



Predictive phage therapy for *Escherichia coli* urinary tract infections: Cocktail selection for therapy based on machine learning models

Marianne Keith^{a,1} , Alba Park de la Torriente^{a,1} , Antonia Chalka^{a,1} , Adriana Vallejo-Trujillo^a , Sean P. McAteer^a , Gavin K. Paterson^{a,b} , Alison S. Low^{a,2} , and David L. Gally^{a,2}

Edited by Bruce Levin, Emory University, Atlanta, GA; received August 8, 2023; accepted February 4, 2024

This study supports the development of predictive bacteriophage (phage) therapy: the concept of phage cocktail selection to treat a bacterial infection based on machine learning (ML) models. For this purpose, ML models were trained on thousands of measured interactions between a panel of phage and sequenced bacterial isolates. The concept was applied to *Escherichia coli* associated with urinary tract infections. This is an important common infection in humans and companion animals from which multidrug-resistant (MDR) bloodstream infections can originate. The global threat of MDR infection has reinvigorated international efforts into alternatives to antibiotics including phage therapy. *E. coli* exhibit extensive genome-level variation due to horizontal gene transfer via phage and plasmids. Associated with this, phage selection for *E. coli* is difficult as individual isolates can exhibit considerable variation in phage susceptibility due to differences in factors important to phage infection including phage receptor profiles and resistance mechanisms. The activity of 31 phage was measured on 314 isolates with growth curves in artificial urine. Random Forest models were built for each phage from bacterial genome features, and the more generalist phage, acting on over 20% of the bacterial population, exhibited F1 scores of >0.6 and could be used to predict phage cocktails effective against previously untested strains. The study demonstrates the potential of predictive ML models which integrate bacterial genomics with phage activity datasets allowing their use on data derived from direct sequencing of clinical samples to inform rapid and effective phage therapy.

E. coli | bacteriophage | machine learning | predictive phage therapy | UTI

Uropathogenic *Escherichia coli* (UPEC) are among the most common causes of urinary tract infections (UTIs) in both humans and companion animals. It is reported that up to 80% of UTIs in humans and 35 to 69% of UTIs in small animal pets are caused by UPEC (1, 2). UTIs are widespread, affecting an estimated 150 million people a year worldwide (3), and 40% of women will develop a UTI during their lifetime (4). It is also reported that 14% of dogs will suffer with a UTI over their lifetime (5). UPEC employ several virulence factors that give them the capacity to colonize, invade, and survive outside of the intestine (3). They can transfer from the intestinal environment, ascend and colonize the urinary tract often causing uncomplicated transient infections that self-resolve. However, the likelihood of antibiotic treatment increases if the infection persists and/or the symptoms are more severe. Bladder (cystitis) infections can sometimes ascend to the kidney(s) (pyelonephritis) and potentially into the bloodstream resulting in bacteremia. Up to 50% of human sepsis cases associated with *Escherichia coli* (*E. coli*) originate from a UTI (6, 7). Conversely and much less commonly, kidney infections and UTIs can result from a bloodstream infection. Recurrent and chronic infection increases the exposure of strains to multiple antibiotics and therefore the potential emergence of multidrug-resistant bacteria (MDR) (3, 8, 9) and treatment failure. Trimethoprim-sulfamethoxazole, fluoroquinolones, and cephalosporins are common first-line antibiotics used to treat UTIs in humans. Antibiotics from these classes are also used to treat pets with UTIs (5, 10). Antibiotic resistance is a growing problem globally and has serious consequences, including increased healthcare costs and the spread of potentially life-threatening infections. High-risk globally disseminated MDR human UPEC strain types such as ST131 and ST410 are now being reported in veterinary settings (11). Alternative non-antibiotic treatments for UTIs are urgently required to help prevent further development of antibiotic-resistant bacteria and offer treatment options for MDR infections. Bacteriophage (phage) are a promising alternative although their routine application as a therapeutic is going to require further advances in our understanding of phage-bacterium interactions (12).

Significance

With the growing challenge of antimicrobial resistance, there is an urgency for alternative treatments for common bacterial diseases including urinary tract infections (UTIs). *Escherichia coli* is the main causative agent of UTIs in both humans and companion animals with multidrug-resistant strains such as the globally disseminated ST131 becoming more common. Bacteriophage (phage) are natural predators of bacteria and potentially an alternative therapy. However, a major barrier for phage therapy is the specificity of phage on bacteria, and therefore, it is difficult to efficiently select the appropriate phage. Here, we demonstrate a genomics-driven approach using machine learning prediction models combined with phage activity clustering to select phage cocktails based only on the genome sequence of the infecting bacterial strain.

Author contributions: A.S.L. and D.L.G. designed research; M.K., A.P.d.I.T., A.C., S.P.M., and A.S.L. performed research; G.K.P. contributed new reagents/analytic tools; M.K., A.P.d.I.T., A.C., A.V.-T., and A.S.L. analyzed data; G.K.P. scientific discussion of ideas and data; and A.C., A.S.L., and D.L.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹M.K., A.P.d.I.T., and A.C. contributed equally to this work.

²To whom correspondence may be addressed. Email: alow@ed.ac.uk or dgally@ed.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2313574121/-/DCSupplemental>.

Published March 13, 2024.

The use of phage, viruses that specifically kill bacteria, is now gaining traction as an alternative to antibiotics for the treatment of bacterial infections. Phage were first described over 100 y ago by Felix d'Herelle (13) and used to treat bacterial infections in this era (14, 15). However, the advent and routine use of antibiotics since the 1940s meant phage therapy development stalled globally. As a result of the rising threat of AMR, there has been a renaissance in phage research with major advances in the characterization of bacterial resistance mechanisms (16, 17) and knowledge of the counteroffensive strategies evolved by phage (18, 19). Phage as a therapy have many advantages; they are naturally occurring, usually highly specific, nontoxic, effective on antibiotic-resistant bacteria, and self-dosing. However, the specificity is also a disadvantage as ideally a therapeutic phage needs to be matched to an infecting strain, traditionally done by manual screening using agar overlay assays. Phage are sought that are “generalist” meaning that they are active on a reasonable proportion of an infecting species, but this is a much greater challenge for some species over others depending on their genome plasticity. This includes *E. coli*, as integrated prophage play a major role in their diversification and often introduce defense mechanisms against other phage to promote their own survival.

To overcome phage specificity issues and potential resistance development, phage therapy has often been developed using cocktails of phage, usually between 2 and 10 phage (20). While phage interference can occur especially if too many phage are included, such cocktails are more likely to include phage that can predate on the infecting strain, and conceptually, there is a reduced likelihood that the bacteria can develop resistance if the phage use different infection pathways (21–23). Even with cocktails, the incredible diversity of *E. coli* associated with UTIs in human and companion animals makes phage selection for effective treatment a major challenge. AI and more specifically machine learning (ML) analysis combined with advances in genomics enables the development of “smarter”, personalized medicine approaches within the field of phage therapy. ML is a computational technique capable of analyzing large and complex datasets (24) which has already been applied for predicting viral hosts (25–27). However, previous studies have focused on species-level host differentiation, whereas phage therapy will require a greater level of resolution for host attribution, down to individual bacterial isolates. To that end, we propose that ML models custom-built around predicting infection efficacy of phage on bacterial isolates of a species can be used to design personalized phage cocktails for patients, aiming to achieve improved efficacy over a generic cocktail.

To help address a knowledge gap of bacterial-phage interactions in clinically relevant conditions and move toward a predictive approach to phage therapy, we have generated a large dataset of phage-UPEC interactions in an artificial urine (AU) medium. This has been used to train ML models for predictive phage therapy but also gives a wealth of information to inform the selection of phage, including activity groups and distribution of resistance mechanisms. Coupled to the direct sequencing of infected urine samples (28), the study shows how bespoke therapeutic phage cocktail selection can be quickly achieved using a bacterial genomics approach and appropriate interaction datasets.

Results

Local Epidemiology of UPEC Strains for Phage Treatment. For this study, focused on UTIs, 314 UPEC strains from both canine ($n = 203$) and human ($n = 111$) patients in the Edinburgh area were whole genome sequenced (Illumina, MicrobesNG). This was required as input data to train the ML models but also enabled an

understanding of local strain diversity. In addition, it facilitated the selection of representative strains for phage enrichment to expand our phage collection from wastewater samples. Fig. 1A shows the phylogenetic tree alongside their source host for the 314 strains as well as a set of 10 validation strains used later in the study. The sequences were uploaded to Enterobase allowing assessment of phylotype, sequence type (ST) and O-antigen. Some phage use components of the lipopolysaccharide including the O-antigen portion as host receptors and other phage activity may cluster with phylogeny so mapping this information against phage activity could be important for rational cocktail design even without predictive ML models.

As expected, most of the *E. coli* strains in the UPEC collection belonged to phylotype B2 and D, the phylotypes most associated with extraintestinal disease in humans and domestic animals (Fig. 1B) (29, 30). The human strains had a higher proportion associated with phylotype D than the canine strains (18.9% of strains compared with 4.4%). The differences in predominant ST and to a lesser extent O-antigen types between the human and the canine-derived isolates (Fig. 1C and D) indicate that different types of *E. coli* strains are associated with UTIs in these two hosts locally and this is likely to impact phage specificity. ST131, ST69, ST95, and ST73 are globally disseminated pathogenic *E. coli* lineages that are associated with UTI and bloodstream infections (31), and these are the most abundant ST groups represented in the human UPEC set. ST73 is also an abundant ST group in the canine strains (6.9%) although ST372 and ST12 represent almost 38% of the canine isolates.

Generation of a Host-Phage Interaction Dataset. To develop ML models and better understand genotype to phage activity phenotype, an interaction dataset is required for training and analysis. While this potentially could be mined from online databases, the methods used would be varied and there would be a lack of negative data. To overcome this, and to test for activity in conditions more closely reflecting those in vivo, we generated an interaction dataset by growing 314 bacterial UPEC strains in an AU medium in 96-well plate growth assays and challenging them with a set of 31 phage. Ultimately phage selected to treat UTIs need to be active in the host environment and to assess phage activity against UPEC the ideal growth medium would be urine but repeated urine collection would introduce too much variability. Our preliminary work for this study therefore included development of an AU medium based on published studies (SI Appendix, Fig. S1A). The AU approximated canine pooled urine in terms of growth dynamics and phage activity and also demonstrated that the activity of certain phage could be lost when the bacteria were cultured in urine and AU compared to Lysogeny/Luria Broth (SI Appendix, Fig. S1B).

Phage were initially selected from our phage library by screening of activity against a small subset of UPEC strains using LB plate assays. Around 60 phage were then assayed in liquid AU against 38 bacterial isolates selected to represent the diversity of the wider UPEC collection. From this data, a final set of 31 phage were selected to generate interaction data for the remaining 284 bacterial isolates (314 in total), >9,000 interactions, each measured in triplicate against a no-phage control. The final phage set were primarily selected as those showing the broadest host range and diverse activity profiles but some narrower range phage and phage with very related activity were also included. An interaction score was determined for each strain-phage combination using the ratio of the area under the curve (AUC) for phage treatment over a no-phage control culture (SI Appendix, Table S1). The scores were plotted as a heatmap to visualize patterns of phage activity across

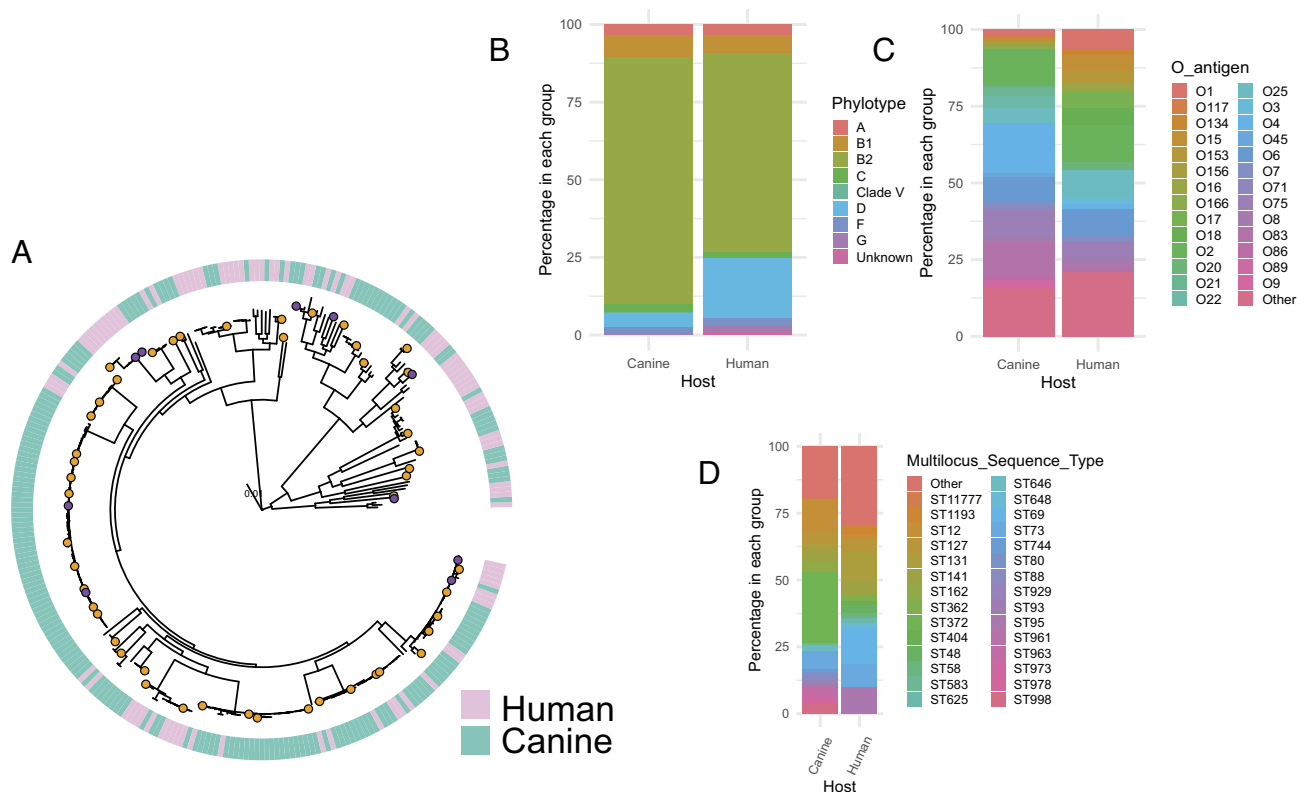


Fig. 1. Phylogeny of UPEC strains used in the study and breakdown of phylotype, O-Antigen, and ST groupings. (A) Phylogenetic tree showing the relationship of the UPEC strains from the collection used in this study (314 strains in the training set and 10 strains from the validation set). The host source (canine or human) is shown in the outer ring. Orange circles show the 50 representative strains used to test the general and bespoke cocktails and the purple circles show the 10 “unseen” strains used to validate the ML models. (B–D) Charts showing the percentage of the canine and human strains ($n = 314$) belonging to the different phylotypes (B), O-Antigen type (C), and ST as defined by Enterobase.

the strain collection (Fig. 2) with any interaction that scored over 100 being given the score of 100 for the purposes of the heatmap. A score of 100 indicates no activity while scores less than 60 were taken as clear evidence of phage interaction and reduced bacterial growth (*Materials and Methods*). With this cutoff, only two phage in our test set showed activity on more than 40% of strains in the bacterial collection, Chap1 (54%) and Nea2 (41%) (Figs. 2 and 3A). Four percent of strains ($n = 14$) were completely resistant to all 31 phage tested (all phage interactions scored higher than 90). After generation of this dataset, these strains were used to enrich for additional phage to broaden the host range of the phage collection.

ML Predictions of Phage Activity. The principal aim of this work was to generate an interaction database that could be used, with features extracted from bacterial whole genome sequences, to train ML models to predict phage activity based on an *E. coli* sequence. This is a “proof-of-principle” study with 31 phage to gain insight into the capacity of the method to predict phage activity, and design effective cocktails, without further manual screening.

The main genomic features used for training ML models were predicted genes generated by a pangenomic analysis and defined as protein variants (PVs). Separate ML models were generated for each phage, resulting in a total of 31 ML models. ML models were initially evaluated with Root Mean Square Error (RMSE) scores, which are the square root of the average of the squared differences between predicted and observed values (available in *SI Appendix, Table S2*). The individual ML models (Fig. 3C and *SI Appendix, Fig. S3*) had average RMSE values between 8 and 23 (working with the 0 to 100 interaction score range). The interaction datasets for some of the phage were very imbalanced (Fig. 3A), which

required a more stringent investigation of model bias. To reflect their use in a clinical setting, observed and predicted interaction scores were classified either side of a threshold score of 60 (below 60 defined as a positive (active) interaction and above 60 a negative (inactive) interaction). A confusion matrix was generated for each phage model, and the ML models were further evaluated on classifier-based metrics; F1 (Fig. 3B), recall and precision (*SI Appendix, Fig. S2*). F1, which is the harmonic mean of precision and recall, was chosen as the performance metric, taking class imbalance into account, and there was a clear trend that more generalist phage yielded more reliable ML models with correspondingly higher F1 scores (Fig. 3B).

The spread of predicted vs. observed scores for the most generalist phage Chap1 ($F1 = 0.83$) is shown in Fig. 3C. Similar figures for the other 30 phage ML models can be found in *SI Appendix, Fig. S3*. The information was also plotted across the phylogeny of the strains, based on core SNP variation, and activity as well as significant positive or negative deviation of the predicted score from the measured score was generally distributed across the phylogeny and not limited to one subcluster (Fig. 3D and *SI Appendix, Fig. S4*).

Assessment of Different Approaches to Select Phage Cocktails.

Clinically, phage are often administered as a cocktail of multiple phage. The advantage of using a phage cocktail is twofold: i) the construction of a general cocktail that could be used “off the shelf” to treat an infection where the inclusion of a range of phage should increase the chance at least one is effective against the infection and ii) a bespoke cocktail where the included phage are matched to the infection and so should all be effective and the inclusion of multiple active phage should help overcome the development of bacterial resistance to the phage treatment. In both these cases, it

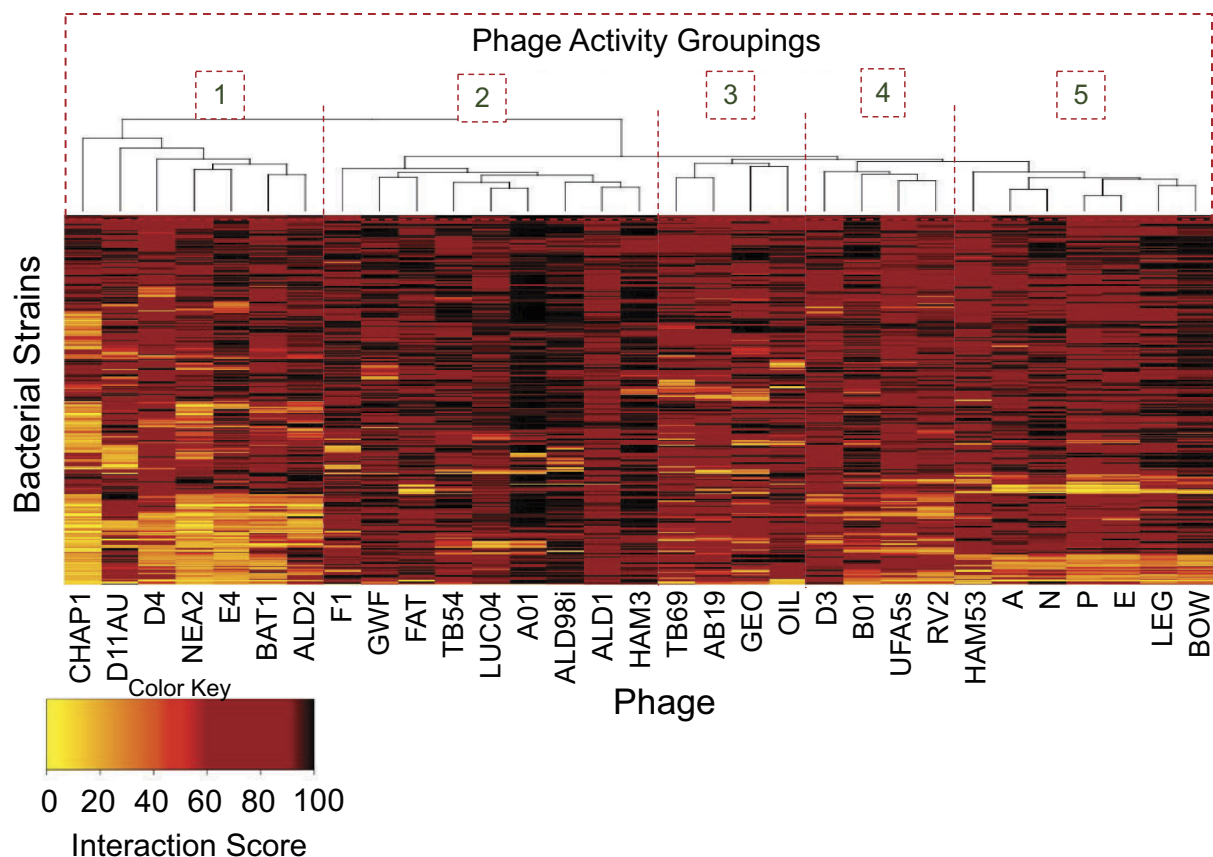


Fig. 2. Heatmap of bacterial ($n = 314$) phage ($n = 31$) interactions based on interaction scores from growth assay in AU. Low interaction scores (yellow) represent high phage activity and good inhibition of bacterial growth. High interaction scores (dark red) represent limited phage activity. Using heatmap2 in gplots (v3.1.3), pairwise comparisons of the data interactions allowed the phage to be clustered into five primary activity groups to aid with cocktail selection.

is useful to include phage with diverse activity and therefore an increased probability they use different mechanisms to infect the host bacteria. Here, the phage were clustered by activity on the bacterial strains generating five primary activity groupings. For this, a matrix with the bacterium-phage interaction scores was used to produce a heatmap (gplots version 3.1.3) and dendrogram based on pairwise comparisons (Fig. 2). Cocktails were generated based on using one phage from each activity group. Chap1, F1, TB69, RV2, and Phage P were selected for a general cocktail, as these were the broadest range phage from each activity group (Figs. 2 and 4A and *SI Appendix, Table S3*). From the dataset of single phage interactions, this cocktail was expected to be effective on 64% (201/314) of strains in the training dataset as at least one phage in the cocktail has activity against the strain. This was validated on a subset of 50 strains from the UPEC collection, selected to represent the diversity across the collection (orange circles on Fig. 1A). As expected, this general cocktail was effective on 64% of the strains tested (Fig. 4A and *SI Appendix, Table S4*). With most of the strains tested, the general cocktail limited the strains to a similar level to the most active single phage in the cocktail on that strain. There was little evidence of synergistic effects from the general cocktail where the score for the cocktail was less than any of the individual phage. In fact, the opposite could be seen where the cocktail score was higher than the best individual phage indicating some antagonist activity.

Bespoke cocktails could also be designed using the phage interaction dataset by choosing a selection of diverse phage active against a strain, a feature that would help combat the development of bacterial phage resistance. Only a small proportion of strains from the test set of 50 were susceptible to more than two phage

in the general cocktail (5 strains, *SI Appendix, Table S4*). A bespoke cocktail for each bacterial strain ensures that all phage included in the cocktail will have some activity against the strain. A few strains from the set of 50 could not have a bespoke cocktail of 5 phage designed due to a limitation of overall phage activity, but assessment was possible for 38 strains. The bespoke cocktails were significantly more effective when compared to the general cocktail with a lower median score (Fig. 4B, $P = 2.081 \times 10^{-5}$). As expected, analysis of the growth curves from these interactions generally showed much less bacterial regrowth in the bespoke cocktails compared to the general cocktail (example shown in Fig. 4C) resulting in lower scores.

A key question is whether the predictive ML models have the potential to inform phage selection of bespoke cocktails and how selection using this method compares to the general cocktail. To assess this and to look at phage activity on strains outside the original 314 in the training dataset, 10 newly isolated clinical UPEC strains (3 canine and 7 human) were genome sequenced, and these sequences run through the ML models to give predicted phage interaction scores (Fig. 4D and *SI Appendix, Table S5*). Subsequently observed data was generated for these unseen strains to compare with the predicted scores. For the majority of interactions where high phage activity is seen in the observed data (dark blue on Fig. 4D) there is corresponding predicted phage activity. From the predicted and observed scores, bespoke cocktails were independently constructed using specific rules (*Materials and Methods*) so they could be compared to each other. Phage cocktail composition for each strain can be found in *SI Appendix* with selection taking account of F1 scores (*SI Appendix, Table S5*). The full set of cocktails (general, bespoke from observed, and bespoke

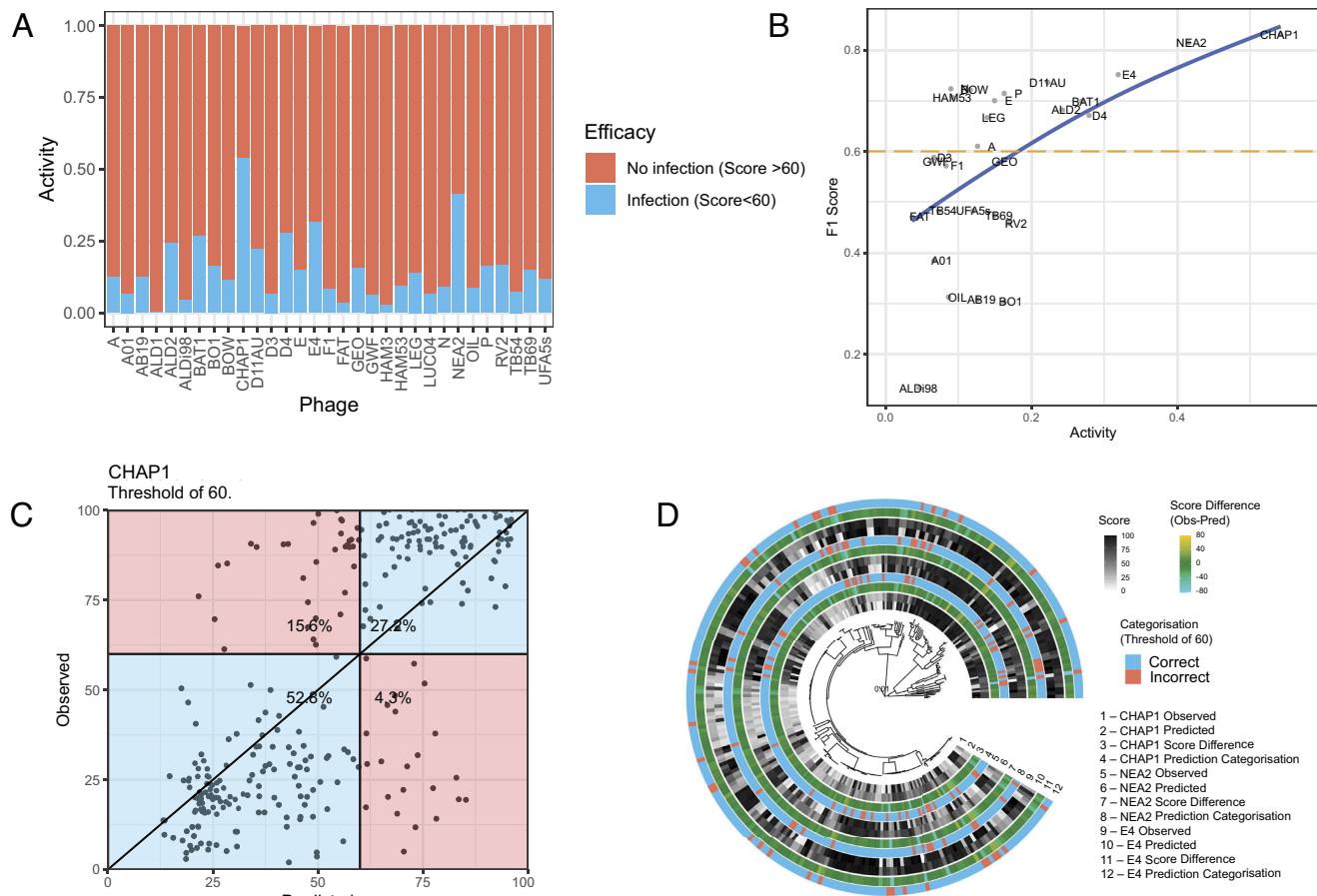


Fig. 3. ML predictions of phage activity. (A) The infectivity of the 31 phage across the *E. coli* dataset. Infections were determined based on an interaction score of less than 60. (B) F1 scores vs. activity (ratio of infected isolates) for PV phage models. The dashed line delineates a threshold of 0.6 for the F1 score as a metric for a “reliable” model. The F1 score was derived as described in the *Materials and Methods*. (C) Scatter plot of the predicted vs. observed interaction scores for phage CHAP1. The diagonal line represents a perfect prediction (observed = predicted). The chart is divided into four color-coded quadrants to represent correct/incorrect predictions (True Negatives as the *Top Right* blue quadrant, True Positives as the *Bottom Left* blue quadrant, False Negatives as the *Bottom Right* red quadrant, and False Positives as the *Top Left* red quadrant). (D) Phylogenetic SNP tree of the isolates used for model training. Exemplar phage models for CHAP1, NEA2, and E4 are shown. For each model, the observed and predicted interaction scores are shown in black and white, followed by the difference between the two scores as a color gradient, followed by whether each isolate would be categorized (score <60 the cutoff value for activity) in the same way by both the observed and predicted scores.

from predicted) were tested against the strains in AU, and as a final validation, in canine urine (CU). The scores from each of these cocktails for duplicate experiments are recorded in *SI Appendix, Table S6*, and the distribution of scores is shown in Fig. 4E. Three of the unseen strains were broadly insensitive to all the phage tested (HU98, HU99, and DTU09) and none of the cocktails worked well on these three strains. This was predicted for strains HU98 and HU99, but the prediction for DTU09 was generally more positive than the observed scores (Fig. 4D and *SI Appendix, Table S5*). The bespoke predicted cocktail performed at a level equivalent to the observed bespoke cocktail and was significantly more effective than the general cocktail in both AU and CU (Fig. 4E).

Discussion

Phage therapy over the last century has more often been successful when a bespoke approach has been used, i.e., where phage are identified that act on the infecting bacteria. While there are some exceptions to this, including “jumbo” phage targeting a wide range of *Salmonella* serovars for a recent advance in food safety (32), the

high-profile news stories and the long-standing work at the Eliava Institute are based on identifying the infecting organism and selecting appropriate phage from extensive phage collections. This highlights the potential of phage therapy but a bespoke approach is much harder to carry out at scale. The selection of appropriate antibiotics is strengthened with knowledge of the antibiotic resistances of the infecting strain, and similarly a diagnostic platform that provides information about phage effectiveness would greatly advance the development of phage treatment. The premise for this study was that genomic information about the infecting strain should be the basis of phage cocktail selection. Toward this, phage activity ML models were generated from bacterial sequence data combined with phage interaction data allowing prediction of phage that are active on “unseen” strains, i.e., those not used in the model training sets. Phage predicted as active from the ML models can then be combined into cocktails for treatment.

E. coli UTIs are an appropriate challenge for phage selection as horizontal gene transfer mediated by phage and plasmids has resulted in an ever-increasing pangenome of >100,000 genes relative to a “soft core” genome of only a few thousand genes. Prophage integration also introduces resistance mechanisms to

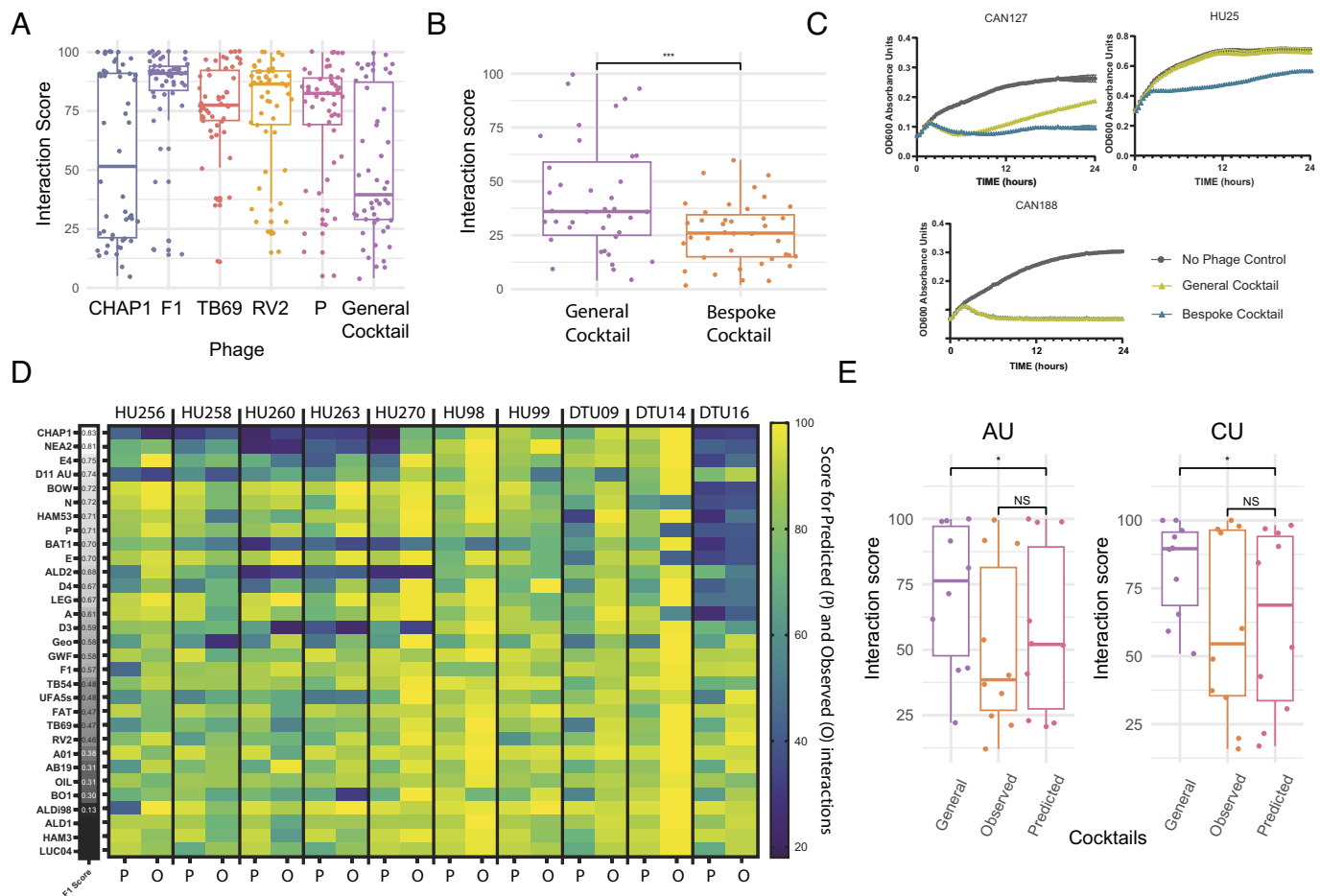


Fig. 4. Phage cocktail analysis. (A) A representative set of 50 UPEC strains from the training set were used to compare a general cocktail and its constituent phage. The box plots show the distribution of interaction scores for the individual phage and for the general cocktail. (B) For the 38 strains (out of the set of 50) that had more than 2 effective phage (from different activity groups) a bespoke cocktail was created and interaction scores recorded (SI Appendix, Tables S3 and S4). A statistical analysis (paired t test) comparing the general to the bespoke cocktail for these 38 strains shows that the bespoke cocktail is more effective ($P = 2.081 \times 10^{-5}$). (C) Representative growth data that were used to generate the interaction scores comparing the general and bespoke cocktails for three strains where there is no difference between the cocktails (CAN188), where there is bacterial resistance evolving for the general cocktail (CAN127) and where the bespoke cocktail is effective over the general cocktail (HU25). (D) A set of 10 strains that had not been in the training set were used to validate the prediction models. The heatmap shows both the predicted (P) and observed (O) interaction scores for these 10 strains against the set of 31 phage (high score (yellow) shows little phage interaction, low score (dark blue) represents strong phage interaction). F1 scores are also plotted on the heatmap (from high to low score) to indicate accuracy of each model for predictive capacity. The numerical scores for these interactions are presented in SI Appendix, Table S5. (E) Using defined rules bespoke cocktails were created for each of the test strains using the predicted interaction scores and separately the observed interaction scores and compared alongside the general cocktail. The specific phages included in each cocktail are shown in SI Appendix Table S5. This was done in AU and canine urine (CU), and the interaction scores were plotted (representative data from one experiment; repeat data available in SI Appendix, Table S6). Analysis of these results (paired t test) showed that the predicted cocktail was more effective than the general cocktail [$P = 0.03502$ (AU), $P = 0.03614$ (CU)]. There was not a statistical difference in effectiveness between the observed or predicted cocktail [$P = 0.1746$ (AU), $P = 0.4288$ (CU)].

phage infection and this accounts for major variation in phage susceptibility along with receptor and metabolic diversity. Due to these genomic determinants, phage activity on a strain should be identifiable from its genome sequence. To date, most of the predictive effort has been focused on possible hosts for phage based on phage genome sequencing, reviewed in ref. 24, in particular, prediction of bacteriophage sequences from metagenomic studies and predictive calling of possible bacterial hosts. These have mainly used nucleotide information including k-mer biases, although one study has taken into account CRISPR sequences (33). These studies are identifying phage interactions at the level of phylum or genus and have not defined phage that would be active on clinically infecting strains of a species based on their sequence data. The idea of predictive ML models based on the different steps in infection has been proposed (34) and our approach presented here, while relatively agnostic in that it uses differential PVs (predicted genes) as features, may well include receptors and resistance mechanisms. However, this is hard to dissect as the majority of features used are

grouped genes with no clear annotation. Future pipelines for phage selection could apply both associative features as well as those based on functional knowledge. Either way, high-quality training sets with phenotypic interaction data will still be critical.

For the majority of methods using genome features for prediction (reviewed for source attribution in ref. 35) there will be significant phylogenetic influence. This will mean that features are used with statistically significant associations with phage activity but which do not have a functional relationship to phage susceptibility. This is not an issue for the ML models if they are used within a similar population structure to the training dataset, but there would be reduced accuracy if the ML models are applied to strains from different structures (36) ML models are also prone to replicating existing biases in their training datasets. This was demonstrated in the prediction scores of phage with very imbalanced datasets (e.g., A01), in which the ML model was trained on mainly high interaction scores indicating no infection and a very small subset of low scores, indicating infection. The resulting

model struggled to differentiate isolates that the phage would actually infect and would be prone to generating False Negatives. From a cocktail perspective, the failure to select a phage that would have been active (False Negative) is actually preferred to False Positives, although a high rate of False Negatives may lead to a paucity of phage predicted as active from which to select a cocktail. We note that with the better ML models presented in this study, the frequency of False Positives and False Negatives falls within a range where if five phage are selected for a cocktail the majority will be correctly predicted. For example, for Chap1, the predictive model generated 7% False Negative scores (22/314), and 10.8% were False Positive scores (34/314).

Our analysis on a set of completely independent sequenced isolates did demonstrate that the predicted cocktails were more effective than a generic cocktail and in fact were comparable with the bespoke cocktails that were based on individual measured activity. This result on unseen isolates provides confidence in the main concept of the study. To really move such predictive phage therapy into the clinic, there are areas the research has highlighted that require further development along with the safety and registration issues that are being tackled more globally by the phage community. A key improvement would be inclusion of more generalist phage in the interaction dataset. Based on the findings illustrated in Fig. 3, the ML models trained on datasets with a class imbalance as high as 85:15 produced practical results for phage selection, as defined by a predefined F1 cutoff of 0.6. For greater reliability, it would be preferable to use phage with a class imbalance less than this and therefore phage that can target at least 30% of the population. More generalist phage will also provide more options for cocktail selection especially if they can be distributed across different activity groups. The work shown is a “proof of concept” and valuable interaction datasets for prediction ML models could now be produced in a short period of time using high throughput platforms, including those that can measure 50- × 96-well plates with appropriate phage and strain collections in place.

Another major consideration is the actual activity of the selected phage in the infection environment, in this case urine which can vary considerably in terms of composition and constituent concentrations, especially pH and osmolarity, as well as factors associated with the host response to infection (37–39). We have tried to mitigate this by measuring phage activity in AU but acknowledge this cannot capture the complexity of real urine. AU is a defined reproducible medium and a closer match to urine than many rich broths that are routinely used for phage activity studies that allow faster bacterial growth. The bespoke phage cocktails formulated in this study based on the predictive ML models showed good activity in CU, comparable with results in AU (Fig. 4D), and this mirrored our preliminary work on the activity of a subset of individual phage in AU vs. pooled CU (SI Appendix, Fig. S1).

Predictive phage therapy would be used in conjunction with diagnostic sequencing. Direct sequencing of nucleic acids extracted from clinical sample has been successfully used for diagnosis for viruses, fungi, and bacteria (40–42) including UTI (43, 44) and we have applied it to direct sequencing of *E. coli* in CU samples with respect to the selection of appropriate antimicrobials (28). There is the challenge of feature extraction for the phage ML models from long read data of more complex samples compared to assemblies of short reads from purified strains, as used in this study. If there are multiple strains associated with the infection, then the ML models may still be useful to select phage, even without their separation at a genome level, although that remains to be tested. The analysis of phage activity with sequenced clinical samples would be used to continually reinforce the learning dataset and improve predictions.

It should be emphasized that the primary aim of this study was to signpost what could be possible for phage therapy with ML or alternative predictive methods based on comprehensive interaction datasets. The study looked at both canine and human UTI *E. coli* isolates presented by patients at local hospitals and used a selection of 31 phage from an initial pool of just over 100. While these are only representative subpopulations, we consider they allowed the concept of predictive phage therapy to be tested with a promiscuous and important pathogen that can exhibit multidrug resistance. We are optimistic that there will be a rapid expansion of advances in this space and the best methods could be combined in pipelines to transition the best approaches into clinical practice. In addition, the wider adoption of phage therapy in the veterinary or human clinic will require changes in licensing and management of expectations of both clinicians and the general public (45).

Materials and Methods

Strain Information. A total of 203 *E. coli* were isolated from UTIs in canine patients at the Royal (Dick) School of Veterinary Studies, Hospital for Small Animals, between Jul. 2017 and Jun. 2019. In addition, 81 *E. coli* were isolated from urine and 30 *E. coli* from blood from patients with bacteremia processed at The Royal Infirmary of Edinburgh (collected between Jan. 2019 and Nov. 2019). These patient isolates were deidentified and assigned a study number prior to collection from the hospital. The canine, human urine, and human blood isolates were designated “CANxx,” “HUxx,” and “HBxx,” respectively. Isolates were stored at –70 °C in 20% glycerol (G5516, Sigma, Germany) and streaked out to obtain pure, individual colonies on LB agar. A final set of 10 strains were collected from the same clinics toward the end of the study for validation of the prediction ML models. (canine, *n* = 3, collected as part of a different study (28) and human, *n* = 7, from The Royal Infirmary of Edinburgh collected Spring 2023).

Phage Isolation and Propagation. Phage used in the study were either isolated from wastewater kindly supplied by the Scottish Environment Protection Agency or provided by The Phage Technology Centre, GmbH, Bönen, Nordrhein-Westfalen, Germany. Phage were propagated by growing the appropriate host strain to an OD₆₀₀ 0.2 to 0.3 in LB or AU in a glass conical flask at 37 °C with shaking at 170 rpm before adding 500 µL of phage plaque filtrate in SM buffer [50 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM MgSO₄, and 0.01% gelatin (v/w)] and continuing to incubate overnight. Infected cultures were centrifuged at 5,000 rpm for 10 min in a benchtop centrifuge to pellet the bacteria, and the phage lysate passed through a 0.22-µm syringe filter before long-term storage at 4 °C. Phage were titrated using double-layer agar overlay plates and diluted to 1 × 10⁶ per mL working stock concentration using SM buffer.

Growth Media and Culture Conditions. One liter volumes of AU were prepared (according to the following protocol, <https://doi.org/10.17504/protocols.io.kxygx3exzg8j/v1>), aliquoted into single-use tubes, and frozen until required. Each aliquot was filter sterilized through a 0.22-µm syringe filter prior to use. CU was collected from healthy dogs, filter sterilized using a 0.22-µm syringe filter, aliquoted, and frozen until use.

Interaction Assays Including AUC Analysis. Phage interaction assays were performed using 31 phage against 314 *E. coli* isolates using a multiplicity of infection (MOI) of 0.01 (~1 phage per 100 bacteria). This MOI ratio was chosen to ensure that complete virus lifecycle infections were analyzed rather than abortive infections which can occur at higher MOIs. A single colony of an *E. coli* isolate was picked to inoculate 5 mL Luria Broth (LB) (LB0102, Formedium Ltd, England) and incubated overnight at 37 °C with shaking at 170 rpm. Fifty microliters of overnight culture was subcultured into 5 mL fresh LB and grown at 37 °C and 170 rpm shaking to an optical density (OD₆₀₀) of 1.0 (equivalent to 10⁸ bacterial cells per mL). During this incubation, 10 µL of SM buffer (for no-phage controls) or 10 µL phage at 1 × 10⁶ per mL (phage stocks were diluted in SM buffer) was added to a flat-bottomed 96-well microplate. One hundred eighty microliters of AU was added to all wells. Ten microliters of *E. coli* at OD₆₀₀ of 1 was added to all wells resulting in a final reaction volume of 200 µL. A gas-permeable, sterile, optically clear plate seal was applied to the plate (4ti-0516/96, Azenta Life Sciences) to control for evaporation

and prevent phage contamination within the plate reader. Microplates were run on 1 of 3 identical Multiskan FC photometers (51119100, Thermo, China) with absorbance measurements at 620 nm taken every 20 min immediately after a 5-s mix over a time course of 18 h and constant incubation at 37 °C.

All interactions and controls were measured with at least three technical repeats and additional controls on each plate for growth and sterility. Media controls comprised 180 µL AU, 10 µL SM buffer, and 10 µL LB in place of bacterial culture. While the number of interactions involved was too large to carry out each interaction as triplicate biological repeats, biological repeats of a limited number of strains were carried out periodically to ensure repeatability of the assay over time and consistency of phage activity. Based on 31 phage measured on 314 bacterial strains with the technical repeats, a total of >27,000 growth curves were produced and analyzed. The Multiskan FC photometers use SkanIt Software (Thermo) which allowed raw data to be exported as a Microsoft Excel file. Phage interaction scores were calculated by comparing the average AUC of the plus phage triplicate wells against the average AUC for the no-phage control wells using PRISM software (GraphPad). [Score = (AUC plus phage/AUC no-phage control) × 100]. For the purposes of plotting heatmap data where the score generated was over 100, it was recorded as a value of 100 representing no phage activity.

Genome Sequencing. Genome sequencing of isolates was provided by MicrobesNG (<http://www.microbesng.com>) using paired Illumina reads. The reads were uploaded to Enterobase and assembled using the Enterobase Tool Kit (<https://github.com/zheminzhou/EToKi>). The assemblies were then annotated with prokka version 1.14.5 (46). Enterobase's Tool Kit was also used to generate a maximum likelihood phylogenetic tree based on core SNP.

ML and Prediction Pipelines. After filtering for fragmented assemblies (<600 contigs), 301 of the 314 isolates were used for ML model training and testing. The main features used for testing were predicted genes, defined as predicted PVs generated from a pangenome analysis of the 301 bacterial sequences using panaroo version 1.2.9 (47). PVs were encoded in a binary format (1/0) to indicate presence/absence in an isolate. This approach was based on previous work that used a similar pipeline for source attribution of *Salmonella Typhimurium* (36).

Random forest regression models were created for each of the 31 phage using the MUVr package (48) using one 75–25 split of training/testing data and 10-fold cross-validation. The predicted value was the interaction score calculated during the interaction assay analysis.

ML model performance for score prediction was evaluated using the rmse statistic, which is calculated by taking the square root of the average of the squared differences between the predicted and actual values. RMSE provides a single numerical value that represents the overall accuracy of a prediction model, with lower values indicating better predictive performance.

Feature reduction was performed recursively using the MUVr package. For each phage, an initial model using all PVs was created. Then, the top-ranking features of a model were selected to create a new model with a reduced number of features, and the RMSE values between the old and the new model were compared. If the ML models had a similar RMSE, the top-ranking features of the new model were used to create another new model, and the process was repeated.

F1, recall, precision, and related statistics were created by classifying each observed and predicted phage interaction based on the score (score < 60: infection; score > 60 no infection). A threshold of 60 was selected as it represents an activity score below which we have high confidence of a positive interaction and this mainly relates to the variation that exists between our experimental repeats. As is standard, those measurements were calculated based on the number of True Positives, True Negatives, False Positives, and False Negative predictions for each ML model. For our binary Infection/No Infection categories, predictions were classified as "True Positives" when the observed and predicted interaction scores for an isolate both indicated infection (i.e., scores < 60). For "True Negatives", both observed and predicted interaction scores indicated no interaction (i.e., scores > 60). "False Positives" occurred when the predicted score indicated infection but

the observed score indicated no infection (i.e., predicted score < 60, observed score > 60). Conversely, "False Negatives" occurred when the predicted score indicated no infection but the observed score indicated infection (i.e., predicted score > 60, observed score < 60).

Precision was calculated by dividing the True Positives with the sum of predicted scores that indicated predictions: Precision = True Positives/(True Positives + False Positives). Recall was calculated by dividing the True Positives with the sum of positive observed scores for infection: Recall = True Positives/(True Positives + False Negatives). F1 is the harmonic mean of precision and recall and was calculated using the formula: $F1 = 2 * (Precision * Recall) / (Precision + Recall)$.

Overall model performance was judged on two metrics: RMSE and F1. RMSE was used to evaluate model performance based on quantitative predictions (activity score), whereas F1 was used to evaluate model performance on a binary categorical scale (Infection/No Infection). These metrics were primarily used to compare the performance of our own phage ML models as well as give an overview of each ML model's accuracy. We are not aware of other studies that allow us to benchmark the performance of our phage ML models more widely.

Phage Cocktail Testing. Cocktail interaction assays were set up in a similar manner to the single phage assays described above. Phage were selected for the bespoke cocktails (both predicted and observed) using the same design rules: i) to include the lowest scoring (highest activity) phage from each interaction grouping (Fig. 2) as long as it scored under 80; ii) if no active phage available from a group, then additional phage(s) were added that were the second lowest scoring phage from activity groupings starting with the remaining phage with the lowest score until 5 phage in the cocktail; and iii) if less than 5 phage available with scores less than 80, then less phage were used in the cocktail. Working stocks of phage cocktails were prepared by mixing equal volumes of each selected phage already prepared at 10⁶ per mL such that the 10 µL volume transferred to each triplicate well of a 96-well plate contained an equivalent total number of virus particles to the single phage assays, so that the overall MOI remained at 0.01. Additionally, the cocktail assays were allowed to proceed for 24 h.

Genomic Data. The genomic data generated for this work have been deposited in Enterobase (https://enterobase.warwick.ac.uk/species/ecoli/search_strains?query=workspace:97040).

Data, Materials, and Software Availability. Genomic data have been deposited in Enterobase (https://enterobase.warwick.ac.uk/species/ecoli/search_strains?query=workspace:97040) (49). All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We are very grateful to Jennifer Harris in the veterinary microbiology laboratory at R(D)SVS for assistance in obtaining canine UPEC strains and Kate Templeton, Martin McHugh, and Azul Zorzoli at the Dept. of Medical Microbiology at the Royal Infirmary, Edinburgh, for sorting approvals and supplying human urine and blood *E. coli* isolates. We would like to acknowledge Joana Alves for assistance with statistical analysis. The work was supported by a Canine Welfare Grant from The Dog's Trust awarded to D.L.G. ("Advanced phage therapy for MDR *E. coli* associated with canine UTIs"), a PhD studentship from the National Council of Science and Technology of Mexico (CONACYT) alongside a joint studentship from the University of Edinburgh and the University of Glasgow awarded to A.P.d.I.T. and a Roslin Institute PhD studentship supported by the BBSRC awarded to A.C. The collection of hospital isolates was approved by Lothian NRS Bioresource SR1094. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author affiliations: ^aThe Roslin Institute, Division of Bacteriology, University of Edinburgh, Edinburgh EH25 9RG, United Kingdom; and ^bRoyal (Dick) School of Veterinary Studies, Easter Bush Pathology, University of Edinburgh, Edinburgh EH25 9RG, United Kingdom

1. I. U. Mysorekar, S. J. Hultgren, Mechanisms of uropathogenic *Escherichia coli* persistence and eradication from the urinary tract. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14170–14175 (2006).
2. N. Nittayasut et al., Multiple and high-risk clones of extended-spectrum cephalosporin-resistant and blaNDM-5-Harboring uropathogenic *Escherichia coli* from cats and dogs in Thailand. *Antibiotics* **10**, 1374 (2021).

3. A. L. Flores-Mireles, J. N. Walker, M. Caparon, S. J. Hultgren, Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
4. R. Kaur, R. Kaur, Symptoms, risk factors, diagnosis and treatment of urinary tract infections. *Postgrad. Med. J.* **97**, 803–812 (2021).
5. J. K. Byron, Urinary tract infection. *Vet. Clin. North Am. Small Anim. Pract.* **49**, 211–221 (2019).

6. A. Sabih, S. W. Leslie, *Complicated Urinary Tract Infections* (StatPearls Publishing, 2023) (15 July 2023).
7. K. M. Hatfield *et al.*, Assessing variability in hospital-level mortality among U.S. medicare beneficiaries with hospitalizations for severe sepsis and septic shock. *Crit. Care Med.* **46**, 1753–1760 (2018).
8. F. Wagenlehner *et al.*, The global prevalence of infections in urology study: A long-term, worldwide surveillance study on urological infections. *Pathogens* **5**, 10 (2016).
9. B. Kot, Antibiotic resistance among uropathogenic *Escherichia coli*. *Pol. J. Microbiol.* **68**, 403–415 (2019).
10. T. m. Sørensen *et al.*, Effects of diagnostic work-up on medical decision-making for canine urinary tract infection: An observational study in Danish small animal practices. *J. Vet. Intern. Med.* **32**, 743–751 (2018).
11. A. L. Zogg, K. Zurfli, S. Schmitt, M. Nüesch-Inderbinen, R. Stephan, Antimicrobial resistance, multilocus sequence types and virulence profiles of ESBL producing and non-ESBL producing uropathogenic *Escherichia coli* isolated from cats and dogs in Switzerland. *Vet. Microbiol.* **216**, 79–84 (2018).
12. S. McCallin, T. M. Kessler, L. Leitner, Management of uncomplicated urinary tract infection in the post-antibiotic era: select non-antibiotic approaches. *Clin. Microbiol. Infect.* **29**, 1267–1271 (2023).
13. F. M. d'Herelle, Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad. Sci. Paris*, **165**, 373–375 (1917).
14. F. d'Herelle, Bacteriophage as a treatment in acute medical and surgical infections. *Bull. N. Y. Acad. Med.* **7**, 329–348 (1931).
15. F. W. Twort, Further investigations on the nature of ultra-microscopic viruses and their cultivation. *J. Hyg. (Lond.)* **36**, 204–235 (1936).
16. A. Bernheim, R. Sorek, The pan-immune system of bacteria: Antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
17. W. P. J. Smith, B. R. Wucher, C. D. Nadell, K. R. Foster, Bacterial defences: Mechanisms, evolution and antimicrobial resistance. *Nat. Rev. Microbiol.* **21**, 519–534 (2023).
18. J. S. Athukoralage, M. F. White, Cyclic nucleotide signaling in phage defense and counter-defense. *Annu. Rev. Virol.* **9**, 451–468 (2022).
19. H. G. Hampton, B. N. J. Watson, P. C. Fineran, The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
20. B. K. Chan, S. T. Abedon, C. Loc-Carrillo, Phage cocktails and the future of phage therapy. *Fut. Microbiol.* **8**, 769–783 (2013).
21. F. L. Gordillo Altamirano, J. J. Barr, Unlocking the next generation of phage therapy: The key is in the receptors. *Curr. Opin. Biotechnol.* **68**, 115–123 (2021).
22. C. Lood, P.-J. Haas, V. van Noort, R. Lavigne, Shopping for phages? Unpacking design rules for therapeutic phage cocktails. *Curr. Opin. Virol.* **52**, 236–243 (2022).
23. M. Merabishvili, J.-P. Pirnay, D. De Vos, "Guidelines to compose an ideal bacteriophage cocktail" in *Bacteriophage Therapy: From Lab to Clinical Practice*, J. Azeredo, S. Sillankorva, Eds. (*Methods in Molecular Biology*, Springer, 2018), pp. 99–110.
24. Y. Nami, N. Imeni, B. Panahi, Application of machine learning in bacteriophage research. *BMC Microbiol.* **21**, 193 (2021).
25. R. Aguas, N. M. Ferguson, Feature selection methods for identifying genetic determinants of host species in RNA viruses. *PLoS Comput. Biol.* **9**, e1003254 (2013).
26. Q. Tang *et al.*, Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Sci. Rep.* **5**, 17155 (2015).
27. M. Wardeh, M. S. C. Blagrove, K. J. Sharkey, M. Baylis, Divide-and-conquer: Machine-learning integrates mammalian and viral traits with network features to predict virus-mammal associations. *Nat. Commun.* **12**, 3954 (2021).
28. N. Ring *et al.*, Rapid metagenomic sequencing for diagnosis and antimicrobial sensitivity prediction of canine bacterial infections. *Microb. Genom.* **9**, mgen001066 (2023).
29. O. Clermont, J. K. Christenson, E. Denamur, D. M. Gordon, The Clermont *Escherichia coli* phylotyping method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **5**, 58–65 (2013).
30. J. R. Johnson, T. A. Russo, Extraintestinal pathogenic *Escherichia coli*: "The other bad E coli". *J. Lab. Clin. Med.* **139**, 155–162 (2002).
31. L. W. Riley, Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* **20**, 380–390 (2014).
32. A. M. Thanki *et al.*, A bacteriophage cocktail delivered in feed significantly reduced *Salmonella* colonization in challenged broiler chickens. *Emerg. Microbes Infect.* **12**, 2217947 (2023).
33. W. Wang *et al.*, A network-based integrated framework for predicting virus-prokaryote interactions. *NAR Genom. Bioinform.* **2**, lqaa044 (2020).
34. C. Lood *et al.*, Digital phagograms: Predicting phage infectivity through a multilayer machine learning approach. *Curr. Opin. Virol.* **52**, 174–181 (2022).
35. N. Lupolova, S. J. Lycett, D. L. Gally, A guide to machine learning for bacterial host attribution using genome sequence data. *Microb. Genom.* **5**, e000317 (2019).
36. A. Chalka, T. J. Dallman, P. Vohra, M. P. Stevens, D. L. Gally, The advantage of intergenic regions as genomic features for machine-learning-based host attribution of *Salmonella* Typhimurium from the USA. *Microb. Genom.* **9**, 001116 (2023).
37. C. Robinson-Cohen *et al.*, Estimation of 24-hour urine phosphate excretion from spot urine collection: Development of a predictive equation. *J. Ren. Nutr. Off. J. Coun. Ren. Nutr. Natl. Kidney Found.* **24**, 194–199 (2014).
38. D. B. Barr *et al.*, Urinary creatinine concentrations in the U.S. population: Implications for urinary biologic monitoring measurements. *Environ. Health Perspect.* **113**, 192–200 (2005).
39. E. N. Taylor, G. C. Curhan, Differences in 24-hour urine composition between black and white women. *J. Am. Soc. Nephrol.* **18**, 654 (2007).
40. K. Lewandowski *et al.*, Metagenomic nanopore sequencing of influenza virus direct from clinical respiratory samples. *J. Clin. Microbiol.* **58**, e00963–19 (2019).
41. A. Ohta, K. Nishi, K. Hirota, Y. Matsuo, Using nanopore sequencing to identify fungi from clinical samples with high phylogenetic resolution. *Sci. Rep.* **13**, 9785 (2023).
42. K. Nilgiriwala *et al.*, Genomic sequencing from sputum for tuberculosis disease diagnosis, lineage determination, and drug susceptibility prediction. *J. Clin. Microbiol.* **61**, e0157822 (2023).
43. K. Schmidt *et al.*, Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* **72**, 104–114 (2017).
44. L. Zhang *et al.*, Rapid detection of bacterial pathogens and antimicrobial resistance genes in clinical urine samples with urinary tract infection by metagenomic nanopore sequencing. *Front. Microbiol.* **13**, 858777 (2022).
45. J. D. Jones, H. J. Stacey, A. Bailey, M. Suleman, R. J. Langley, Managing patient and clinician expectations of phage therapy in the United Kingdom. *Antibiot. Basel Switz.* **12**, 502 (2023).
46. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinform. Oxf. Engl.* **30**, 2068–2069 (2014).
47. G. Tonkin-Hill *et al.*, Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
48. L. Shi, J. A. Westerhuis, J. Rosén, R. Landberg, C. Brunius, Variable selection and validation in multivariate modelling. *Bioinformatics* **35**, 972–980 (2019).
49. A. Chalka, A. Vallejo-Trujillo, A. Low, Workspace - *ecoli_train*. Enterobase. https://enterobase.warwick.ac.uk/species/ecoli/search_strains?query=workspace:97040. Deposited 1 June 2022.