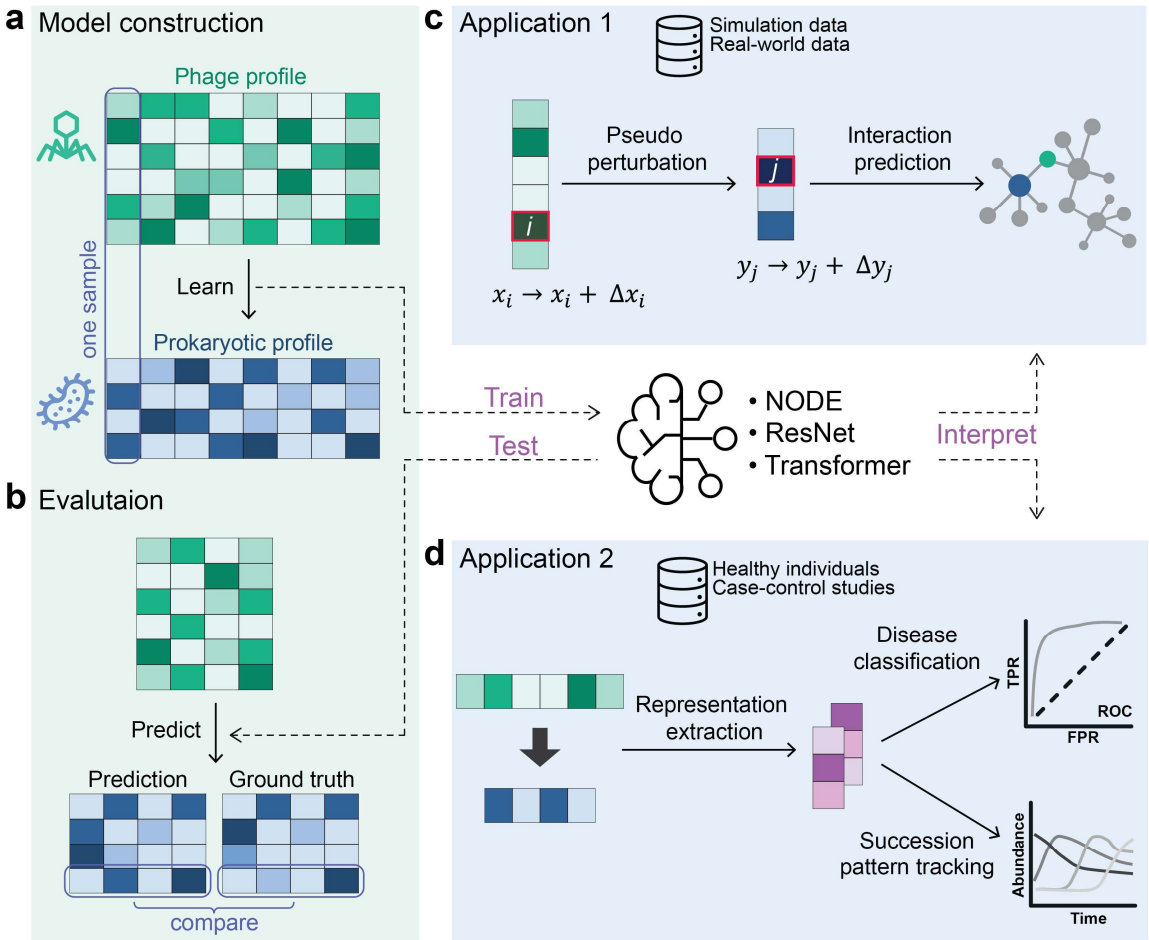


Fig 1: The workflow of the PHILM framework. PHILM is designed to apply deep learning models to discover phage-host interactions from metagenomic data. **a**, Deep learning models are utilized to train on paired phage and prokaryotic relative abundance profiles from the same set of samples. Across all panels, green- and blue-colored matrices represent the phage profiles and prokaryotic profiles, respectively. None of the entities 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. Eight samples in the training set are used for training deep learning models. **b**, A well-trained deep learning model can predict prokaryotic microbial profiles in the test set. For each prokaryotic taxon, the Pearson correlation coefficient (PCC) between its predicted and true abundances across samples in the test set is computed to quantify the models' performance. **c**, The first application of PHILM is to conduct a sensitivity analysis that perturbs the abundances of input features and predicts alterations in the output. In this way, it helps reveal possible interactions between the input (i.e., phage taxon) and output (i.e., prokaryotic taxon). **d**, The second application involves extracting latent representations from trained models that implicitly encode

772 informative PHIs, which serve as novel features to predict human host phenotypes such as
773 gender in healthy individuals and disease conditions between patients and healthy controls, and
774 to track succession patterns. ResNet: Residual neural network; NODE: Neural ordinary differential
775 equations.

776

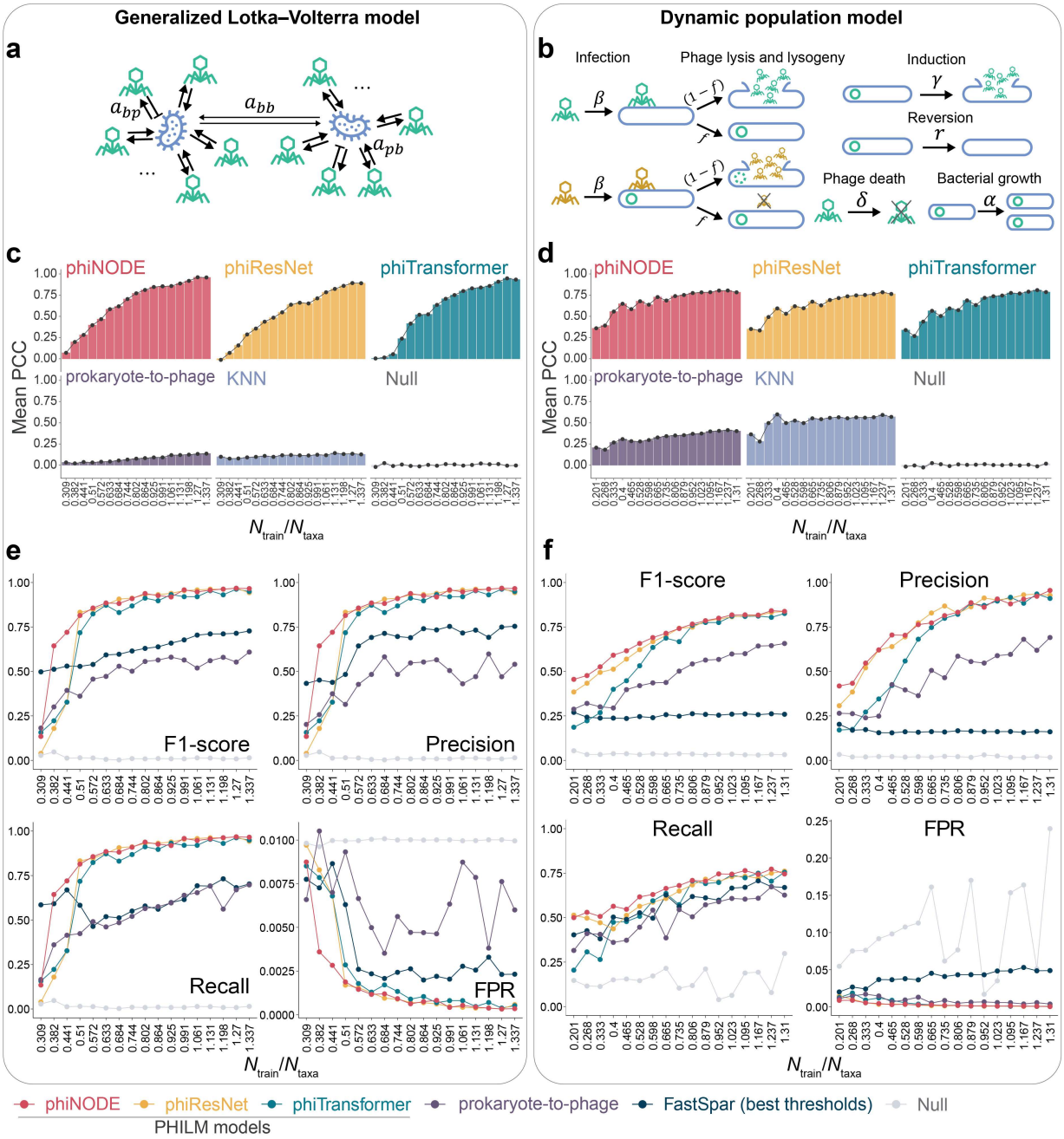


Fig 2: Validation of PHILM using synthetic data. a, Schematic of data generation using the generalized Lotka-Volterra model (GLV). a_{pb} and a_{bp} represent the phage-prokaryote interactions, while a_{bb} presents the inter-prokaryotic interactions. Each prokaryote can be infected by 10 phages, and 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 reached steady states. **b**, Schematic of data generation using the dynamic population model (DP). Although each prokaryote is set to be targeted by 10 phages, phages are allowed to infect multiple

785 hosts. The DP model was simulated until all phage and prokaryotic abundances reached dynamic
786 balances. Each phage has a lysogen fraction (f), indicating its infections leading to lysogeny.
787 Phage with f=0 are obligate lytic. More details can be found in the Methods section. **c-d**,
788 Performance of phiNODE, phiResNet, phiTransformer, prokaryote-to-phage method, and a k-
789 nearest neighbors (KNN) model to predict prokaryotic profiles (phage profiles for prokaryote-to-
790 phage method) on the test sets in simulation data from (**c**) GLV and (**d**) DP models with different
791 ratios of training sample size/phage and prokaryote taxa number, $N_{\mathrm{train}}/N_{\mathrm{taxa}}$. The mean PCCs
792 for all prokaryotic abundances (phage abundances for the prokaryote-to-phage method) across
793 samples are used to reflect the models' prediction power. **e-f**, Performance of deep learning-
794 based methods and a co-abundance-based method (i.e., FastSpar) to predict PHIs using
795 synthetic data generated by (**e**) GLV and (**f**) DP models with varying $N_{\mathrm{train}}/N_{\mathrm{taxa}}$ ratios. F1-score,
796 precision, recall, and false positive rate (FPR) were used to evaluate the model's performance.
797 Note that deep learning methods adopted single-host prediction on GLV model-generated data
798 and multi-host prediction on DP model-generated data. For single-host prediction on DP model
799 data, please refer to **Supplementary Fig. 5**.

800

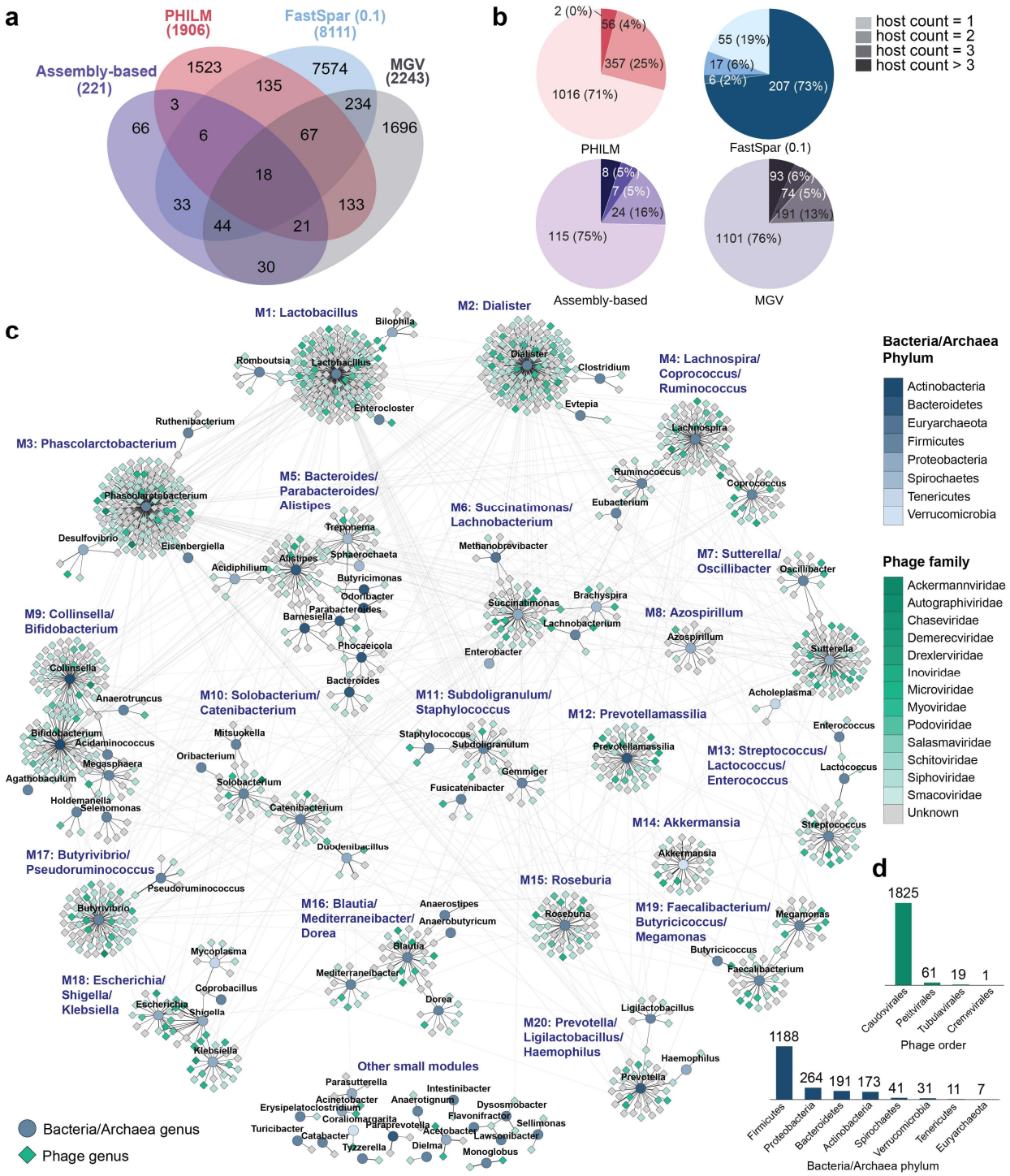


Fig 3: Reconstruction of phage-host interactions in healthy individuals using PHILM. a, Venn diagram illustrating the overlaps among PHIs predicted by FastSpar (using a ±0.1 threshold), the assembly-based approach, PHILM, and the interactions recorded in the MGV (Metagenomic Gut Virus) catalogue. **b**, Pie charts illustrating the distribution of phage specificities toward the

806 hosts using different approaches. **c**, PHI networks inferred by PHILM, where circles represent
807 bacteria/archaea genera and diamonds represent phage genera. Phage families and
808 bacteria/archaea phyla are displayed in different colors, and the thickness of the edges indicates
809 the interaction strengths. Nodes that belong to the same modules are clustered together. Edges
810 within modules are shown in black, while those between modules are displayed in light gray. **d**,
811 Distributions of phage orders (top) and bacteria/archaea phyla (bottom) among the interactions
812 identified by PHILM.

813

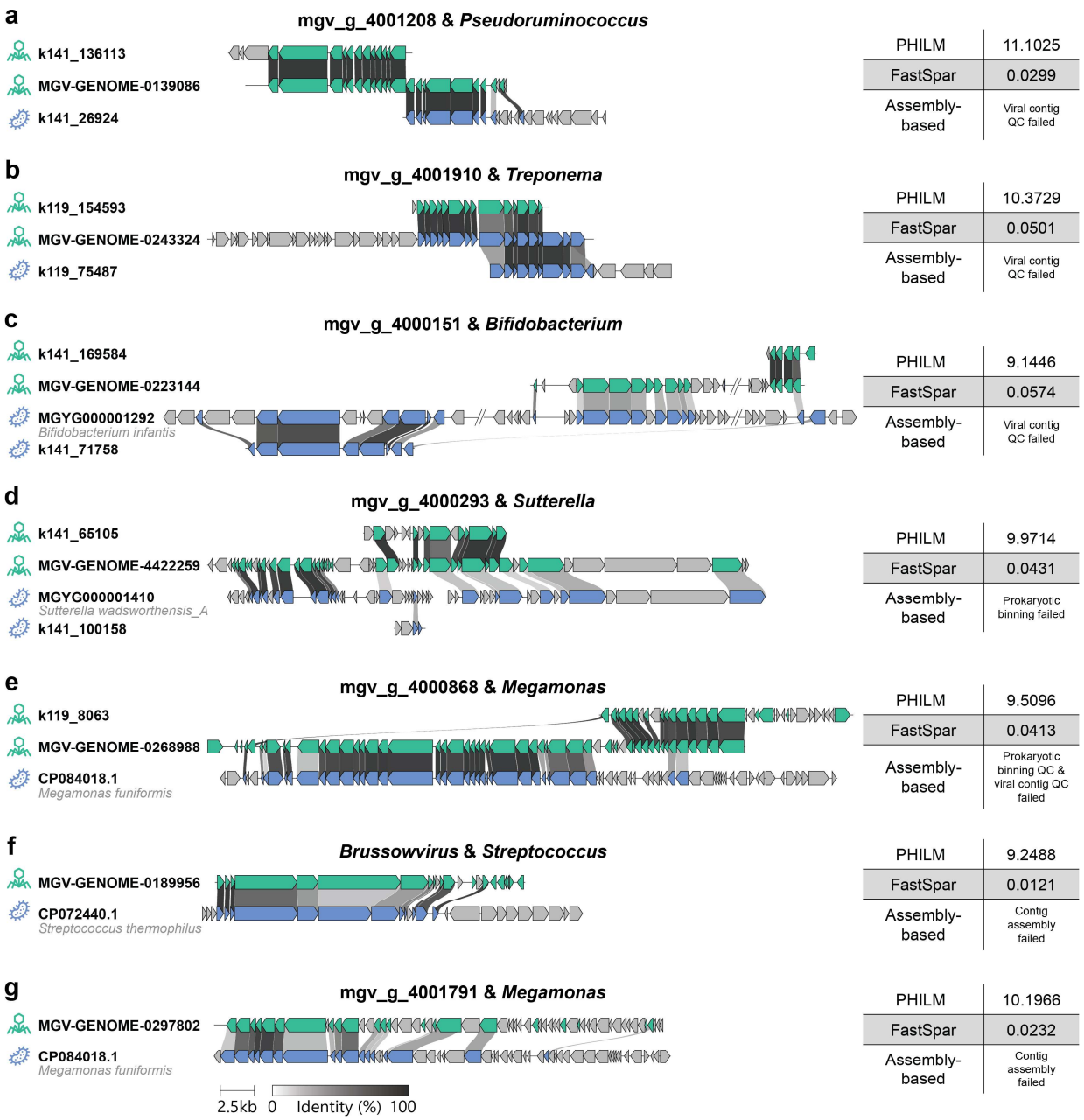


Fig 4: Examples illustrating PHILM's higher sensitivity in recovering interactions supported by genomic features. a-c, Examples show interactions that were not supported by FastSpar or the assembly-based approach, yet are supported by the Metagenomic Gut Virus (MGV) catalogue. **d-e**, Examples illustrate that the interactions were not identified by the assembly-based approach due to poor prokaryotic binning. The MGV catalogue reported multiple prokaryotic host genera for the phage genera, while PHILM prioritized host genera for these phages in the gut of healthy individuals. **f-g**, Examples demonstrate an interaction detected by PHILM in samples that lacked any assembled phage contigs. Although neither FastSpar nor the

823 assembly-based approach identified this interaction, it is supported by the MGV catalogue. The
824 superkingdom icons and genome or contig accessions are shown to the left of each sequence.

825

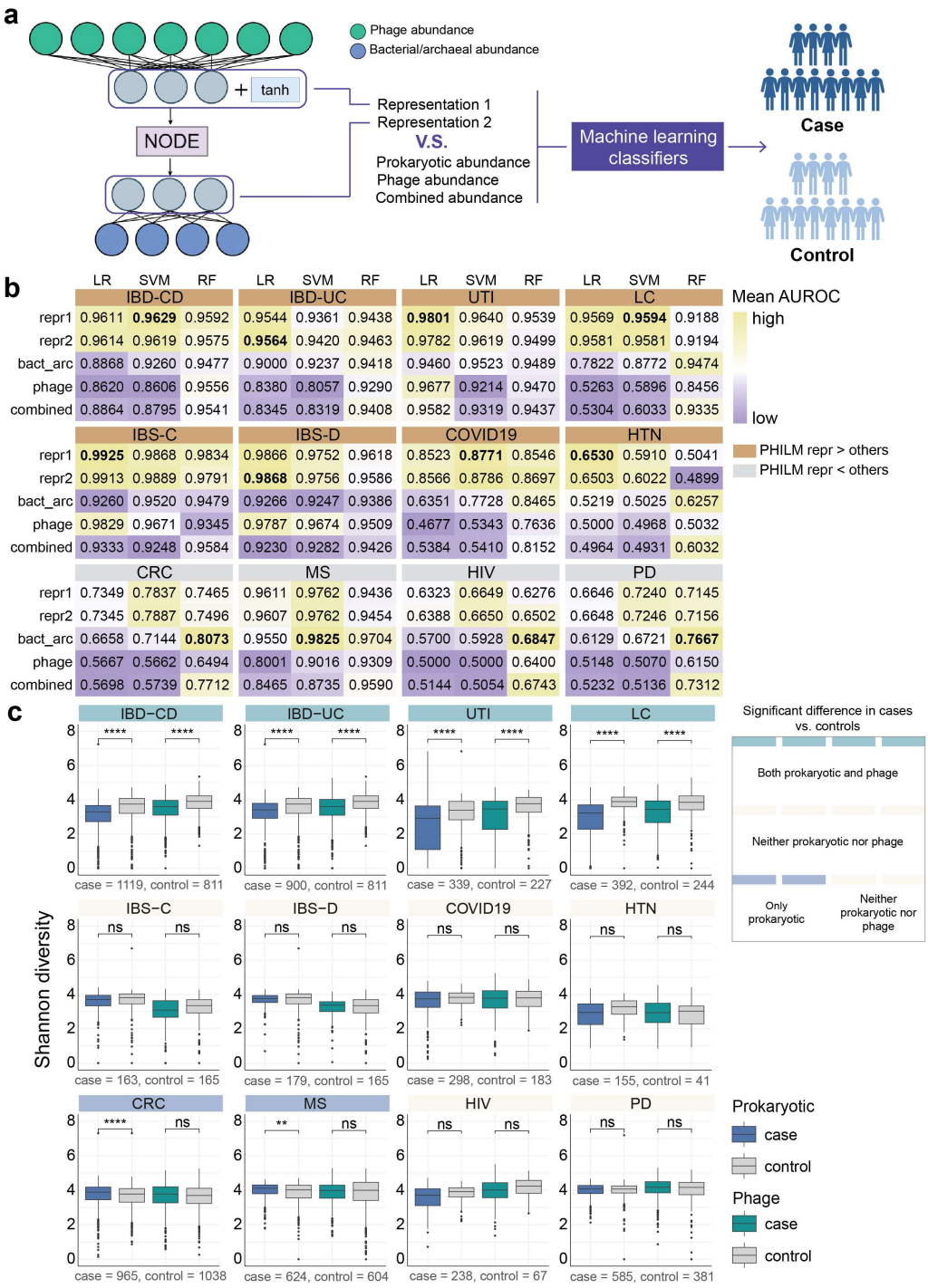
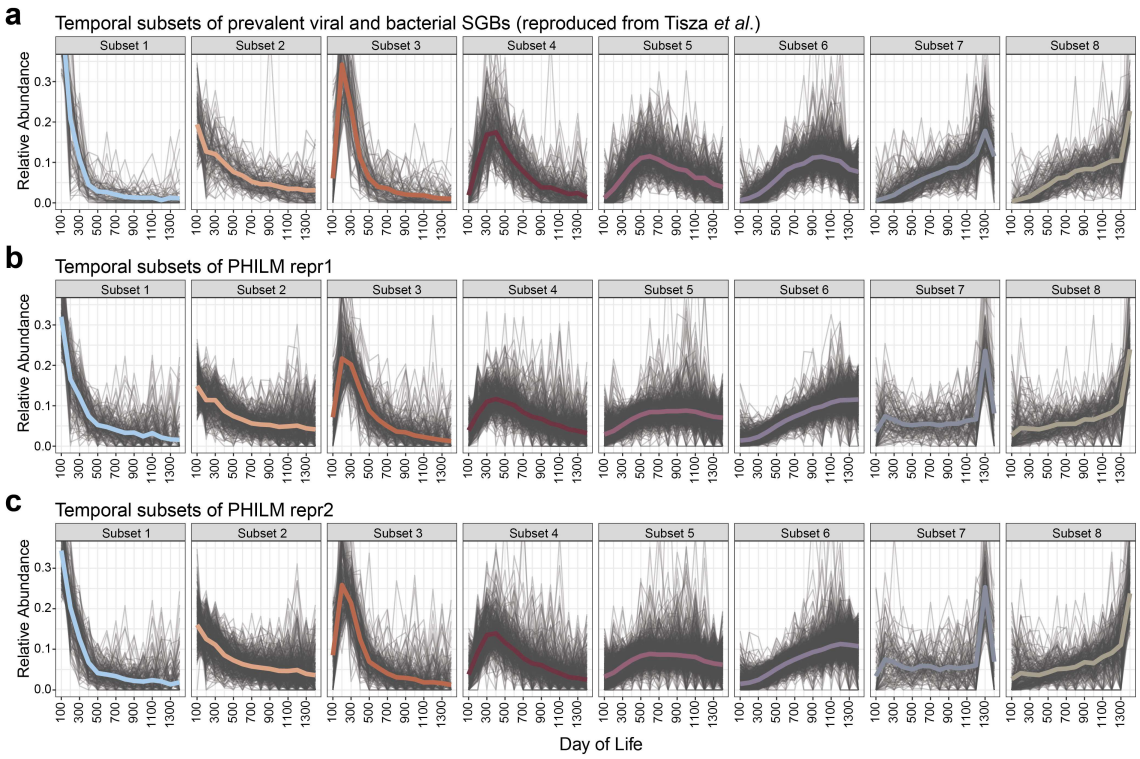


Fig 5: Application of PHILM's latent representations for case-control classification across diseases. a, For a trained PHILM deep learning model (phiNODE shown as an exemplar model) to predict disease-related prokaryotic profiles from phage profiles, its hidden layers implicitly reflecting PHIs can be 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 PHILM-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 PHILM-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 ($p > 0.01$), ** ($p \leq 0.01$), *** ($p \leq 0.001$), and **** ($p \leq 0.0001$). 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.



852

853 **Fig 6: Application of PHILM's latent representations to reconstruct phage-host dynamics**
854 **in the early-life gut microbiome. a**, Temporal subsets of prevalent viral and bacterial species-
855 level genome bins (SGBs), reproduced from Tisza *et al*. Each trajectory shows the abundance
856 changes of a single SGB over days of life. Gray lines represent individual SGBs, and colored lines
857 indicate their average trends. Viral and bacterial SGBs are plotted together, rather than separately
858 in the original study. **b-c**, Temporal subsets of PHILM's latent representations 1 and 2, respectively.
859 Values in representation vectors were inverse-centered log-ratio-transformed into abundance-like
860 values to facilitate comparison with the original raw abundances.

861

862 # References

863 1. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 103
864 (2020).
865 2. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*
866 **551**, 457–463 (2017).
867 3. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev*
868 *Microbiol* **15**, 579–590 (2017).
869 4. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359
870 (2015).
871 5. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's
872 Biogeochemical Cycles. *Science* **320**, 1034–1039 (2008).
873 6. Shreiner, A. B., Kao, J. Y. & Young, V. B. The gut microbiome in health and in disease: *Current*
874 *Opinion in Gastroenterology* **31**, 69–75 (2015).
875 7. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and
876 resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
877 8. Wu, G. *et al.* A core microbiome signature as an indicator of health. *Cell* **187**, 6550-6565.e11 (2024).
878 9. Pittayanon, R. *et al.* Differences in Gut Microbiota in Patients With vs Without Inflammatory Bowel
879 Diseases: A Systematic Review. *Gastroenterology* **158**, 930-946.e1 (2020).
880 10. Pittayanon, R. *et al.* Gut Microbiota in Patients With Irritable Bowel Syndrome—A Systematic Review.
881 *Gastroenterology* **157**, 97–108 (2019).
882 11. Chanda, D. & De, D. Meta-analysis reveals obesity associated gut microbial alteration patterns and
883 reproducible contributors of functional shift. *Gut Microbes* **16**, 2304900 (2024).
884 12. Zheng, J. *et al.* Noninvasive, microbiome-based diagnosis of inflammatory bowel disease. *Nat Med*
885 **30**, 3555–3567 (2024).
886 13. Liu, N.-N. *et al.* Multi-kingdom microbiota analyses identify bacterial–fungal interactions and
887 biomarkers of colorectal cancer across cohorts. *Nat Microbiol* **7**, 238–250 (2022).
888 14. Shkoporov, A. N. *et al.* The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell*
889 *Host & Microbe* **26**, 527-541.e5 (2019).

890 15. Cao, Z. *et al.* The gut virome: A new microbiome component in health and disease. *eBioMedicine* **81**,
891 104113 (2022).
892 16. Scanlan, P. D. Bacteria–Bacteriophage Coevolution in the Human Gut: Implications for Microbial
893 Diversity and Functionality. *Trends in Microbiology* **25**, 614–623 (2017).
894 17. Popescu, M., Van Belleghem, J. D., Khosravi, A. & Bollyky, P. L. Bacteriophages and the Immune
895 System. *Annu. Rev. Virol.* **8**, 415–435 (2021).
896 18. Rasmussen, T. S. *et al.* Bacteriophage-mediated manipulation of the gut microbiome – promises and
897 presents limitations. *FEMS Microbiology Reviews* **44**, 507–521 (2020).
898 19. Norman, J. M. *et al.* Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel
899 Disease. *Cell* **160**, 447–460 (2015).
900 20. Nakatsu, G. *et al.* Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival
901 Outcomes. *Gastroenterology* **155**, 529-541.e5 (2018).
902 21. Gogokhia, L. *et al.* Expansion of Bacteriophages Is Linked to Aggravated Intestinal Inflammation and
903 Colitis. *Cell Host & Microbe* **25**, 285-299.e8 (2019).
904 22. Leal Rodríguez, C. *et al.* The infant gut virome is associated with preschool asthma risk
905 independently of bacteria. *Nat Med* **30**, 138–148 (2024).
906 23. Ritz, N. L. *et al.* The gut virome is associated with stress-induced changes in behaviour and immune
907 responses in mice. *Nat Microbiol* **9**, 359–376 (2024).
908 24. Fan, G. *et al.* Alterations in the gut virome are associated with type 2 diabetes and diabetic
909 nephropathy. *Gut Microbes* **15**, 2226925 (2023).
910 25. Guo, J. *et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA
911 viruses. *Microbiome* **9**, 37 (2021).
912 26. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of
913 microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*
914 **8**, 90 (2020).
915 27. Camargo, A. P. *et al.* Identification of mobile genetic elements with geNomad. *Nat Biotechnol* **42**,
916 1303–1312 (2024).

917 28. Johansen, J. *et al.* Genome binning of viral entities from bulk metagenomics data. *Nat Commun* **13**,
918 965 (2022).
919 29. Chen, L. & Banfield, J. F. COBRA improves the completeness and contiguity of viral genomes
920 assembled from metagenomes. *Nat Microbiol* **9**, 737–750 (2024).
921 30. Nayfach, S. *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut
922 microbiome. *Nat Microbiol* **6**, 960–970 (2021).
923 31. Roux, S. *et al.* iPHoP: An integrated machine learning framework to maximize host prediction for
924 metagenome-derived viruses of archaea and bacteria. *PLoS Biol* **21**, e3002083 (2023).
925 32. Pinto, Y., Chakraborty, M., Jain, N. & Bhatt, A. S. Phage-inclusive profiling of human gut microbiomes
926 with Phanta. *Nat Biotechnol* **42**, 651–662 (2024).
927 33. Bharti, R. & Grimm, D. G. Current challenges and best-practice protocols for microbiome analysis.
928 *Briefings in Bioinformatics* **22**, 178–193 (2021).
929 34. Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L. & Prosperi, M. Challenges in the analysis
930 of viral metagenomes. *Virus Evol* **2**, vew022 (2016).
931 35. Gregory, A. C. *et al.* The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity
932 in the Human Gut. *Cell Host & Microbe* **28**, 724-740.e8 (2020).
933 36. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive
934 expansion of human gut bacteriophage diversity. *Cell* **184**, 1098-1109.e9 (2021).
935 37. Camargo, A. P. *et al.* IMG/VR v4: an expanded database of uncultivated virus genomes within a
936 framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Research* **51**,
937 D733–D743 (2023).
938 38. Shaw, J. & Yu, Y. W. Rapid species-level metagenome profiling and containment estimation with
939 sylph. *Nat Biotechnol* (2024) doi:10.1038/s41587-024-02412-y.
940 39. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput*
941 *Biol* **8**, e1002687 (2012).
942 40. Watts, S. C., Ritchie, S. C., Inouye, M. & Holt, K. E. FastSpar: rapid and scalable correlation
943 estimation for compositional data. *Bioinformatics* **35**, 1064–1066 (2019).

944 41. Edwards, R. A., McNair, K., Faust, K., Raes, J. & Dutilh, B. E. Computational approaches to predict
945 bacteriophage–host relationships. *FEMS Microbiology Reviews* **40**, 258–272 (2016).
946 42. Michel-Mata, S., Wang, X., Liu, Y. & Angulo, M. T. Predicting microbiome compositions from species
947 assemblages through deep learning. *iMeta* **1**, e3 (2022).
948 43. Wang, X.-W. *et al.* Identifying keystone species in microbial communities using deep learning. *Nat*
949 *Ecol Evol* **8**, 22–31 (2023).
950 44. Wang, T. *et al.* Predicting metabolomic profiles from microbial composition through neural ordinary
951 differential equations. *Nat Mach Intell* **5**, 284–293 (2023).
952 45. Shtossel, O., Koren, O., Shai, I., Rinott, E. & Louzoun, Y. Gut microbiome-metabolome interactions
953 predict host condition. *Microbiome* **12**, 24 (2024).
954 46. Tisza, M. J. *et al.* Longitudinal phage–bacteria dynamics in the early life gut microbiome. *Nat*
955 *Microbiol* **10**, 420–430 (2025).
956 47. Ram, G. & Solomon, S. Generalized Lotka-Volterra (glv) Models. in (1997).
957 48. An Illustrated Guide to Theoretical Ecology. *Journal of Mammalogy* **82**, 247–248 (2001).
958 49. Kimchi, O., Meir, Y. & Wingreen, N. S. Lytic and temperate phage naturally coexist in a dynamic
959 population model. *The ISME Journal* **18**, wrae093 (2024).
960 50. Dikareva, E. *et al.* An extended catalog of integrated prophages in the infant and adult fecal
961 microbiome shows high prevalence of lysogeny. *Front. Microbiol.* **14**, 1254535 (2023).
962 51. Kim, M.-S., Park, E.-J., Roh, S. W. & Bae, J.-W. Diversity and Abundance of Single-Stranded DNA
963 Viruses in Human Feces. *Appl Environ Microbiol* **77**, 8062–8070 (2011).
964 52. Sausset, R., Petit, M. A., Gaboriau-Routhiau, V. & De Paepe, M. New insights into intestinal phages.
965 *Mucosal Immunology* **13**, 205–215 (2020).
966 53. Hoyles, L. *et al.* Characterization of virus-like particles associated with the human faecal and caecal
967 microbiota. *Research in Microbiology* **165**, 803–812 (2014).
968 54. Styles, K. M., Brown, A. T. & Sagona, A. P. A Review of Using Mathematical Modeling to Improve Our
969 Understanding of Bacteriophage, Bacteria, and Eukaryotic Interactions. *Front. Microbiol.* **12**, 724767
970 (2021).

971 55. Pasolli, E. *et al.* Accessible, curated metagenomic data through ExperimentHub. *Nat Methods* **14**,
972 1023–1024 (2017).
973 56. Nayfach, S. *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol* **39**, 499–509 (2021).
974 57. Arboleya, S., Watkins, C., Stanton, C. & Ross, R. P. Gut Bifidobacteria Populations in Human Health
975 and Aging. *Front. Microbiol.* **7**, (2016).
976 58. Tzemah-Shahar, R. *et al.* Signs of aging in midlife: physical function and sex differences in
977 microbiota. *GeroScience* **46**, 1477–1488 (2023).
978 59. Torma, F. *et al.* Alterations of the gut microbiome are associated with epigenetic age acceleration and
979 physical fitness. *Aging Cell* **23**, e14101 (2024).
980 60. Squillario, M. *et al.* Gut-microbiota in children and adolescents with obesity: inferred functional
981 analysis and machine-learning algorithms to classify microorganisms. *Sci Rep* **13**, 11294 (2023).
982 61. Chen, X. *et al.* Characterization of the gut microbiota in Chinese children with overweight and obesity
983 using 16S rRNA gene sequencing. *PeerJ* **9**, e11439 (2021).
984 62. Bischoff, S. C. *et al.* Gut Microbiota Patterns Predicting Long-Term Weight Loss Success in
985 Individuals with Obesity Undergoing Nonsurgical Therapy. *Nutrients* **14**, 3182 (2022).
986 63. Hu, H.-J. *et al.* Obesity Alters the Microbial Community Profile in Korean Adolescents. *PLoS ONE* **10**,
987 e0134333 (2015).
988 64. MetaHIT Consortium (additional members) *et al.* Enterotypes of the human gut microbiome. *Nature*
989 **473**, 174–180 (2011).
990 65. Monshizadeh, M., Hong, Y. & Ye, Y. Multitask knowledge-primed neural network for predicting missing
991 metadata and host phenotype based on human microbiome. *Bioinformatics Advances* **5**, vbae203
992 (2024).
993 66. Shen, J., Zhang, D. & Liang, B. Prediction of host age and sex classification through gut microbes
994 based on machine learning. *Biochemical Engineering Journal* **178**, 108280 (2022).
995 67. Akiyama, S. *et al.* Multi-biome analysis identifies distinct gut microbial signatures and their crosstalk
996 in ulcerative colitis and Crohn's disease. *Nat Commun* **15**, 10291 (2024).
997 68. Clooney, A. G. *et al.* Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel
998 Disease. *Cell Host & Microbe* **26**, 764-778.e5 (2019).

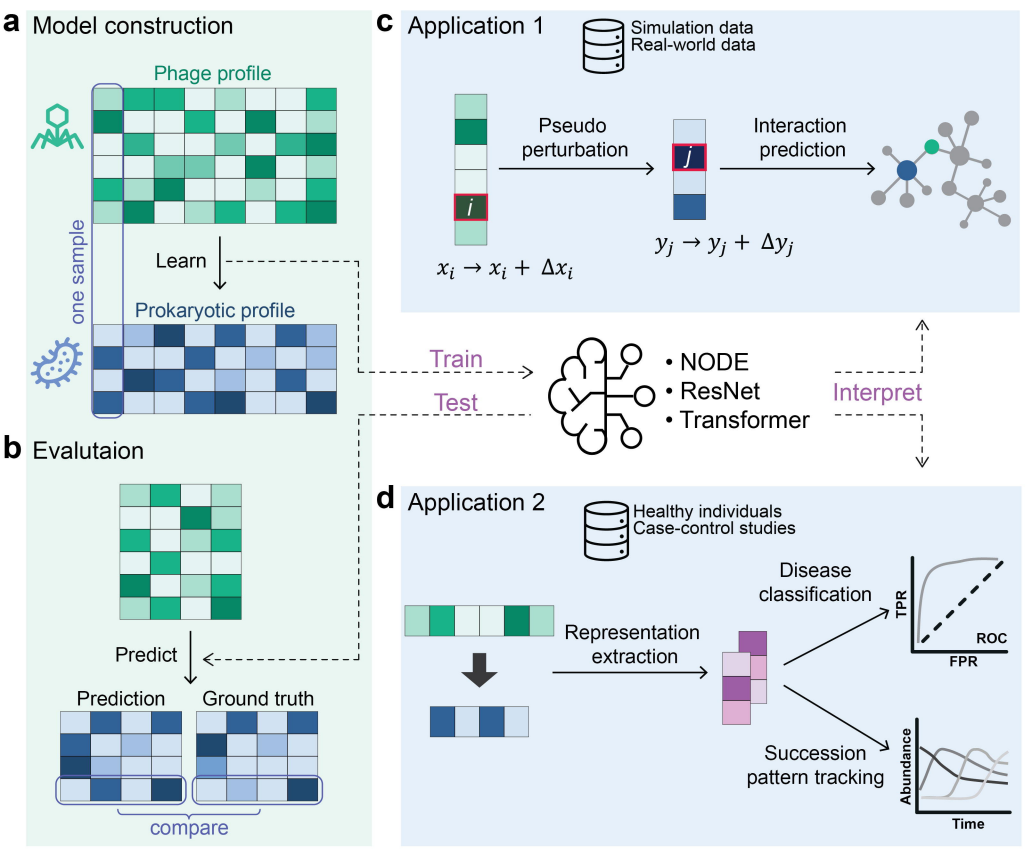
999 69. Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases.
1000 *Nature* **569**, 655–662 (2019).
1001 70. Zhang, Y. *et al.* Discovery of bioactive microbial gene products in inflammatory bowel disease. *Nature*
1002 **606**, 754–760 (2022).
1003 71. Mitchell, A. L. *et al.* EBI Metagenomics in 2017: enriching the analysis of microbial communities, from
1004 sequence reads to assemblies. *Nucleic Acids Research* **46**, D726–D735 (2018).
1005 72. Zheng, J. *et al.* Noninvasive, microbiome-based diagnosis of inflammatory bowel disease. *Nat Med*
1006 **30**, 3555–3567 (2024).
1007 73. Franzosa, E. A. *et al.* Gut microbiome structure and metabolic activity in inflammatory bowel disease.
1008 *Nat Microbiol* **4**, 293–305 (2018).
1009 74. Weng, Y. J. *et al.* Correlation of diet, microbiota and metabolite networks in inflammatory bowel
1010 disease. *J of Digest Diseases* **20**, 447–459 (2019).
1011 75. Yachida, S. *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes
1012 of the gut microbiota in colorectal cancer. *Nat Med* **25**, 968–976 (2019).
1013 76. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive
1014 biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).
1015 77. Vogtmann, E. *et al.* Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-
1016 Genome Shotgun Sequencing. *PLoS ONE* **11**, e0155362 (2016).
1017 78. Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are
1018 specific for colorectal cancer. *Nat Med* **25**, 679–689 (2019).
1019 79. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular*
1020 *Systems Biology* **10**, 766 (2014).
1021 80. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma–carcinoma sequence.
1022 *Nat Commun* **6**, 6528 (2015).
1023 81. Zackular, J. P., Rogers, M. A. M., Ruffin, M. T. & Schloss, P. D. The Human Gut Microbiome as a
1024 Screening Tool for Colorectal Cancer. *Cancer Prevention Research* **7**, 1112–1121 (2014).
1025 82. Yang, J. *et al.* Establishing high-accuracy biomarkers for colorectal cancer by comparing fecal
1026 microbiomes in patients with healthy families. *Gut Microbes* **11**, 918–929 (2020).

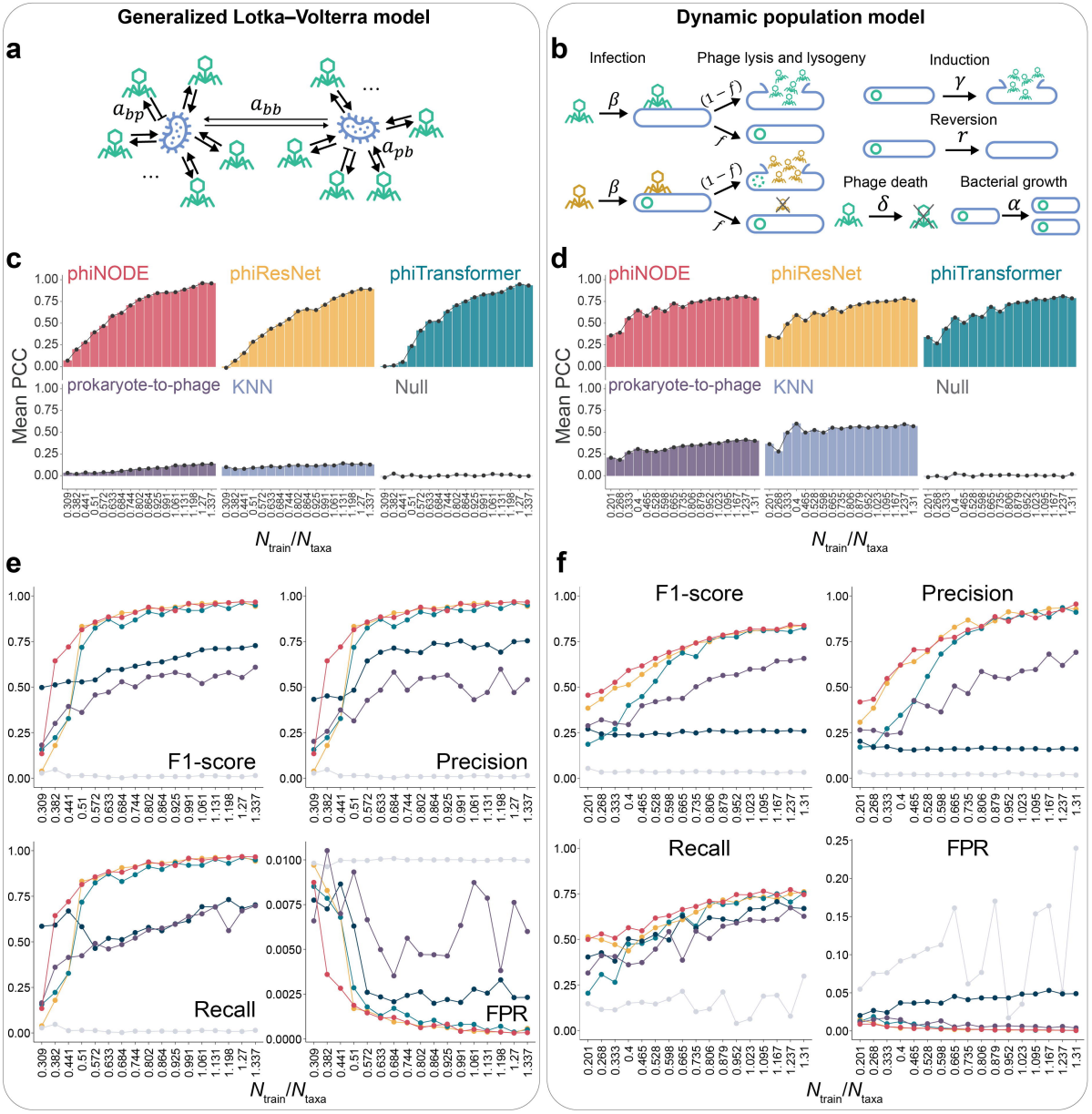
1027 83. Thomas, A. M. *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross-cohort
1028 microbial diagnostic signatures and a link with choline degradation. *Nat Med* **25**, 667–678 (2019).
1029 84. Gupta, A. *et al.* Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut
1030 Microbiome of Colorectal Cancer Patients in India. *mSystems* **4**, 10.1128/msystems.00438-19 (2019).
1031 85. Liu, N.-N. *et al.* Multi-kingdom microbiota analyses identifybacterial–fungal interactions and
1032 biomarkers of colorectal cancer acrosscohorts. *Nat Microbiol* **7**, 238–250 (2022).
1033 86. Yang, Y. *et al.* Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun*
1034 **12**, 6757 (2021).
1035 87. Zuo, W., Michail, S. & Sun, F. Metagenomic Analyses of Multiple Gut Datasets Revealed the
1036 Association of Phage Signatures in Colorectal Cancer. *Front. Cell. Infect. Microbiol.* **12**, 918010
1037 (2022).
1038 88. Thirion, F. *et al.* The gut microbiota in multiple sclerosis varies with disease activity. *Genome Med* **15**,
1039 1 (2023).
1040 89. Boktor, J. C. *et al.* Integrated Multi-Cohort Analysis of the Parkinson's Disease Gut Metagenome.
1041 *Movement Disorders* **38**, 399–409 (2023).
1042 90. Wallen, Z. D. *et al.* Metagenomics of Parkinson's disease implicates the gut microbiome in multiple
1043 disease mechanisms. *Nat Commun* **13**, 6958 (2022).
1044 91. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64 (2014).
1045 92. Loomba, R. *et al.* Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of
1046 Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metabolism* **25**, 1054-1062.e5
1047 (2017).
1048 93. Odenwald, M. A. *et al.* Bifidobacteria metabolize lactulose to optimize gut metabolites and prevent
1049 systemic infection in patients with liver disease. *Nat Microbiol* **8**, 2033–2049 (2023).
1050 94. Worby, C. J. *et al.* Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent
1051 urinary tract infections in women. *Nat Microbiol* **7**, 630–639 (2022).
1052 95. Mars, R. A. T. *et al.* Longitudinal Multi-omics Reveals Subset-Specific Mechanisms Underlying
1053 Irritable Bowel Syndrome. *Cell* **182**, 1460-1473.e17 (2020).

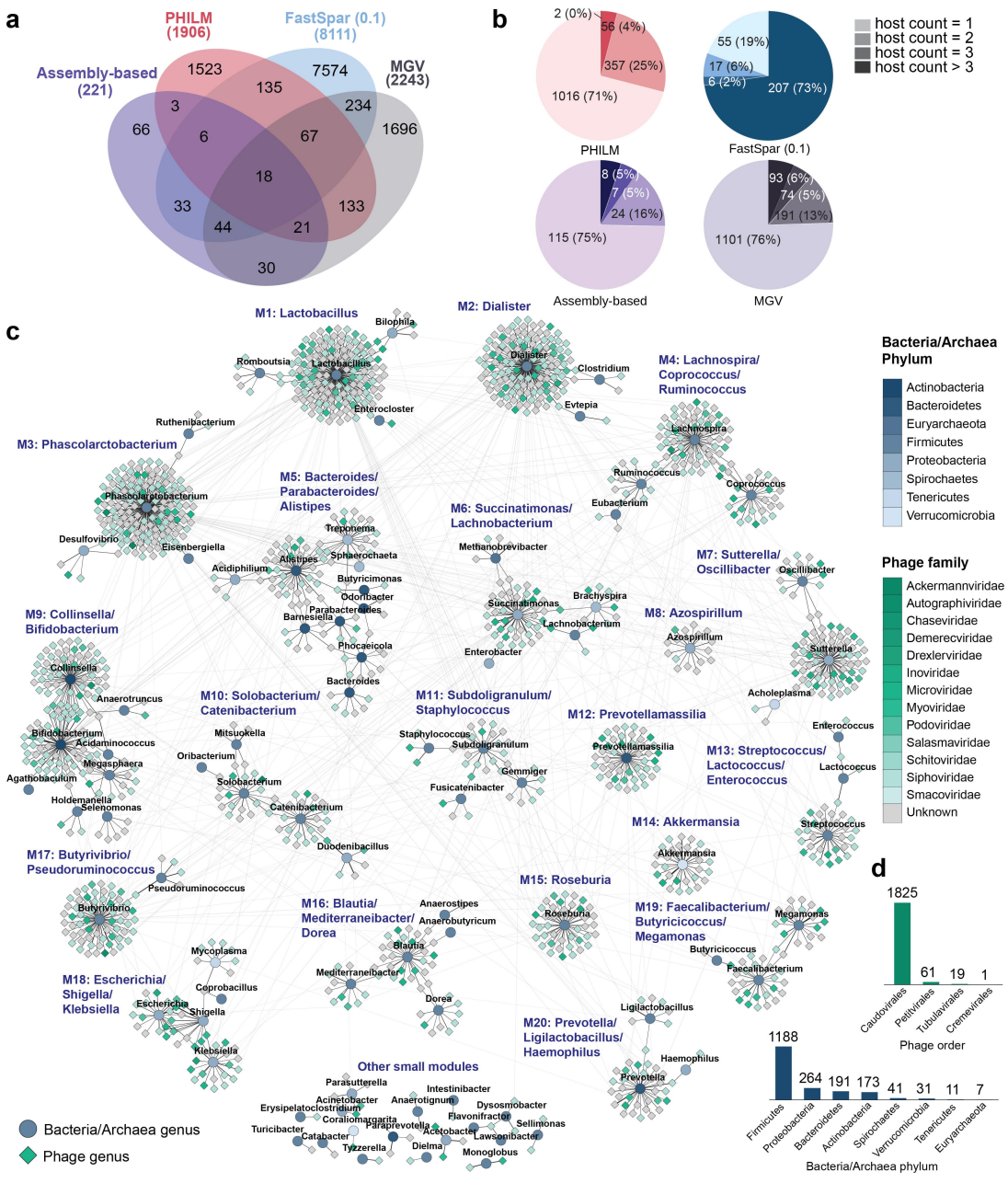
1054 96. Nagata, N. *et al.* Human Gut Microbiota and Its Metabolites Impact Immune Responses in COVID-19
1055 and Its Complications. *Gastroenterology* **164**, 272–288 (2023).
1056 97. Yeoh, Y. K. *et al.* Gut microbiota composition reflects disease severity and dysfunctional immune
1057 responses in patients with COVID-19. *Gut* **70**, 698–706 (2021).
1058 98. Noguera-Julian, M. *et al.* Gut Microbiota Linked to Sexual Preference and HIV Infection.
1059 *EBioMedicine* **5**, 135–146 (2016).
1060 99. Bai, X. *et al.* Whole-Genome Metagenomic Analysis of the Gut Microbiome in HIV-1-Infected
1061 Individuals on Antiretroviral Therapy. *Front. Microbiol.* **12**, 667718 (2021).
1062 100. Lu, W. *et al.* Association Between Gut Microbiota and CD4 Recovery in HIV-1 Infected Patients.
1063 *Front. Microbiol.* **9**, 1451 (2018).
1064 101. Rocafort, M. *et al.* Evolution of the gut microbiome following acute HIV-1 infection. *Microbiome* **7**,
1065 73 (2019).
1066 102. Li, J. *et al.* Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome*
1067 **5**, 14 (2017).
1068 103. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
1069 *Bioinformatics* **30**, 2114–2120 (2014).
1070 104. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–
1071 359 (2012).
1072 105. Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems*
1073 **30**, (2017).
1074 106. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016*
1075 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, Las Vegas,
1076 NV, USA, 2016). doi:10.1109/CVPR.2016.90.
1077 107. Chen, R. T., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential
1078 equations. *Advances in neural information processing systems* **31**, (2018).
1079 108. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. Preprint at
1080 https://doi.org/10.48550/ARXIV.1412.6980 (2014).

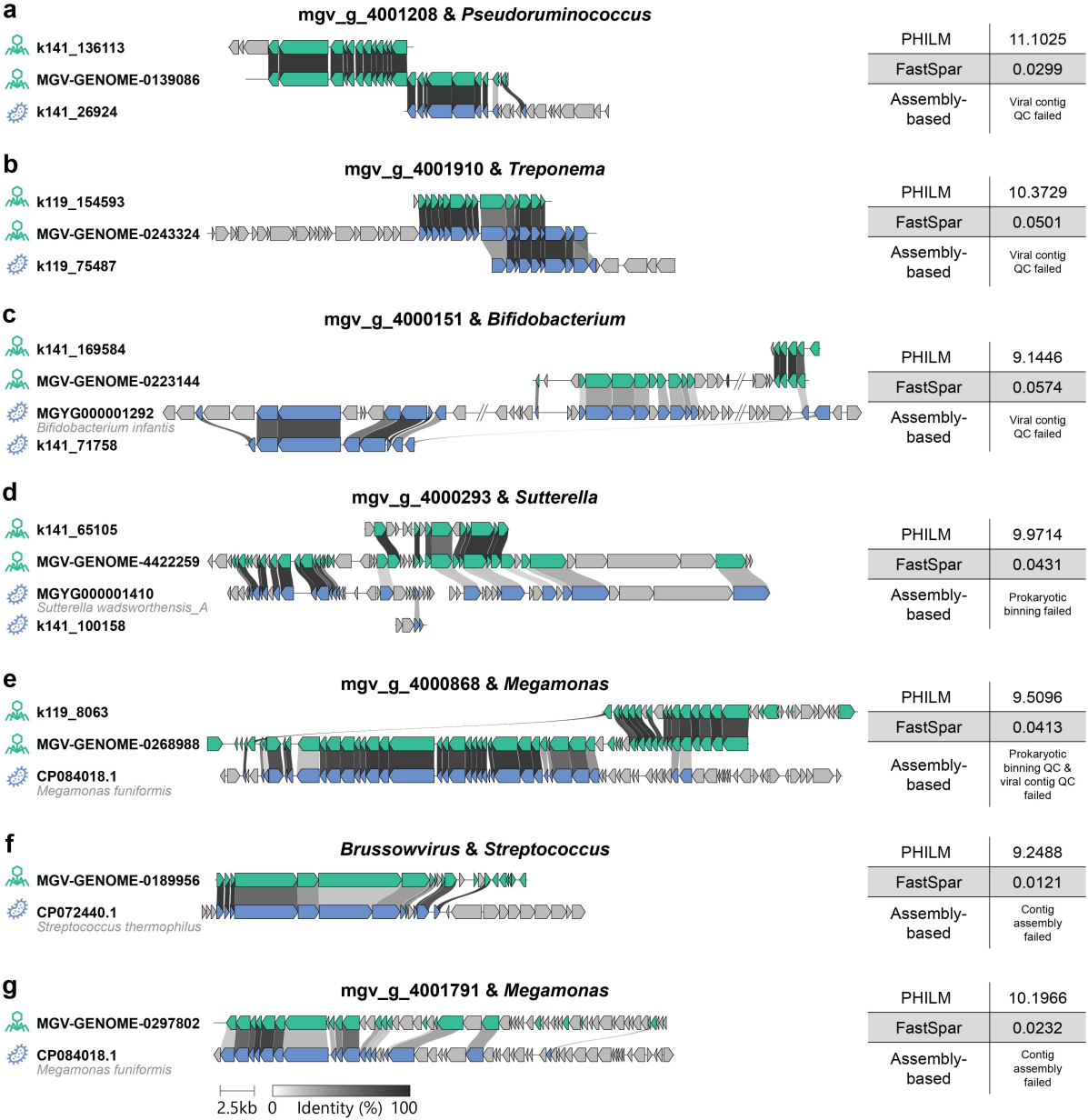
1081 109. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A Next-generation
1082 Hyperparameter Optimization Framework. in *Proceedings of the 25th ACM SIGKDD International*
1083 *Conference on Knowledge Discovery & Data Mining* 2623–2631 (ACM, Anchorage AK USA, 2019).
1084 doi:10.1145/3292500.3330701.
1085 110. Paszke, A. *et al.* Automatic differentiation in PyTorch. (2017).
1086 111. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular
1087 Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
1088 112. Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLay: community structure analysis of
1089 biological networks. *Bioinformatics* **26**, 3135–3137 (2010).
1090 113. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome.
1091 *Nat Biotechnol* **39**, 105–114 (2021).
1092 114. Youngblut, N. & Shen, W. nick-youngblut/gtdb_to_taxdump: Zenodo release. Zenodo
1093 https://doi.org/10.5281/ZENODO.3696964 (2020).
1094 115. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for
1095 large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* **31**, 1674–
1096 1676 (2015).
1097 116. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome
1098 reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
1099 117. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the
1100 quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.*
1101 **25**, 1043–1055 (2015).
1102 118. Von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust
1103 taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol*
1104 **20**, 217 (2019).
1105 119. Hiseni, P., Rudi, K., Wilson, R. C., Hegge, F. T. & Snipen, L. HumGut: a comprehensive human gut
1106 prokaryotic genomes collection filtered by metagenome data. *Microbiome* **9**, 165 (2021).

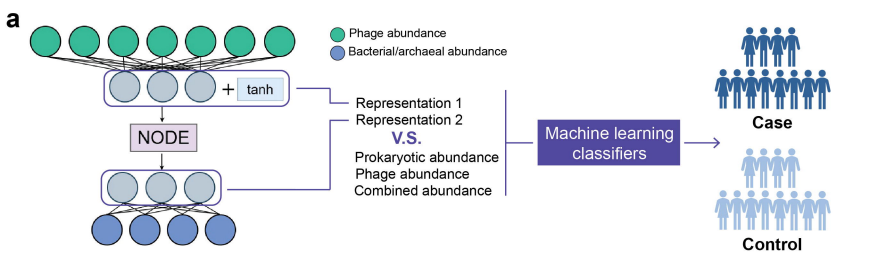
1107 120. Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A. & Sørensen, S. J.
1108 CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. *The*
1109 *CRISPR Journal* **3**, 462–469 (2020).
1110 121. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled
1111 viral genomes. *Nat Biotechnol* **39**, 578–585 (2021).
1112 122. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
1113 123. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning*
1114 *Research* **12**, 2825–2830 (2011).
1115
1116


a
Model construction
Phage profile
one sample
Learn
Prokaryotic profile
b
Evalutaion
Predict
Prediction
Ground truth
compare
Train
Test
• NODE
• ResNet
• Transformer
Interpret
c
Application 1
Simulation data
Real-world data
Pseudo perturbation
Interaction prediction
$x_i \rightarrow x_i + \Delta x_i$
$y_j \rightarrow y_j + \Delta y_j$
d
Application 2
Healthy individuals
Case-control studies
Representation extraction
Disease classification
TPR
FPR
ROC
Succession pattern tracking
Abundance
Time


Generalized Lotka–Volterra model
Dynamic population model
a
b
Infection
Phage lysis and lysogeny
Induction
Reversion
Phage death
Bacterial growth
c
d
phiNODE
phiResNet
phiTransformer
prokaryote-to-phage
KNN
Null
Mean PCC
N_{train}/N_{taxa}
e
f
F1-score
Precision
Recall
FPR
phiNODE
phiResNet
phiTransformer
prokaryote-to-phage
FastSpar (best thresholds)
Null
PHILM models


a
PHILM (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
host count = 1
host count = 2
host count = 3
host count > 3
2 (0%)
56 (4%)
357 (25%)
1016 (71%)
PHILM
55 (19%)
17 (6%)
6 (2%)
207 (73%)
FastSpar (0.1)
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
M2: Dialister
Dialister
Clostridium
Evtepia
M4: Lachnospira/ Coprococcus/ Ruminococcus
Lachnospira
Ruminococcus
Coprococcus
Eubacterium
Ruthenibacterium
M3: Phascolarctobacterium
Phascolarctobacterium
Desulfovibrio
Eisenbergiella
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
M17: Butyrivibrio/ Pseudoruminococcus
Pseudoruminococcus
Butyrivibrio
M15: Roseburia
Roseburia
M16: Blautia/ Mediterraneibacter/ Dorea
Anaerostipes
Anaerobutyricum
Blautia
Mediterraneibacter
Dorea
M19: Faecalibacterium/ Butyricicoccus/ Megamonas
Megamonas
Butyricicoccus
Faecalibacterium
M18: Escherichia/ Shigella/ Klebsiella
Mycoplasma
Coprobacillus
Escherichia
Shigella
Klebsiella
M20: Prevotella/ Ligilactobacillus/ Haemophilus
Ligilactobacillus
Haemophilus
Prevotella
Other small modules
Parasutterella
Intestinibacter
Acinetobacter
Anaerotignum
Dysosmobacter
Erysipelatoclostridium
Flavonifractor
Coralionargarita
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
PHILM 11.1025
FastSpar 0.0299
Assembly-based Viral contig QC failed
b
mgv_g_4001910 & Treponema
k119_154593
MGV-GENOME-0243324
k119_75487
PHILM 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
PHILM 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
PHILM 9.9714
FastSpar 0.0431
Assembly-based Prokaryotic binning failed
e
mgv_g_4000868 & Megamonas
k119_8063
MGV-GENOME-0268988
CP084018.1
Megamonas funiformis
PHILM 9.5096
FastSpar 0.0413
Assembly-based Prokaryotic binning QC & viral contig QC failed
f
Brussowvirus & Streptococcus
MGV-GENOME-0189956
CP072440.1
Streptococcus thermophilus
PHILM 9.2488
FastSpar 0.0121
Assembly-based Contig assembly failed
g
mgv_g_4001791 & Megamonas
MGV-GENOME-0297802
CP084018.1
Megamonas funiformis
PHILM 10.1966
FastSpar 0.0232
Assembly-based Contig assembly failed
2.5kb
0
Identity (%)
100



b

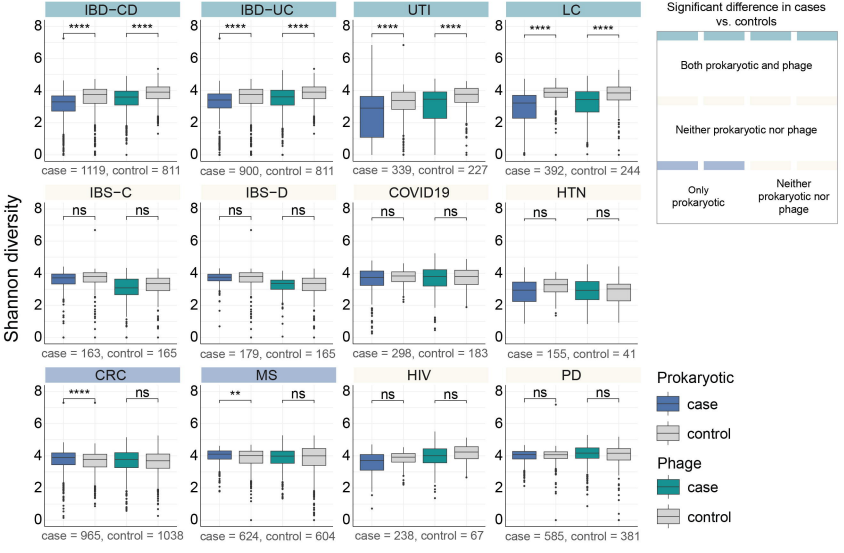
Mean AUROC: high – low

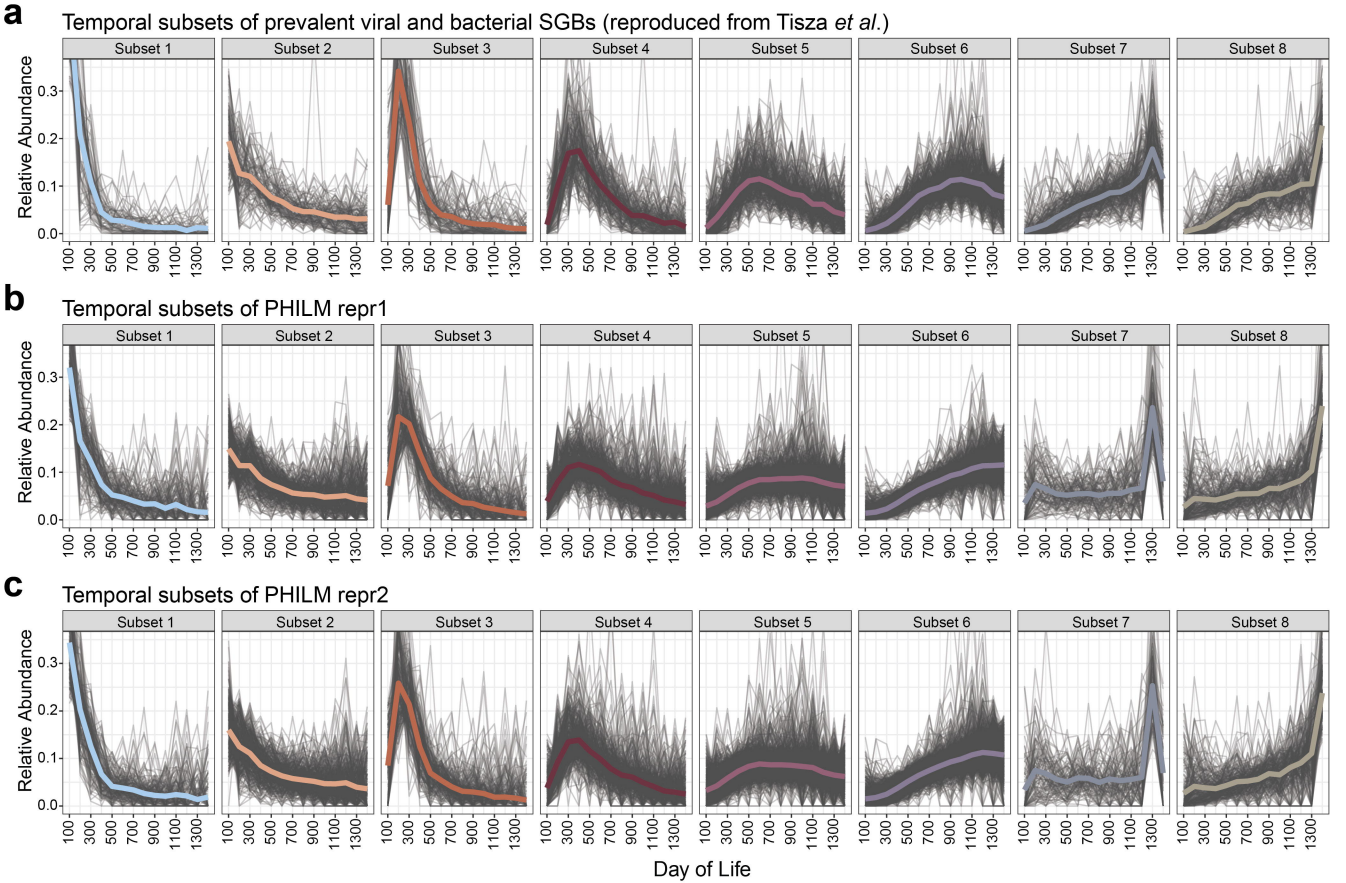
PHILM repr > others

PHILM repr < others

	LR	SVM	RF	LR	SVM	RF	LR	SVM	RF	LR	SVM	RF
	IBD-CD			IBD-UC			UTI			LC		
repr1	0.9611	**0.9629**	0.9592	0.9544	0.9361	0.9438	**0.9801**	0.9640	0.9539	0.9569	**0.9594**	0.9188
repr2	0.9614	0.9619	0.9575	**0.9564**	0.9420	0.9463	0.9782	0.9619	0.9499	0.9581	0.9581	0.9194
bact_arc	0.8868	0.9260	0.9477	0.9000	0.9237	0.9418	0.9460	0.9523	0.9489	0.7822	0.8772	0.9474
phage	0.8620	0.8606	0.9556	0.8380	0.8057	0.9290	0.9677	0.9214	0.9470	0.5263	0.5896	0.8456
combined	0.8864	0.8795	0.9541	0.8345	0.8319	0.9408	0.9582	0.9319	0.9437	0.5304	0.6033	0.9335
	IBS-C			IBS-D			COVID19			HTN		
repr1	**0.9925**	0.9868	0.9834	0.9866	0.9752	0.9618	0.8523	**0.8771**	0.8546	**0.6530**	0.5910	0.5041
repr2	0.9913	0.9889	0.9791	**0.9868**	0.9756	0.9586	0.8566	0.8786	0.8697	0.6503	0.6022	0.4899
bact_arc	0.9260	0.9520	0.9479	0.9266	0.9247	0.9386	0.6351	0.7728	0.8465	0.5219	0.5025	0.6257
phage	0.9829	0.9671	0.9345	0.9787	0.9674	0.9509	0.4677	0.5343	0.7636	0.5000	0.4968	0.5032
combined	0.9333	0.9248	0.9584	0.9230	0.9282	0.9426	0.5384	0.5410	0.8152	0.4964	0.4931	0.6032
	CRC			MS			HIV			PD		
repr1	0.7349	0.7837	0.7465	0.9611	0.9762	0.9436	0.6323	0.6649	0.6276	0.6646	0.7240	0.7145
repr2	0.7345	0.7887	0.7496	0.9607	0.9762	0.9454	0.6388	0.6650	0.6502	0.6648	0.7246	0.7156
bact_arc	0.6658	0.7144	**0.8073**	0.9550	**0.9825**	0.9704	0.5700	0.5928	**0.6847**	0.6129	0.6721	**0.7667**
phage	0.5667	0.5662	0.6494	0.8001	0.9016	0.9309	0.5000	0.5000	0.6400	0.5148	0.5070	0.6150
combined	0.5698	0.5739	0.7712	0.8465	0.8735	0.9590	0.5144	0.5054	0.6743	0.5232	0.5136	0.7312

c




a
Temporal subsets of prevalent viral and bacterial SGBs (reproduced from Tisza et al.)
b
Temporal subsets of PHILM repr1
c
Temporal subsets of PHILM repr2
Subset 1
Subset 2
Subset 3
Subset 4
Subset 5
Subset 6
Subset 7
Subset 8
Relative Abundance
0.0
0.1
0.2
0.3
100
300
500
700
900
1100
1300
Day of Life