Supplementary Materials for

Temporal encoding of bacterial identity and traits in growth dynamics

Carolyn Zhang[1], Wenchen Song[2, 3], Helena R. Ma[1], Xiao Peng[1], Deverick J. Anderson[4], Vance G. Fowler Jr.[5], Joshua T. Thaden[5], Minfeng Xiao[2, 3], Lingchong You[1, 6, 7*]

[1]Department of Biomedical Engineering, Duke University, Durham, North Carolina, 27708, USA

[2]BGI-Shenzhen, Shenzhen, 518083, China

[3]Shenzhen Key Laboratory of Unknown Pathogen Identification, Shenzhen, 518083, China

[4]Duke Center for Antimicrobial Stewardship and Infection Prevention, Duke University School of Medicine, Durham, North Carolina, 27708, USA

[5]Division of Infectious Diseases and International Health, Department of Medicine, Duke University School of Medicine, Durham, NC, 27710, USA

[6]Center for Genomic and Computational Biology, Duke University, Durham, North Carolina, 27708, USA

[7]Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, North Carolina, 27708, USA

*Corresponding author. Department of Biomedical Engineering, Duke University, CIEMAS 2355, 101 Science Drive, Box 3382, Durham, North Carolina, 27708, USA

Tel.: +1 (919)660-8408; Fax: +1 (919)668-0795; E-mail: lingchong.you@duke.edu

Table of Contents

1. Development of phenotype-based predictions

We examined the extent to which growth dynamics store information relevant for strain identification and other characteristics. For the subsequent analysis, we used Support Vector Machine (SVM), a supervised learning algorithm.

1.1 Strain prediction – clinical isolates: Supplementary Tables 1.1-1.4, Supplementary Figure 1.1

Here, we described the accuracy of strain identification of the clinical isolate library (203 strains) using one of two methods: (1) 3-fold cross validation with holdout and (2) 4-fold cross validation. For the 3-fold cross validation procedure, we reported the accuracy of a validation set. For the 4-fold cross validation procedure, we reported the average test set accuracy across the four folds. The latter, which is reported in the main text, prevents skewing of the accuracy by certain replicates by averaging results using all replicates as the test set **(Supplementary Tables 1.1-1.4)**. The first is to demonstrate the generality of our models as the reported validation accuracy is on data the model has not previously seen. Due to the similarity between the results, for the main text and subsequent analysis we utilized the 4-fold cross validation procedure due to its consistency.

For both methods, we reported the optimal parameter set using multiclass Support Vector Machines. The top k strains predicted are checked against the true label (k = 1 or k = 5). If multiple parameter sets resulted in the highest accuracy, only one is shown. Parameters were selected from the following options: kernel function (linear, quadratic, rbf), kkt violation level (0, 0.05, 0.1), box constraint of the soft margin C (10, 100, 1000), and rbf scaling factor σ (1, 10, 20). Here, the label for each strain was decided according to phylogenetic analysis (described in **Methods** and **Supplementary Section 3.2**). All isolates with a pairwise distance of 0 were considered as the same strain while those with a distance greater than 0 were labeled to be unique strains. The phylogenetic tree separated the 244 isolates into 203 strains. To choose 4 replicates as the final dataset, we randomly chose 4 replicates out of all replicates (from multiple isolates) if more than a single isolate made up a strain. The randomness inherent to this process can lead to some variability in the reported accuracy.

The raw experimental data consisted of OD_{600} measurements, a measure of cell density, as a function of time (99 time points in increments of 10 minutes). In this section, we compared the use of the growth curves (cell density as a function of time), time derivative of cell density (weighted growth rate as a function of time), growth rate, and other metrics as the covariates to train the SVM model **(Supplementary Tables 1.1-1.4)**. The other metrics are area under the growth curve (AUC) and maximum time derivative of cell density (μ) as the predictor in **Supplementary Table 1.4 and Figure 1.1**. Here, we calculated the AUC of the growth curves after smoothing with a median filter (window = 3) using trapezoidal numerical integration. Additionally, μ was defined as the maximum value of the time derivative of the growth curve, after the same smoothing protocol. By using either of these metrics or both in combination, we showed that the time derivative of cell density curves **(Supplementary Table 1.1)** were a better predictor of strain identity for the clinical isolates than the other data processing approaches.

MATLAB files in package (https://github.com/youlab/strain_prediction_CZ): predictStrain_clinicalIsolates.m (main file), importClinicalIsolates.m, multi_class_svm_cv_topk.m, multi_class_svm_ci.m, clinSVMOpt.m, predictStrain_clinicalIsolates_changetime.m (**Supplementary Figure 1.1**), predictStrain_clinicalIsolates_traditionalGR.m (main file), importClinicalIsolates_traditionalGR.m

Supplementary Table 1.1

The features used are one of two metrics: (1) the time derivative of growth curves and (2) the growth rate. Additionally, we compare the use of phylogeny based on a core set of genes (MLST) and SNPs. When using either approach for strain definition, the conclusion remains the same. Here, we demonstrate two approaches to estimating the predictive capability of growth dynamics.

In the first, we use a 4-fold cross validation in which the average test accuracy of the 4 folds is reported. The top k (k = 1) predictions are used to predict each sample in the test set. The accuracy for all combinations of growth conditions is compared to the accuracy associated with random chance; the top three conditions are highlighted in bold.

In the second, we use 1 replicate as the held-out validation set and 3-fold cross validation is applied to the other 3 replicates to optimize the SVM hyper-parameters. A final model uses these optimized parameters to predict the accuracy of the validation set. The average test accuracy for the 3-fold cross validation is reported along with the accuracy of the validation set. The similarity in the hold out test accuracy with the average cross validation accuracy demonstrates the generalizability of the model.

SVM Parameters	**Average Test Accuracy 4-fold CV[1]**	**Hold Out Test Accuracy 3-fold CV[2]**	**Growth Condition**
Predictions below are based on time derivative of growth curves using a SNP approach to strain definition (203 unique strains).			
Random chance	**0.49%**	**0.49%**	**----------top 1----------**
Rbf kernel, C = 100, kkt = 0, $\boldsymbol{\sigma}$ = 10	91.50%	CV: 88.18% Test: 92.12%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	92.12%	CV: 90.64% Test: 83.74%	Phage treatment
Rbf kernel, C= 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	91.50%	CV: 88.43% Test: 83.25%	100x dilution
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	91.01%	CV: 87.68% Test: 80.30%	Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	95.94%	CV: 94.58% Test: 96.06%	10,000x dilution + Phage treatment
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	96.31%	CV: 95.07% Test: 98.03%	10,000x dilution + 100x dilution

Rbf kernel, C = 100, kkt = 0, σ = 20	96.31%	CV: 95.24% Test: 94.58%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	96.80%	CV: 95.57% Test: 95.57%	Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 10	96.18%	CV: 94.42% Test: 97.54%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 1000, kkt = 0.1, σ = 20	96.18%	CV: 93.76% Test: 95.57%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	**97.78%**	**CV: 95.89% Test: 97.54%**	**10,000x dilution + Phage treatment + 100x dilution**
Linear kernel, C = 10, kkt = 0	96.80%	CV: 96.22% Test: 96.06%	10,000x dilution + Phage treatment + Carbenicillin treatment
Quadratic kernel, C = 10, kkt = 0	**97.41%**	**CV: 95.73% Test: 96.55%**	**10,000x dilution + 100x dilution + Carbenicillin treatment**
Rbf kernel, C = 10, kkt = 0, σ = 20	96.31%	CV: 95.73% Test:95.57%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	**97.53%**	**CV: 96.39% Test: 97.00%**	**10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment**
Predictions below are based on growth rate time courses using a SNP approach to strain definition (203 unique strains).			
Random chance	**0.49%**	**0.49%**	**----------top 1----------**
Rbf kernel, C = 100, kkt = 0, σ = 10	83.50%	CV: 78.82% Test: 80.79%	10,000x dilution
Rbf kernel, C = 100, kkt = 0, σ = 10	84.48%	CV: 80.13% Test: 82.76%	Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	84.98%	CV: 81.78% Test: 80.79%	100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 10	60.47%	CV: 55.50% Test: 53.70%	Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	92.24%	CV: 87.03% Test: 92.61%	10,000x dilution + Phage treatment
Rbf kernel, C = 100, kkt = 0, σ = 10	93.47%	CV: 89.82% Test: 92.61%	10,000x dilution + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 10	86.58%	CV: 78.82% Test: 85.22%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	92.73%	CV: 90.64% Test: 91.13%	Phage treatment + 100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 20	86.45%	CV: 82.43% Test: 84.73%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 10	88.55%	CV: 82.43% Test: 87.68%	100x dilution + Carbenicillin treatment

Rbf kernel, C = 100, kkt = 0, σ = 20	**94.83%**	**CV: 93.10% Test: 94.09%**	**10,000x dilution + Phage treatment + 100x dilution**
Rbf kernel, C = 100, kkt = 0, σ = 20	92.98%	CV: 87.68% Test: 93.10%	10,000x dilution + Phage treatment + Carbenicillin treatment
Rbf kernel, C = 100, kkt =0 , σ = 20	**94.83%**	**CV: 92.12% Test: 92.12%**	**10,000x dilution + 100x dilution + Carbenicillin treatment**
Rbf kernel, C = 10, kkt = 0, σ = 10	93.35%	CV: 91.30% Test: 92.61%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	**95.32%**	**CV: 92.94% Test: 96.06%**	**10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment**
Predictions below are based on time derivative of growth curves using an MLST approach to strain definition (41 unique strains).			
Random chance	**2.4%**	**2.4%**	**----------top 1----------**
Rbf kernel, C = 10, kkt = 0, σ = 20	97.56%	CV: 96.75% Test: 95.12%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	97.56%	CV: 96.75% Test: 95.12%	Phage treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	96.34%	CV: 90.24% Test: 92.68%	100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	96.34%	CV: 96.75% Test: 90.24%	Carbenicillin treatment
Rbf kernel, C =10, kkt = 0, σ = 10	**99.39%**	**CV: 99.19% Test: 100%**	**10,000x dilution + Phage treatment**
Rbf kernel, C = 10, kkt = 0, σ = 10	98.78%	CV: 98.37% Test: 97.56%	10,000x dilution + 100x dilution
Quadratic kernel, C = 10, kkt = 0	98.17%	CV: 99.19% Test: 92.68%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	98.17%	CV: 98.37% Test: 95.12%	Phage treatment + 100x dilution
Quadratic kernel, C =10, kkt = 0	98.17%	CV: 99.19% Test: 92.68%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	96.95%	CV: 94.31% Test: 92.68%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0.05, σ = 20	**100%**	**CV: 100% Test: 100%**	**10,000x dilution + Phage treatment + 100x dilution**
Linear kernel, C = 10, kkt = 0	**98.78%**	**CV: 99.19% Test: 97.56%**	**10,000x dilution + Phage treatment + Carbenicillin treatment**
Rbf kernel, C = 10, kkt = 0.1, σ = 20	98.17%	CV: 99.19% Test: 92.68%	10,000x dilution + 100x dilution + Carbenicillin treatment

Rbf kernel, C = 10, kkt = 0, σ = 20	98.17%	CV: 99.19% Test: 95.12%	Phage treatment + 100x dilution + Carbenicillin treatment
Linear kernel, C = 10, kkt = 0	**98.78%**	**CV: 99.19% Test: 95.12%**	**10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment**
Predictions below are based on growth rate time courses using an MLST approach to strain definition (41 unique strains).			
Random chance	**2.4%**	**2.4%**	**----------top 1----------**
Rbf kernel, C = 100, kkt = 0, σ = 20	94.51%	CV: 95.94% Test: 90.24%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, σ = 10	88.41%	CV: 83.74% Test: 90.24%	Phage treatment
Rbf kernel, C = 100, kkt = 0, σ = 10	92.07%	CV: 88.62% Test: 87.80%	100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 10	70.12%	CV: 65.85% Test: 58.54%	Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	94.51%	CV: 94.31% Test: 97.56%	10,000x dilution + Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	**98.78%**	**CV: 96.75% Test: 100%**	**10,000x dilution + 100x dilution**
Rbf kernel, C = 10, kkt = 0, σ = 10	96.34%	CV: 95.94% Test: 97.56%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	93.90%	CV: 90.24% Test: 92.68%	Phage treatment + 100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 20	90.24%	CV: 90.24% Test: 82.93%	Phage treatment + Carbenicillin treatment
Rbf kernel, C =100 , kkt = 0, σ = 20	97.56%	CV: 90.24% Test: 97.56%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	**98.17%**	**CV: 95.35% Test: 100%**	**10,000x dilution + Phage treatment + 100x dilution**
Rbf kernel, C = 10, kkt = 0, σ = 20	95.73%	CV: 95.12% Test: 100%	10,000x dilution + Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	98.17%	CV: 96.75% Test: 100%	10,000x dilution + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	96.34%	CV: 95.12% Test: 97.56%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	**98.17%**	**CV: 96.75% Test: 100%**	**10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment**

Supplementary Table 1.2

The features used are one of two metrics: (1) the time derivative of growth curves and (2) the growth rate. We run the 4-fold cross validation protocol such that the top k (k = 5) predictions are used to predict each sample in the test set. The accuracy for all combinations of growth conditions are compared to the accuracy associated with random chance; the top three conditions are highlighted in bold.

SVM Parameters	Average Test Accuracy	Growth Condition
Predictions below are based on time derivative of growth curves using a SNP approach to strain definition (203 unique strains).		
Random chance	**2.46%**	**----------top 5----------**
Quadratic kernel, C = 10, kkt = 0	97.04%	10,000x dilution
Rbf kernel, C = 100, kkt = 0.1, σ = 10	96.67%	Phage treatment
Rbf kernel, C= 100, kkt = 0, σ = 20	97.04%	100x dilution
Rbf kernel, C = 10, kkt = 0.1, σ = 20	96.18%	Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 10	97.78%	10,000x dilution + Phage treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 10	97.78%	10,000x dilution + 100x dilution
Quadratic kernel, C = 10, kkt = 0	97.66%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 10	97.66%	Phage treatment + 100x dilution
Quadratic kernel, C = 10, kkt = 0	97.78%	Phage treatment + Carbenicillin treatment
Quadratic kernel, C = 10, kkt = 0	**98.03%**	**100x dilution + Carbenicillin treatment**
Linear kernel, C = 10, kkt = 0	**98.15%**	**10,000x dilution + Phage treatment + 100x dilution**
Linear kernel, C = 10, kkt = 0	97.91%	10,000x dilution + Phage treatment + Carbenicillin treatment
Linear kernel, C = 10, kkt = 0	97.78%	10,000x dilution + 100x dilution + Carbenicillin treatment
Linear kernel, C = 10, kkt = 0	97.41%	Phage treatment + 100x dilution + Carbenicillin treatment
Linear kernel, C = 10, kkt = 0	**98.28%**	**10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment**
Predictions below are based on growth rate time courses using a SNP approach to strain definition (203 unique strains).		
Random chance	**2.46%**	**----------top 5----------**
Rbf kernel, C = 10, kkt = 0, σ = 10	92.24%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, σ = 10	93.72%	Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	93.84%	100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 10	81.28%	Carbenicillin treatment

Rbf kernel, C = 100, kkt = 0, σ = 20	96.18%	10,000x dilution + Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	96.06%	10,000x dilution + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	95.69%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	96.55%	Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	95.32%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	95.32%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	96.80%	10,000x dilution + Phage treatment + 100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 20	**97.54%**	**10,000x dilution + Phage treatment + Carbenicillin treatment**
Rbf kernel, C = 10, kkt = 0, σ = 20	**97.78%**	**10,000x dilution + 100x dilution + Carbenicillin treatment**
Rbf kernel, C = 100, kkt = 0, σ = 20	97.17%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	**98.03%**	**10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment**

Supplementary Table 1.3

We use the smoothed growth curves as the features to the 4-fold cross validation procedure (k = 1). The accuracy for all combinations of growth conditions is compared to the accuracy associated with random chance; the top three conditions are highlighted in bold. Additionally, we compare the use of phylogeny based on a core set of genes (MLST) and SNPs. When using either approach for strain definition, the conclusion remains the same.

SVM Parameters	Average Test Accuracy	Growth Condition
Predictions below use a SNP approach to strain definition (203 unique strains).		
Random chance	**0.49%**	**----------top 1----------**
Rbf kernel, C = 1000, kkt = 0.1, σ = 20	87.93%	10,000x dilution
Rbf kernel, C = 1000, kkt = 0.1, σ = 20	83.25%	Phage treatment
Rbf kernel, C= 100, kkt = 0.1, σ = 10	77.96%	100x dilution
Rbf kernel, C = 1000, kkt = 0.1, σ = 10	81.40%	Carbenicillin treatment
Rbf kernel, C = 1000, kkt = 0.1, σ = 10	92.61%	10,000x dilution + Phage treatment
Rbf kernel, C = 1000, kkt = 0.1, σ = 10	92.86%	10,000x dilution + 100x dilution
Rbf kernel, C = 1000, kkt = 0.05, σ = 20	92.61%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 20	90.64%	Phage treatment + 100x dilution
Rbf kernel, C = 1000, kkt = 0.1, σ = 20	90.52%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 20	90.02%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 1000, kkt = 0.1, σ = 20	**94.58%**	**10,000x dilution + Phage treatment + 100x dilution**
Rbf kernel, C = 100, kkt = 0.1, σ = 20	94.34%	10,000x dilution + Phage treatment + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 20	93.35%	10,000x dilution + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 20	**94.46%**	**Phage treatment + 100x dilution + Carbenicillin treatment**
Rbf kernel, C = 100, kkt = 0.05, σ = 20	**95.69%**	**10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment**
Predictions below use a MLST approach to strain definition (41 unique strains).		
Random chance	**2.4%**	**----------top 1----------**
Linear kernel, C = 10, kkt = 0	96.34%	10,000x dilution
Linear kernel, C = 10, kkt = 0	93.29%	Phage treatment
Rbf kernel, C = 1000, kkt = 0, σ = 20	93.90%	100x dilution
Rbf kernel, C = 1000, kkt = 0, σ = 20	91.46%	Carbenicillin treatment

Rbf kernel, C = 10, kkt = 0, σ = 10	**97.56%**	**10,000x dilution + Phage treatment**
Rbf kernel, C = 100, kkt = 0, σ = 20	**98.17%**	**10,000x dilution + 100x dilution**
Linear kernel, C = 10, kkt = 0	93.90%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	96.95%	Phage treatment + 100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 20	93.29%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 1000, kkt =0 , σ = 20	94.51%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	**99.39%**	**10,000x dilution + Phage treatment + 100x dilution**
Rbf kernel, C = 100, kkt = 0, σ = 20	95.12%	10,000x dilution + Phage treatment + Carbenicillin treatment
Linear kernel, C = 10, kkt = 0	95.12%	10,000x dilution + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	94.51%	Phage treatment + 100x dilution + Carbenicillin treatment
Linear kernel, C = 10, kkt = 0	96.95%	10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment

Supplementary Table 1.4

We apply the 4-fold cross validation procedure with SVM to a dataset using AUC, μ (maximum of time derivative of growth curves), or a combination of AUC and μ as the covariates (k = 1). SVM hyperparameters are selected from the following options: kernel function (rbf), kkt violation level (0, 0.05, 0.1), box constraint of the soft margin C (10, 100), and rbf scaling factor σ (1, 10, 20). The accuracy for all combinations of growth conditions is compared to the accuracy associated with random chance; the top three conditions are highlighted in bold. Additionally, we compare the use of phylogeny based on a core set of genes (MLST) and SNPs. When using either approach for strain definition, the conclusion remains the same.

Average Test Accuracy			**Growth Condition**
μ	**AUC**	**AUC + μ**	
Predictions below use a SNP approach to strain definition (203 unique strains).			
0.49%			**-------random chance top 1-------**
1.72%	7.27%	14.29%	10,000x dilution
2.09%	9.48%	19.95%	Phage treatment
2.71%	7.64%	15.27%	100x dilution
2.34%	11.58%	22.78%	Carbenicillin treatment
8.25%	32.02%	46.55%	10,000x dilution + Phage treatment
8.62%	21.01%	45.32%	10,000x dilution + 100x dilution
9.61%	40.64%	48.28%	10,000x dilution + Carbenicillin treatment
9.61%	26.60%	49.14%	Phage treatment + 100x dilution
8.99%	45.57%	53.94%	Phage treatment + Carbenicillin treatment
9.98%	39.90%	50.00%	100x dilution + Carbenicillin treatment
15.89%	57.14%	**66.26%**	10,000x dilution + Phage treatment + 100x dilution
16.87%	**67.36%**	65.52%	10,000x dilution + Phage treatment + Carbenicillin treatment
18.47%	59.61%	63.67%	10,000x dilution + 100x dilution + Carbenicillin treatment
19.95%	**65.76%**	**70.20%**	Phage treatment + 100x dilution + Carbenicillin treatment
26.11%	**77.96%**	**77.46%**	10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment
Predictions below use a MLST approach to strain definition (41 unique strains).			
2.4%			**-------random chance top 1-------**
13.42%	25.00%	42.07%	10,000x dilution
14.02%	24.39%	57.32%	Phage treatment
11.59%	18.90%	40.24%	100x dilution
14.63%	38.41%	57.32%	Carbenicillin treatment
30.49%	67.07%	75.00%	10,000x dilution + Phage treatment
29.27%	62.20%	70.12%	10,000x dilution + 100x dilution
32.93%	70.73%	76.22%	10,000x dilution + Carbenicillin treatment
36.59%	53.05%	81.10%	Phage treatment + 100x dilution
37.20%	70.12%	88.41%	Phage treatment + Carbenicillin treatment

35.98%	68.90%	75.61%	100x dilution + Carbenicillin treatment
46.34%	78.66%	88.41%	10,000x dilution + Phage treatment + 100x dilution
50.00%	**85.37%**	**90.85%**	10,000x dilution + Phage treatment + Carbenicillin treatment
45.12%	**84.76%**	82.93%	10,000x dilution + 100x dilution + Carbenicillin treatment
54.27%	82.32%	**90.24%**	Phage treatment + 100x dilution + Carbenicillin treatment
62.80%	**89.63%**	**93.29%**	10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment

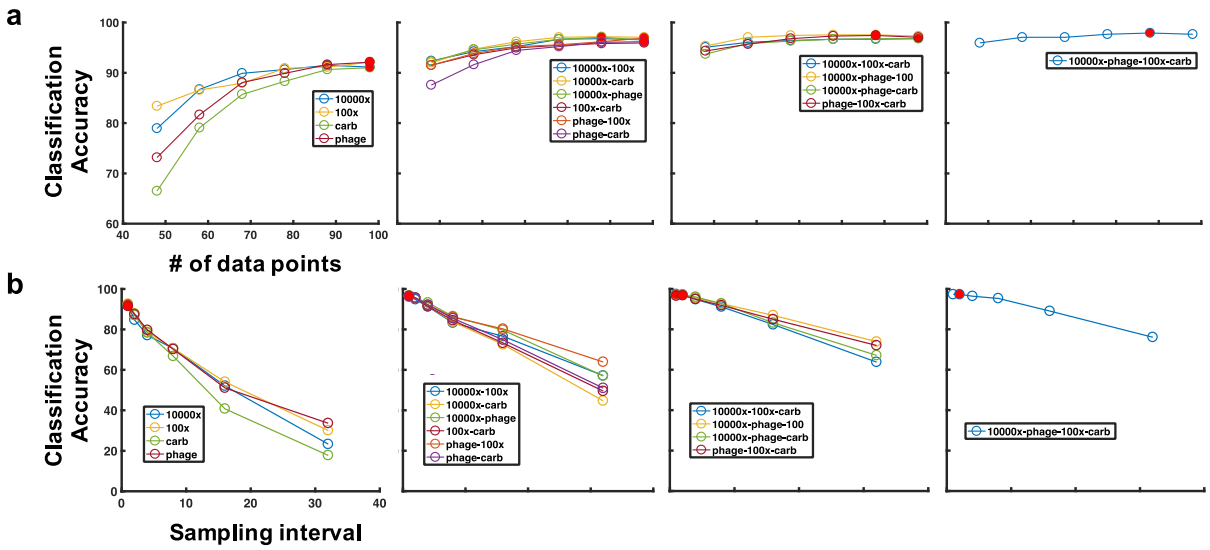
Supplementary Figure 1.1

The impact of experimental time span and sampling frequency on strain identification.

We vary the time span and sampling frequency to examine the effect of controlling the number of time points on strain identification accuracy (features are time derivative of growth curves). Here, we define strains through phylogeny based on SNPs. From left to right, the number of growth conditions used as the training dataset increases. In each plot, the red filled points indicate the number of data points resulting in the greatest classification accuracy for each dataset. In both cases, increasing the number of growth conditions can prevent the decrease in classification accuracy as the time span shortens or as the sampling interval increases.

(a) Time span varies

(b) Sampling interval varies



1.2 Antibiotic resistance prediction: Supplementary Tables 1.5-1.8, Supplementary Figure 1.2-1.3

In this section, we described the accuracy of antibiotic resistance predictions with the corresponding optimal parameter set using SVM for four antibiotics spanning four classes – SAM, GM, SXT, and CIP **(Supplementary Tables 1.5-1.8)**. For all antibiotics, we compared the predictive accuracy using two types of features: (1) time derivative of growth curve and (2) growth rate. The results using the first are reported in the main text.

We modified the traditional cross validation procedure and take the average (accuracy, true positive rate, and true negative rate) across 244 models where each model was trained on the replicates of 243 isolates and the test set consists of the replicates for 1 isolate. This provided an estimate of the predictive potential for resistance given a larger dataset. If multiple parameter sets resulted in the highest accuracy then only one is shown. In this case, we defined the model with the highest accuracy as the one where the average of the accuracy, true positive rate, and true negative rate is the highest. Parameters were selected from the following options: kernel function (rbf), kkt violation level (0, 0.1), box constraint of the soft margin C (10, 100, 1000), and rbf scaling factor σ (1, 10, 20). We additionally visualized these results in terms of ROC curves. Using the time derivative of growth curves, those for SAM and SXT are found in the main text and those for GM and CIP are in **Supplementary Figure 1.2a**. Using growth rate as the features, those for all four antibiotics are in **Supplementary Figure 1.2b**.

MATLAB files in package (https://github.com/youlab/strain_prediction_CZ):
predictResistance_clinicalIsolates.m (main file), clinResSVMOpt.m,
Figure5b_resistance_prediction_GC_roc.m (main file),
predictResistance_clinicalIsolates_traditionalGR.m (main file), clinResSVMOpt_mod.m,
Figure5b_resistance_prediction_GC_roc_traditionalGR.m (main file)

Supplementary Table 1.5

Results for SAM. The features used are one of two metrics: (1) time derivative of growth curve and (2) growth rate. For SAM, 135 out of 244 isolates are classified as resistant according to standard disk diffusion.

SVM Parameters	Average Accuracy	Average TPR	Average TNR	Growth Condition
Predictions below are based on time derivative of growth curves.				
Rbf kernel, C = 10, kkt = 0.1, $\boldsymbol{\sigma}$ = 20	68.65%	68.89%	68.35%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 20	66.91%	69.63%	63.53%	Phage treatment

Rbf kernel, C = 1000, kkt = 0, **σ** = 10	66.50%	74.49%	54.59%	100x dilution
Rbf kernel, C = 100, kkt = 0, **σ** = 20	72.54%	72.96%	72.02%	Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, **σ** = 20	70.80%	72.96%	68.12%	10,000x dilution + Phage treatment
Rbf kernel, C = 10, kkt = 0.1, **σ** = 10	67.21%	72.04%	61.24%	10,000x dilution + 100x dilution
Rbf kernel, C = 10, kkt = 0, **σ** = 10	81.45%	84.07%	78.21%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, **σ** = 10	62.70%	68.52%	55.50%	Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, **σ** = 10	77.77%	81.67%	72.94%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, **σ** = 10	70.90%	77.78%	62.39%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, **σ** = 10	68.95%	73.70%	63.07%	10,000x dilution + Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, **σ** = 10	82.58%	88.52%	75.23%	10,000x dilution + Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, **σ** = 10	81.25%	87.04%	74.08%	10,000x dilution + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, **σ** = 10	77.66%	84.81%	68.81%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, **σ** = 20	82.99%	87.78%	77.06%	10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment
Predictions below are based on growth rate time courses.*				
Rbf kernel, C = 10, kkt = 0, **σ** = 20	68.14%	67.59%	68.81%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, **σ** = 20	55.94%	67.04%	42.20%	100x dilution
Rbf kernel, C = 10, kkt = 0, **σ** = 20	66.39%	68.15%	64.22%	10,000x dilution + 100x dilution

***Predictions are not reported for conditions when SVM is unable to converge (within 10,000,000 iterations).**

Supplementary Table 1.6

Results for GM. The features used are one of two metrics: (1) time derivative of growth curve and (2) growth rate. For GM, only ~15% of the isolates (39/244) are positive for resistance according to standard disk diffusion, resulting in an imbalanced dataset. So, the optimal parameter set is determined by the average true positive rate.

SVM Parameters	Average Accuracy	Average TPR	Average TNR	Growth Condition
Predictions below are based on time derivative of growth curves.				
Rbf kernel, C = 10, kkt = 0, σ = 20	71.93%	40.38%	77.93%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	68.95%	33.33%	75.73%	Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	62.81%	28.21%	69.39%	100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	72.13%	33.97%	79.39%	Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	77.97%	20.51%	88.90%	10,000x dilution + Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	75.31%	19.23%	85.98%	10,000x dilution + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	77.46%	19.23%	88.54%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	72.75%	16.03%	83.54%	Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	75.41%	17.31%	86.46%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	74.49%	19.87%	84.88%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	80.33%	15.38%	92.68%	10,000x dilution + Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	80.02%	12.82%	92.80%	10,000x dilution + Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	79.10%	10.90%	92.07%	10,000x dilution + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	78.07%	11.54%	90.73%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	80.74%	8.97%	94.39%	10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment
Predictions below are based on growth rate time courses.*				

Rbf kernel, C = 1000, kkt = 0, σ = 20	74.18%	28.85%	82.80%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, σ = 1	59.94%	37.82%	64.15%	100x dilution
Rbf kernel, C = 100, kkt = 0, σ = 20	68.44%	31.41%	75.49%	10,000x dilution + 100x dilution

***Predictions are not reported for conditions when SVM is unable to converge (within 10,000,000 iterations).**

Supplementary Table 1.7

Results for SXT. The features used are one of two metrics: (1) time derivative of growth curve and (2) growth rate. For SXT, 140 out of 244 isolates are classified as resistant according to standard disk diffusion.

SVM Parameters	Average Accuracy	Average TPR	Average TNR	Growth Condition
Predictions below are based on time derivative of growth curves.				
Rbf kernel, C = 10, kkt = 0, σ = 20	59.94%	63.57%	55.05%	10,000x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	60.04%	61.07%	58.65%	Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	55.53%	55.00%	56.25%	100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	68.34%	73.57%	61.30%	Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	64.34%	68.75%	58.41%	10,000x dilution + Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	58.61%	61.96%	54.09%	10,000x dilution + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	68.55%	71.61%	64.42%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0.1, σ = 20	61.58%	65.54%	56.25%	Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 10	74.08%	80.00%	66.11%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	67.32%	72.86%	59.86%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	61.99%	71.96%	48.56%	10,000x dilution + Phage treatment + 100x dilution

Rbf kernel, C = 100, kkt = 0, σ = 20	69.47%	73.93%	63.46%	10,000x dilution + Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	65.47%	72.14%	56.49%	10,000x dilution + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0, σ = 20	70.70%	77.14%	62.02%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	68.85%	75.18%	60.34%	10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment
Predictions below are based on growth rate time courses.*				
Rbf kernel, C = 10, kkt = 0, σ = 20	64.96%	64.46%	65.62%	10,000x dilution
Rbf kernel, C = 10, kkt = 0.1, σ = 20	57.38%	59.11%	55.05%	100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	58.81%	60.54%	56.49%	10,000x dilution + 100x dilution

***Predictions are not reported for conditions when SVM is unable to converge (within 10,000,000 iterations).**

Supplementary Table 1.8

Results for CIP. The features used are one of two metrics: (1) time derivative of growth curve and (2) growth rate. For CIP, 146 out of 244 isolates are classified as resistant according to standard disk diffusion

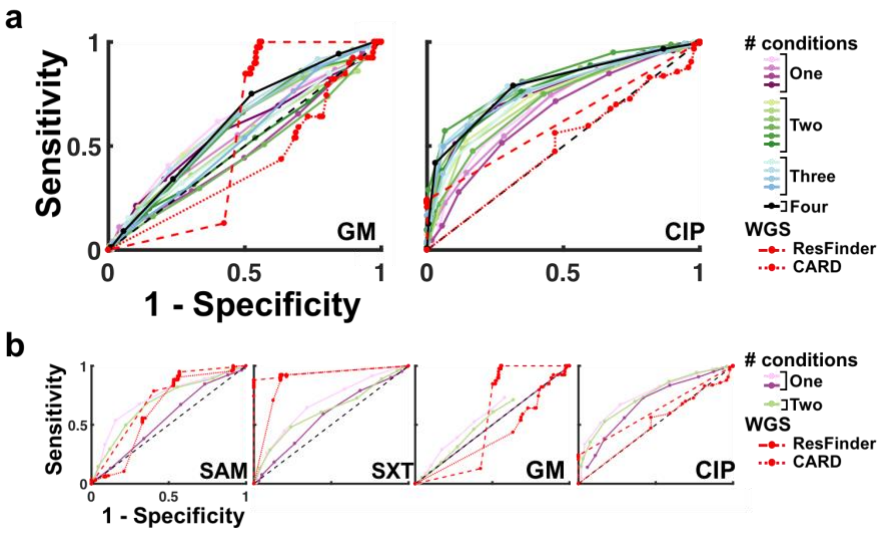
SVM Parameters	Average Accuracy	Average TPR	Average TNR	Growth Condition
Predictions below are based on time derivative of growth curves.				
Rbf kernel, C = 10, kkt = 0, σ = 20	65.16%	67.12%	62.24%	10,000x dilution
Rbf kernel, C = 1000, kkt = 0, σ = 10	66.50%	74.49%	54.59%	Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	63.93%	71.40%	52.81%	100x dilution
Rbf kernel, C = 10, kkt = 0.1, σ = 20	71.41%	68.84%	75.26%	Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	69.36%	74.32%	61.99%	10,000x dilution + Phage treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	68.03%	75.00%	57.65%	10,000x dilution + 100x dilution

Rbf kernel, C = 10, kkt = 0, σ = 20	72.23%	75.86%	66.84%	10,000x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	67.73%	74.83%	57.14%	Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 10	74.49%	80.82%	65.05%	Phage treatment + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 20	71.72%	76.03%	65.31%	100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	71.31%	77.23%	62.50%	10,000x dilution + Phage treatment + 100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 10	75.31%	78.60%	70.41%	10,000x dilution + Phage treatment + Carbenicillin treatment
Rbf kernel, C = 100, kkt = 0.1, σ = 20	70.90%	76.54%	62.50%	10,000x dilution + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	73.98%	80.14%	64.80%	Phage treatment + 100x dilution + Carbenicillin treatment
Rbf kernel, C = 10, kkt = 0, σ = 10	74.69%	78.94%	68.37%	10,000x dilution + Phage treatment + 100x dilution + Carbenicillin treatment
Predictions below are based on growth rate time courses.*				
Rbf kernel, C = 10, kkt = 0, σ = 20	69.36%	66.95%	72.96%	10,000x dilution
Rbf kernel, C = 100, kkt = 0, σ = 10	66.91%	72.95%	57.91%	100x dilution
Rbf kernel, C = 10, kkt = 0, σ = 20	67.11%	69.69%	63.27%	10,000x dilution + 100x dilution

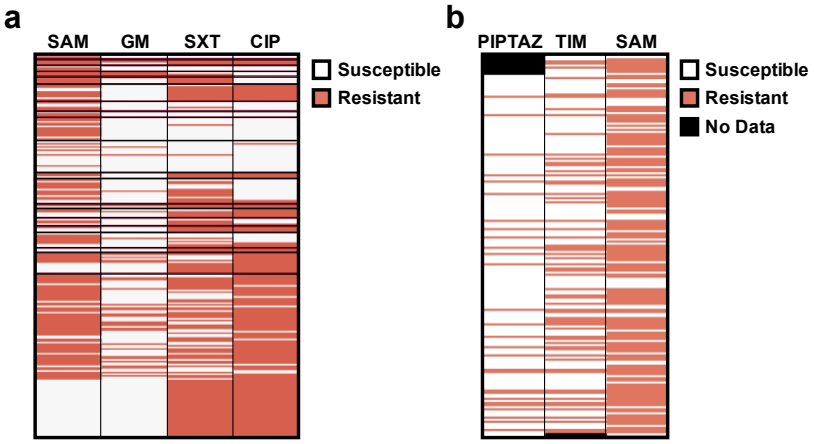
***Predictions are not reported for conditions when SVM is unable to converge (within 10,000,000 iterations).**

Supplementary Figure 1.2
Compare phenotype-based and WGS-based resistance prediction.

We compare ROC curves for the phenotype-based predictions (time derivative of growth curve) and genotype-based predictions (in red).
(a) Features are the time derivative of growth curve.
(b) Features are the growth rate.



Supplementary Figure 1.3
Compare resistance profile for clinical isolate library.
(a) We compare the resistance profiles (SAM, GM, SXT, CIP) of isolates defined as the same strain based on an MLST approach to strain definition. Each set of isolates clustered as one strain is separated by a black horizontal line.
(b) We compare the resistance profiles (PIPTAZ: piperacillin-tazobactam, TIM: ticarcillin-clavulanate, SAM) of the 185 clinical isolates sourced from blood samples due to antibiotic resistance information for additional antibiotics in the Beta-lactam class. This shows that resistance to one antibiotic in a class does not necessarily imply resistance to others in the same class.



1.3 Limitations of method for strain prediction - Keio Collection: Supplementary Tables 1.9-1.10

The dataset we utilized in this section was derived from published work in which growth curves of 97% of the Keio collection were collected with a total of 49 time points per strain.[1] Because the collection consisted of isolates each with single gene knockouts, we assumed each isolate to be a unique strain. Unlike the experimental protocol we design **(Methods)**, this dataset is more limited in terms of the number of growth conditions, the resolution of the growth dynamics, and the number of replicates; simultaneously, this dataset contained an order of magnitude greater number of strains (a total of 3,866 strains).

Here, the features used were one of two metrics related: (1) time derivative of growth curve and (2) growth rate. The following analysis included accuracy of strain identification with 3-fold cross validation along with optimal parameter set using multiclass Support Vector Machines. If multiple parameter sets resulted in the highest accuracy then only one was shown. Parameters were selected from the following options: kernel function (rbf), kkt violation level (0, 0.05, 0.1), box constraint of the soft margin C (10, 100), and rbf scaling factor σ (1, 10, 20).

Matlab files in package (https://github.com/youlab/strain_prediction_CZ):
predictStrain_keio_singlePlate.m, predictStrain_keio_singlePlate_traditionalGR.m, predictStrain_keio_allPlates.m

Supplementary Table 1.9

The features used are one of two metrics: (1) time derivative of growth curve and (2) growth rate. We display the accuracy for 3-fold cross validation with the corresponding optimal SVM parameter set on a few sample plates and display the accuracy of the average across the test sets in comparison to the prediction due to random chance (k = 1 or 10).

SVM Parameters	k	Average Test Accuracy	Plate	Number of Strains	Random Chance
Predictions below are based on time derivative of growth curves.					
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	1	10.53%	5	76	1.32%
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	10	41.67%			13.2%
Rbf kernel, C = 100, kkt = 0, $\boldsymbol{\sigma}$ = 10	1	14.21%	7	68	1.47%
Rbf kernel, C = 100, kkt = 0, $\boldsymbol{\sigma}$ = 10	10	47.06%			14.7%
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 1	1	2.75%	9	85	1.18%
Rbf quadratic, C = 10, kkt = 0	10	18.43%			11.8%
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	1	14.05%	11	76	1.32%
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 10	10	41.67%			13.2%
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 20	1	11.62%	13	66	1.52%
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 20	10	46.97%			15.2%
Rbf kernel, C = 1000, kkt = 0, $\boldsymbol{\sigma}$ = 10	1	70.13%	15	77	1.30%

Rbf kernel, C = 10, kkt = 0, σ = 10	10	78.79%			13.0%
Linear kernel, C = 10, kkt = 0	1	1.85%	17	72	1.39%
Rbf kernel, C = 10, kkt = 0, σ = 20	10	20.37%			13.9%
Rbf kernel, C = 100, kkt = 0, σ = 10	1	12.82%	19	78	1.28%
Rbf kernel, C = 10, kkt = 0, σ = 20	10	42.31%			12.8%
Predictions below are based on growth rate time courses.					
Rbf kernel, C = 100, kkt = 0, σ = 10	1	14.04%	5	76	1.32%
Rbf kernel, C = 100, kkt = 0, σ = 20	10	48.68%			13.2%
Rbf kernel, C = 1000, kkt = 0, σ = 10	1	17.65%	7	68	1.47%
Rbf kernel, C = 100, kkt = 0, σ = 20	10	54.90%			14.7%
Rbf kernel, C = 1000, kkt = 0, σ = 10	1	3.92%	9	85	1.18%
Rbf kernel, C = 100, kkt = 0, σ = 10	10	26.67%			11.8%
Rbf kernel, C = 100, kkt = 0, σ = 20	1	17.54%	11	76	1.32%
Rbf kernel, C = 100, kkt = 0, σ = 20	10	55.26%			13.2%
Rbf kernel, C = 1000, kkt = 0, σ = 10	1	18.18%	13	66	1.52%
Rbf kernel, C = 100, kkt = 0, σ = 20	10	54.55%			15.2%
Rbf kernel, C = 100, kkt = 0, σ = 10	1	47.62%	15	77	1.30%
Rbf kernel, C = 10, kkt = 0, σ = 10	10	80.09%			13.0%
Rbf kernel, C = 10, kkt = 0, σ = 10	1	2.78%	17	72	1.39%
Rbf kernel, C = 10, kkt = 0, σ = 10	10	22.69%			13.9%
Rbf kernel, C = 1000, kkt = 0, σ = 10	1	7.27%	19	78	1.28%
Rbf kernel, C = 10, kkt = 0, σ = 10	10	35.90%			12.8%

Supplementary Table 1.10

We display the accuracy for 3-fold cross validation with the corresponding optimal SVM parameter set. Here, we run the cross validation protocol on 3,866 strains in the Keio collection such that the top k (k = 10, 50, and 100) predictions identify each sample in the test set. The features used are one of two metrics: (1) time derivative of growth curve and (2) growth rate.

SVM Parameters	Average Test Accuracy	k	Random chance
Predictions below are based on time derivative of growth curves.			
Rbf kernel, C = 10, kkt = 0.05, σ = 1	12.69%	10	0.26%
Rbf kernel, C = 10, kkt = 0.05, σ = 1	12.69%	50	1.29%
Rbf kernel, C = 10, kkt = 0, σ = 10	33.57%	100	2.59%
Predictions are based on growth rate time courses.			
Rbf kernel, C = 100, kkt = 0.05, σ = 10	14.05%	10	0.26%
Rbf kernel, C = 100, kkt = 0.05, σ = 10	28.30%	50	1.29%
Rbf kernel, C = 100, kkt = 0.05, σ = 20	37.91%	100	2.59%

1.4 Strain prediction on environmental isolates: Supplementary Table 1.11

We described the accuracy of strain identification with 3-fold cross validation along with an optimal parameter set using multiclass Support Vector Machines for a library of 143 unique environmental isolates collected across Duke University. We have 12 replicates of all strains, so we used a hold out cross validation procedure. We set aside 3 replicates per strain as the validation set. A 3-fold cross validation procedure was applied to the other 9 replicates per strain such that per fold 3 replicates were used as the test set and 6 replicates were used as the training set. We reported the average accuracy across these test sets as well as the accuracy of the validation set. If multiple parameter sets resulted in the highest accuracy then only one was shown. Parameters were selected from the following options: kernel function (quadratic, rbf), kkt violation level (0, 0.05, 0.1), box constraint of the soft margin C (10, 100, 1000), and rbf scaling factor σ (1, 5, 10, 20). In **Supplementary Table 1.11**, we described these predictions using two metrics as the features: (1) the time derivative of the growth curve and (2) the growth rate. For this particular dataset, growth rate performed better than the time derivative, the results of both features were reported in the main text.

Matlab files in package (https://github.com/youlab/strain_prediction_CZ):
predict_EnvironmentalIsolates.m (main file), envSVMOpt.m, importEnvironmentalIsolates.m, predict_EnvironmentalIsolates_traditionalGR.m (main file), importEnvironmentalIsolates_traditionalGR.m

Supplementary Table 1.11

We describe the predictions of genetic identity for the environmental isolate library (143 strain classes). The features used are one of two metrics: (1) time derivative of growth curve and (2) growth rate.

SVM Parameters	Average Test Accuracy of 3-fold CV	Hold Out Test Accuracy (validation set)	# of strains	k	Random chance
Predictions below are based on time derivative of growth curves.					
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 5	78.71%	82.75%	143	1	**0.70%**
Predictions are based on growth rate time courses.					
Rbf kernel, C = 10, kkt = 0, $\boldsymbol{\sigma}$ = 5	86.56%	90.68%	143	1	**0.70%**

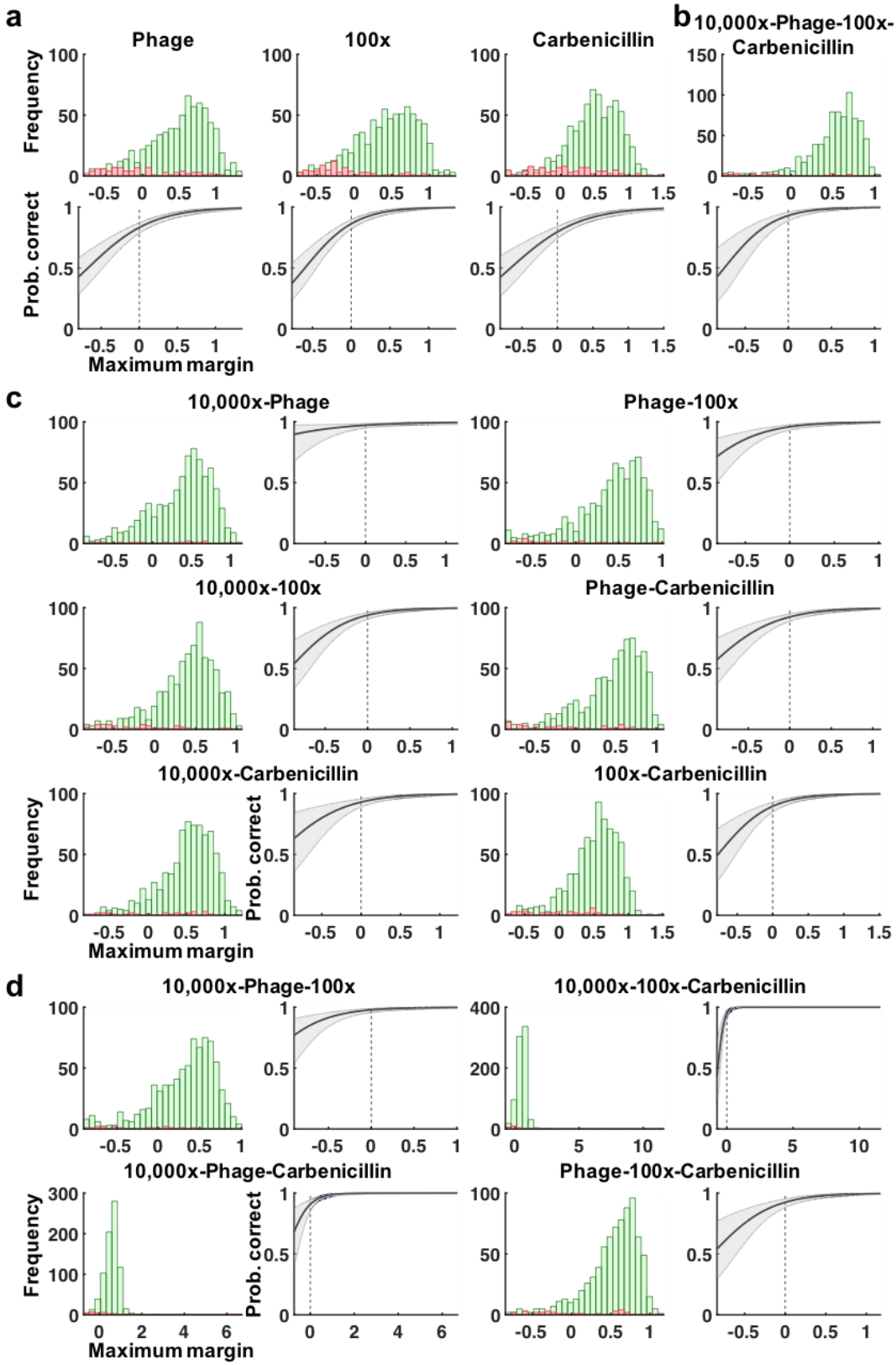
1.5 An empirical confidence interval for support vector machines: Supplementary Figure 1.4

In this section, we described the development of an estimate for the confidence of the model predictions for the strain-level classification. Since we used a one-versus-all SVM approach, we generate one classifier per class for a total of $\boldsymbol{N}$ classifiers. With this model, classifier $\boldsymbol{i}$ draws a hyperplane between the training set of class $\boldsymbol{i}$ and the training set for all other classes. So, each sample in the test set is associated with $\boldsymbol{N}$ margins (distance between the data point and the hyperplane). To predict the class of the unknown samples, we choose the class with the maximum margin (out of $\boldsymbol{N}$ margins). Generally, a larger margin indicates a higher confidence. As such, for each sample, we define a metric, the maximum margin, as the greatest value of the set of $\boldsymbol{N}$ margins. We use the maximum margins of the test sets (using the 4-fold cross validation procedure described in **Supplementary Section 1.1**) to generate a frequency distribution of all predictions and compare this distribution to that of all incorrectly predicted samples. We label each of the values in these vectors according to whether they were classified correctly (1 is correctly classified and 0 is misclassified) and use logistic regression to plot the probability of the prediction being correct.

Matlab files in package (https://github.com/youlab/strain_prediction_CZ):
confidenceInterval_clinicalIsolates.m (main file)

Supplementary Figure 1.4
Empirical generation of confidence interval for strain identification SVM models.

We illustrate the density curve of the maximum margin for all predictions in the test set and the estimated confidence for given maximum margins (features are time derivative of growth curves). We overlay the frequency distribution of all correctly classified points in the test set (green) and the frequency distribution of the misclassified points (red). Given these distributions, we plot the logistic regression curve. **(a)** Single growth conditions;**(b)** Quadruple combination growth condition; **(c)** Double combination growth conditions; **(d)** Triple combination growth conditions


a
Phage
100x
Carbenicillin
b
10,000x-Phage-100x-Carbenicillin
Frequency
Prob. correct
Maximum margin
c
10,000x-Phage
Phage-100x
10,000x-100x
Phage-Carbenicillin
10,000x-Carbenicillin
100x-Carbenicillin
d
10,000x-Phage-100x
10,000x-100x-Carbenicillin
10,000x-Phage-Carbenicillin
Phage-100x-Carbenicillin

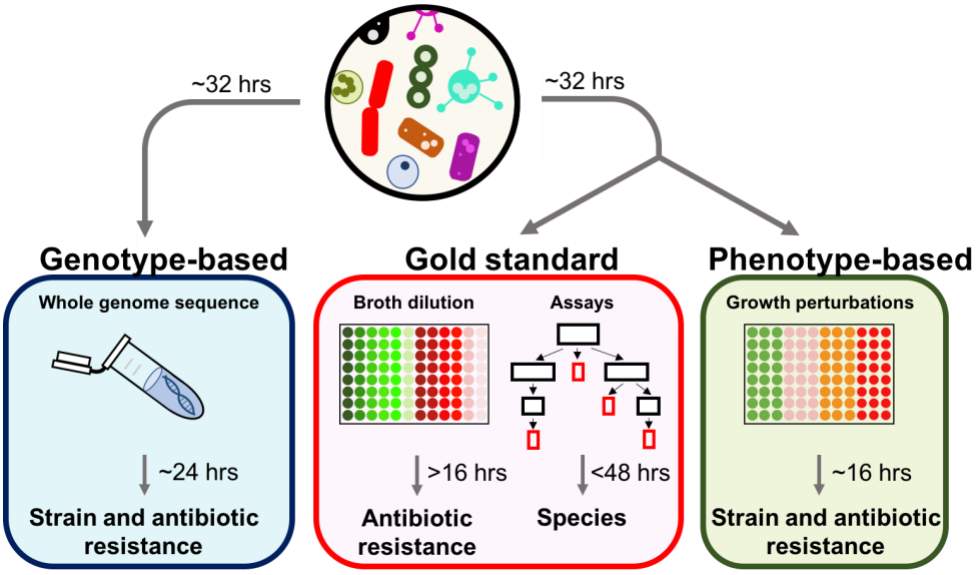
1.6 A comparison to current technologies: Supplementary Figure 1.5

Although the experimental method we have described in the main text is a proof of concept approach, we compare the results and workflow to that of other current technologies**.** Specifically, we examined the relative accuracies of antibiotic resistance prediction using WGS and growth dynamics. In **Supplementary Figure 1.5**, we illustrate the workflow of our approach in comparison to the current standard (used in the clinic) and other alternative approaches (sequence-based techniques).

Supplementary Figure 1.5

Comparison of current bacterial characterization strategies and our new framework.

Under the current protocol, the phenotype-based technique we developed has an estimated time to strain-level and antibiotic resistance identification of ~48 hours. In contrast, WGS requires ~56 hours and the current Gold Standard require ~80 hours to get the same information. In addition, the organism identification resolution is lower for the Gold Standard while WGS has a significantly higher associated cost.



1.7 Strain prediction - the effect of biological replicates on predictive power: Supplementary Table 1.12, Supplementary Figure 1.6

Using the methods described in the main text, we use a high throughput liquid handling protocol to devise a second dataset of growth curves for a subset of the clinical isolates. For this dataset, we generated growth curves for the clinical isolate library with 3 biological replicates each of which had 4 technical replicates (each growth curve had 145 data points). Here, we described the accuracy of strain identification of the clinical isolate library (194 strains using an SNP based approach or 41 strains using a MLST approach) using one of two methods: (1) 2-fold cross validation with holdout and (2) 3-fold cross validation. When multiple isolates were classified as being the same strain, one isolate was chosen at random to represent the strain. For the 2-fold cross validation procedure, we treated the 4 technical replicates per biological replicate as a distinct fold and held out all 4 technical replicates for the third biological replicate for the validation set. The accuracy across the two folds as well as the validation set are reported in **Supplementary table 1.12.** This procedure demonstrates the ability of the models to generalize to growth dynamics from biological replicates not present in the training dataset. Similarly, we used the 4 technical replicates for each biological replicate as one of the 3 folds for the 3-fold cross validation procedure. The average test set accuracy across the three folds is reported in **Supplementary table 1.12**. The features used to train the model were those described in **Supplementary Section 1.1**.

For both cross validation methods, we reported the optimal parameter set using multiclass Support Vector Machines. The top k strains predicted are checked against the true label (k = 1). If multiple parameter sets resulted in the highest accuracy, only one is shown. Parameters were selected from the following options: kernel function (linear, quadratic, rbf), kkt violation level (0, 0.05, 0.1), box constraint of the soft margin C (10, 100, 1000), and rbf scaling factor σ (1, 10, 20). Here, the label for each strain was decided according to phylogenetic analysis (described in **Methods** and **Supplementary Section 3**).

MATLAB files in package (https://github.com/youlab/strain_prediction_CZ):
importClinicalIsolates_replicates.m, clinSVMOpt_bioRep.m,
importClinicalIsolates_replicates_traditionalGR.m, SuppFigure_visualizeData_clin_bioRep.m

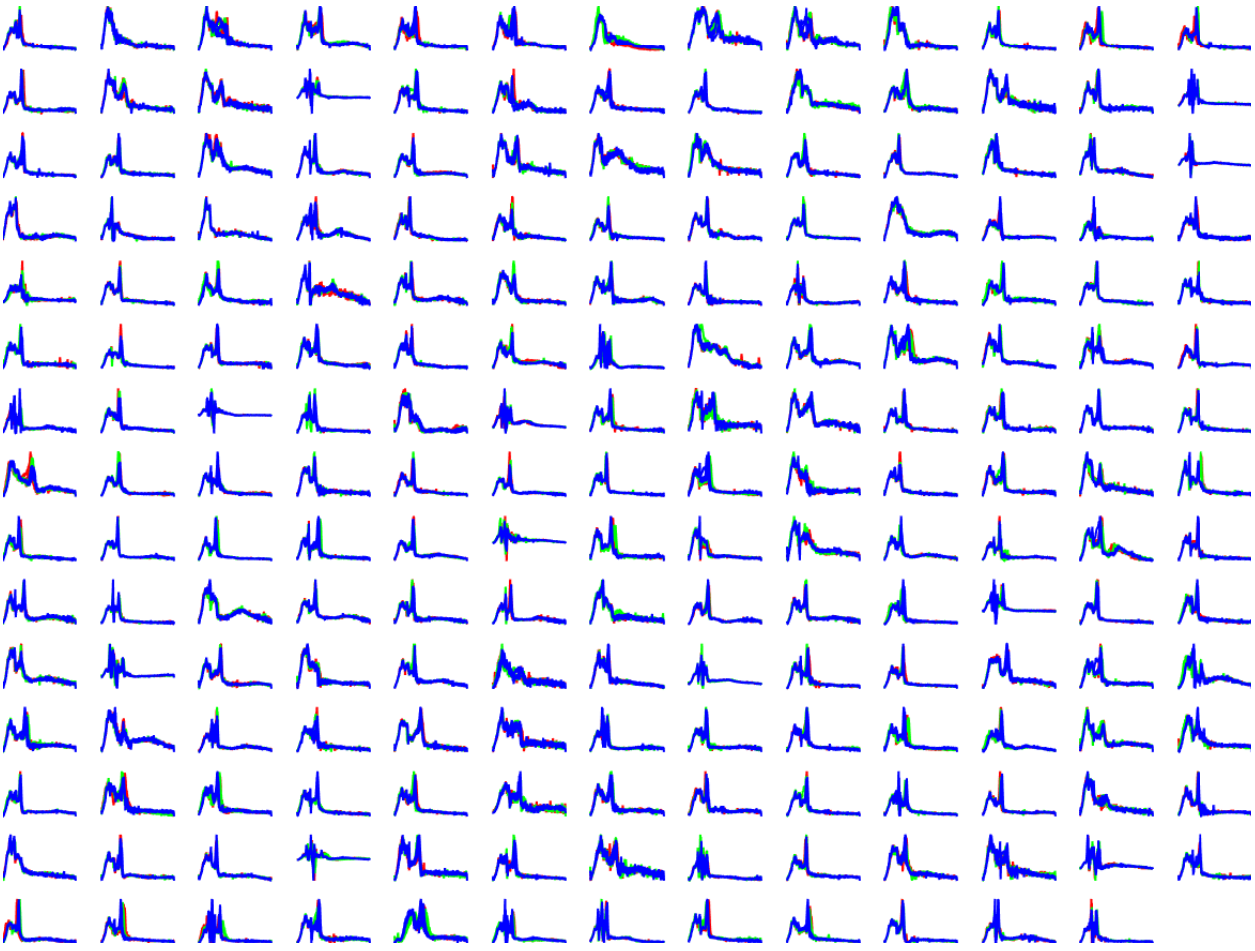
Supplementary Table 1.12

SVM Parameters	Average Test Accuracy 3-fold CV[1]	Hold Out Test Accuracy 2-fold CV[2]	Growth Condition
SNP approach to strain definition (194 unique strains).			
Random chance	**0.52%**	**0.52%**	**----------top 1----------**
Rbf kernel, C = 10, kkt = 0, σ = 10	**82.56%**	**CV: 81.06% Test: 74.87%**	**100x dilution – time derivative of growth curves**
Rbf kernel, C = 100, kkt = 0, σ = 10	81.01%	CV: 81.38% Test: 70.75%	100x dilution – growth rate time courses

Rbf kernel, C = 100, kkt = 0, σ = 10	69.07%	CV: 65.72% Test: 60.18%	100x dilution – smoothed growth curve
MLST approach to strain definition (41 unique strains).			
Random chance	**2.4%**	**2.4%**	**----------top 1----------**
Rbf kernel, C = 10, kkt = 0, σ = 10	**95.33%**	**CV: 97.26% Test: 90.85%**	**100x dilution – time derivative of growth curves**
Rbf kernel, C = 100, kkt = 0, σ = 20	93.70%	CV: 97.87% Test: 79.88%	100x dilution – growth rate time courses
Rbf kernel, C = 100, kkt = 0, σ = 10	89.43%	CV: 91.16% Test: 78.05%	100x dilution – smoothed growth curve

Supplementary Figure 1.6
A visualization of the features: biological replicates of the clinical isolates.

The time derivative of growth curves for the 194 clinical strains (as defined by SNPs) are illustrated with 3 biological replicates (in red, green, or blue), each of which has 4 technical replicates.



2. An overview of the data

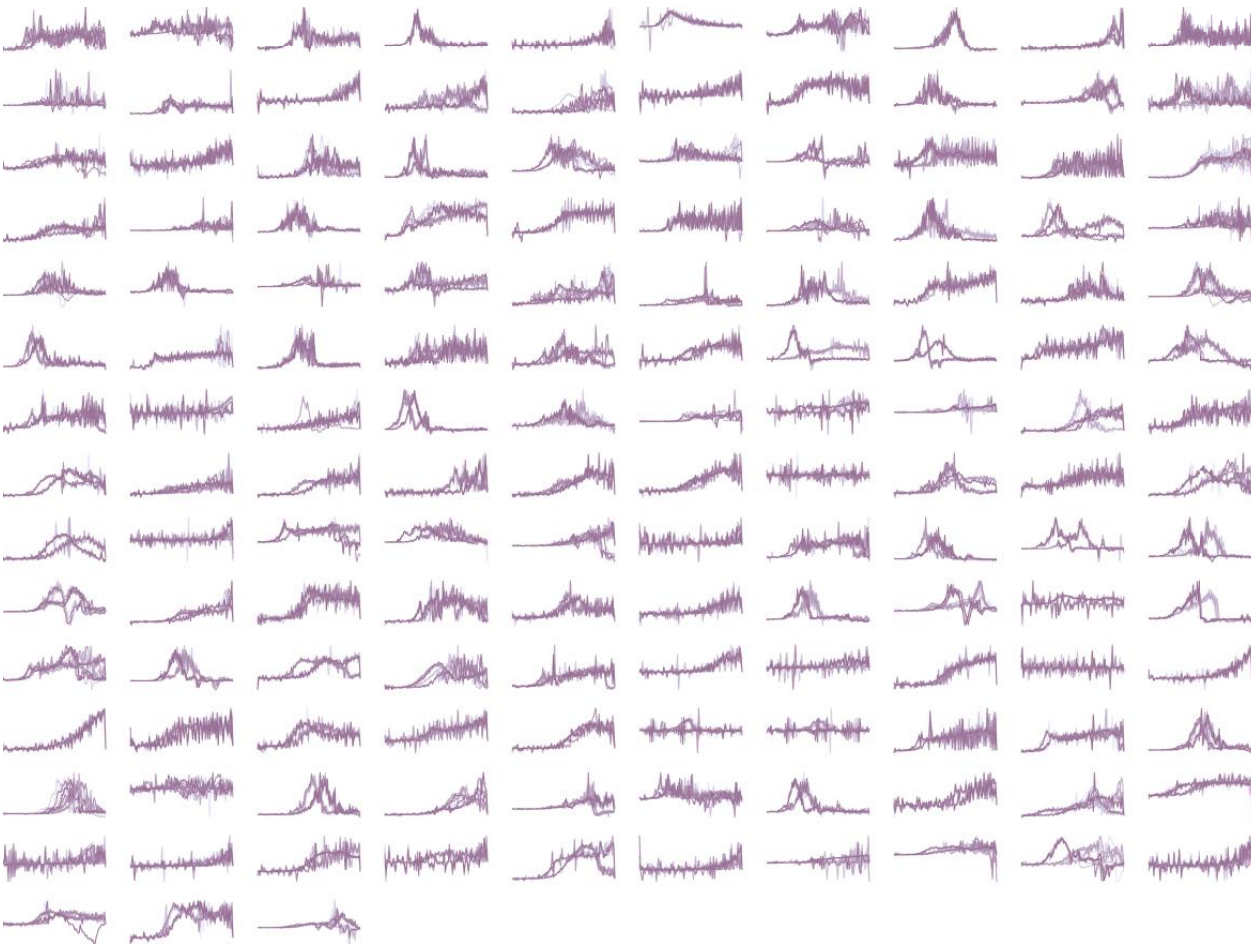
Using the methods described in the main text, we have generated growth dynamics for 203 unique clinical strains in replicates of 4 under 4 growth conditions (10,000x dilution, phage, 100x dilution, and carbenicillin) and 143 unique environmental strains in replicates of 12 under a single growth condition (10,000x dilution). The latter is illustrated by **Figure 2** and the former is visualized in **Supplementary Figure 2.1**.

MATLAB files in package (https://github.com/youlab/strain_prediction_CZ):
SuppFigure_visualizeData.m (main file)

2.1 An overview of the phenotypic landscape: Supplementary Figure 2.1

Supplementary Figure 2.1
A visualization of the features: environmental isolates.

The time derivative of growth curves for the 143 environmental strains are illustrated in replicates of 12.



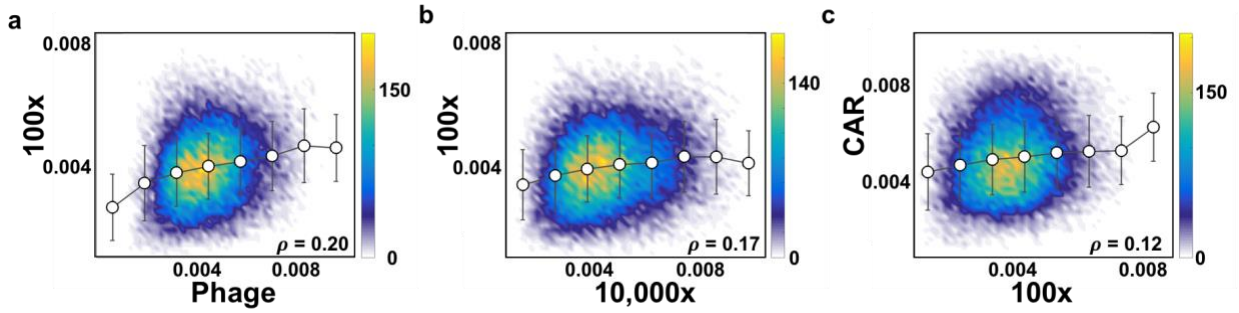
2.2 A comparison between the genetic and phenotypic landscape: Supplementary Figure 2.2-2.3, Supplementary Table 2.1-2.3

In this section, we compared the genetic and phenotypic landscapes of the clinical and environmental isolate libraries. We defined the phenotypic distance between pairs of strains as the Euclidean distance between the average of all replicates (growth dynamics) for both strains. Here, we demonstrated these correlations with two types of features: (1) time derivative of growth curves and (2) growth rate. We saw a lack of correlation between phylogenetic distance and phenotypic distance for both clinical and environmental isolate libraries **(Supplementary Table 2.1, Table 2.3, and Supplementary Figure 2.3)**. In contrast, there was a relatively stronger correlation between some phenotypes **(Supplementary Table 2.2 and Supplementary Figure 2.2)**. As a positive control, we compared the Spearman correlation between the growth dynamics of two replicates and show that the correlation is significantly stronger, as anticipated **(Supplementary Table 2.1).**

Matlab files in package (https://github.com/youlab/strain_prediction_CZ):
Supp_correlationsClinicalIsolates.m, Figure4_correlation_envIsolates.m,
Supp_correlationsClinicalIsolates_traditionalGR.m,
Figure4_correlation_envIsolates_traditionalGR.m

Supplementary Figure 2.2
An examination of the correlation between phenotypic landscapes.

We visualize additional correlations between pairs of growth conditions with a density plot and report the corresponding Spearman correlation coefficient. Here, the growth phenotype is defined by the time derivative of the growth curves.
(a) Phage treatment vs 100x dilution ($p = 2.56 \times 10^{-7}$)
(b) 10,000x dilution vs 100x dilution ($p = 2.68 \times 10^{-7}$)
(c) 100x dilution vs Carbenicillin treatment ($p = 8.48 \times 10^{-5}$)

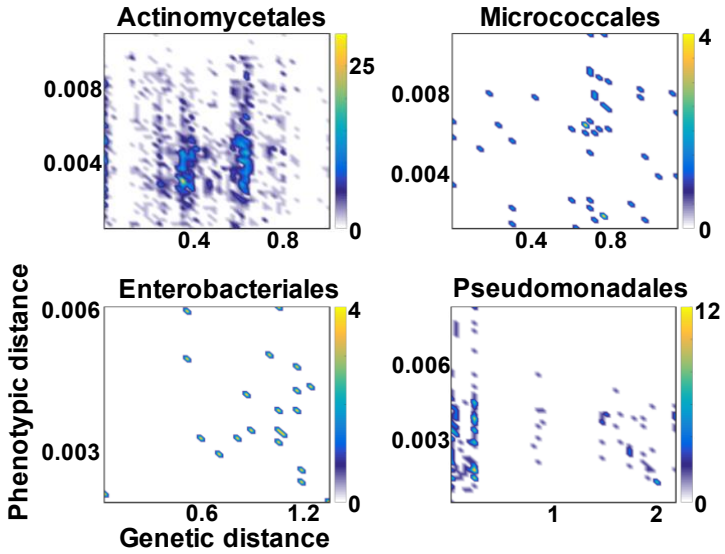


Supplementary Figure 2.3

The correlation between phenotypic and genetic landscapes of environmental isolates.

We visualize four additional correlations on the taxonomic level of order with a density plot and report the corresponding Spearman correlation coefficient (and significance). Here, the growth phenotype is defined by the time derivative of the growth curves.

Actinomycetales (ρ = 0.064, p = 0.13)
Micrococcales (ρ =0.14, p = 0.29)
Pseudomonadales (ρ = -0.03 p = 0.62)
Enterobacteriales (ρ = -0.12, p = 0.65)



Supplementary Table 2.1

Results from Mantel test describe correlation between phylogenetic distance and phenotypic distance (average of 4 replicates).[2] Here, we demonstrate these correlations with two types of features: (1) time derivative of growth curves and (2) growth rate. Additionally, we compare the use of phylogeny based on a core set of genes (MLST) and SNPs. When using either approach for strain definition, the conclusion remains the same.

Matrix 1	Matrix 2	Spearman correlation (ρ)	p-value
Phenotype is the time derivative of growth curves and using a SNP approach to phylogenetic distance.			
Phylogeny	Phenotype (10,000x)	-0.0258	0.81
Phylogeny	Phenotype (Phage)	-0.0475	0.92
Phylogeny	Phenotype (100x)	0.1242	6.62×10^{-5}
Phylogeny	Phenotype (Carbenicillin)	-0.0302	0.80
Phylogeny	Phenotype (10,000x, Phage)	-0.0491	0.92

Phylogeny	Phenotype (10,000x, 100x)	0.0397	0.13
Phylogeny	Phenotype (10,000x, Carbenicillin)	-0.0363	0.85
Phylogeny	Phenotype (Phage, 100x)	0.0383	0.14
Phylogeny	Phenotype (Phage, Carbenicillin)	-0.0411	0.86
Phylogeny	Phenotype (100x, Carbenicillin)	0.0376	0.15
Phylogeny	Phenotype (10,000x, Phage, 100x)	0.0035	0.46
Phylogeny	Phenotype (10,000x, Phage, Carbenicillin)	-0.0480	0.91
Phylogeny	Phenotype (10,000x, 100x, Carbenicillin)	0.0096	0.39
Phylogeny	Phenotype (Phage, 100x, Carbenicillin)	0.0098	0.40
Phylogeny	Phenotype (10,000x, Phage, 100x, Carbenicillin)	-0.0094	0.60
Phenotype is the time derivative of growth curves and using a MLST approach to phylogenetic distance.			
Phylogeny	Phenotype (10,000x)	-0.0206	0.77
Phylogeny	Phenotype (Phage)	-0.042	0.90
Phylogeny	Phenotype (100x)	0.12	0.001
Phylogeny	Phenotype (Carbenicillin)	-0.023	0.74
Phylogeny	Phenotype (10,000x, Phage)	-0.04	0.91
Phylogeny	Phenotype (10,000x, 100x)	0.04	0.094
Phylogeny	Phenotype (10,000x, Carbenicillin)	-0.024	0.76
Phylogeny	Phenotype (Phage, 100x)	0.041	0.12
Phylogeny	Phenotype (Phage, Carbenicillin)	-0.031	0.81
Phylogeny	Phenotype (100x, Carbenicillin)	0.041	0.12
Phylogeny	Phenotype (10,000x, Phage, 100x)	0.0086	0.40
Phylogeny	Phenotype (10,000x, Phage, Carbenicillin)	-0.036	0.85
Phylogeny	Phenotype (10,000x, 100x, Carbenicillin)	0.019	0.29
Phylogeny	Phenotype (Phage, 100x, Carbenicillin)	0.017	0.31
Phylogeny	Phenotype (10,000x, Phage, 100x, Carbenicillin)	3.8×10^{-4}	0.50
Phenotype is the growth rate time courses and using a SNP approach to phylogenetic distance.			
Phylogeny	Phenotype (10,000x)	0.039	0.1718
Phylogeny	Phenotype (Phage)	0.0084	0.4159
Phylogeny	Phenotype (100x)	-0.0066	0.5636
Phylogeny	Phenotype (Carbenicillin)	0.0366	0.1780
Phylogeny	Phenotype (10,000x, Phage)	0.0309	0.2204
Phylogeny	Phenotype (10,000x, 100x)	0.0346	0.2001
Phylogeny	Phenotype (10,000x, Carbenicillin)	0.0398	0.1652
Phylogeny	Phenotype (Phage, 100x)	0.0036	0.4608
Phylogeny	Phenotype (Phage, Carbenicillin)	0.0474	0.1138
Phylogeny	Phenotype (100x, Carbenicillin)	0.0351	0.1787
Phylogeny	Phenotype (10,000x, Phage, 100x)	0.0302	0.2359
Phylogeny	Phenotype (10,000x, Phage, Carbenicillin)	0.0498	0.1109
Phylogeny	Phenotype (10,000x, 100x, Carbenicillin)	0.0393	0.1661

Phylogeny	Phenotype (Phage, 100x, Carbenicillin)	0.0459	0.1149
Phylogeny	Phenotype (10,000x, Phage, 100x, Carbenicillin)	0.0493	0.1130
Phenotype is the growth rate time courses and using a MLST approach to phylogenetic distance.			
Phylogeny	Phenotype (10,000x)	0.017	0.33
Phylogeny	Phenotype (Phage)	0.030	0.23
Phylogeny	Phenotype (100x)	-0.0072	0.57
Phylogeny	Phenotype (Carbenicillin)	0.018	0.30
Phylogeny	Phenotype (10,000x, Phage)	0.018	0.32
Phylogeny	Phenotype (10,000x, 100x)	0.010	0.39
Phylogeny	Phenotype (10,000x, Carbenicillin)	0.014	0.35
Phylogeny	Phenotype (Phage, 100x)	0.022	0.29
Phylogeny	Phenotype (Phage, Carbenicillin)	0.034	0.18
Phylogeny	Phenotype (100x, Carbenicillin)	0.016	0.34
Phylogeny	Phenotype (10,000x, Phage, 100x)	0.015	0.34
Phylogeny	Phenotype (10,000x, Phage, Carbenicillin)	0.028	0.23
Phylogeny	Phenotype (10,000x, 100x, Carbenicillin)	0.013	0.36
Phylogeny	Phenotype (Phage, 100x, Carbenicillin)	0.031	0.20
Phylogeny	Phenotype (10,000x, Phage, 100x, Carbenicillin)	0.026	0.25

Supplementary Table 2.2

Results from Mantel test describe correlation between phenotypic distance between pairs of growth conditions (average of 4 replicates).[2] A positive control is highlighted in italics to illustrate the expected high correlation between replicates. Here, we demonstrate these correlations with two types of features: (1) time derivative of growth curves and (2) growth rate.

Matrix 1	Matrix 2	Spearman correlation (ρ)	p-value
Phenotype is the time derivative of growth curves.			
Phenotype (10,000x)	Phenotype (Phage)	0.2854	3.16×10^{-7}
Phenotype (10,000x)	Phenotype (100x)	0.1666	2.68×10^{-7}
Phenotype (10,000x)	Phenotype (Carbenicillin)	0.2462	2.30×10^{-7}
Phenotype (Phage)	Phenotype (100x)	0.1972	2.56×10^{-7}
Phenotype (Phage)	Phenotype (Carbenicillin)	0.2360	5.70×10^{-7}
Phenotype (100x)	Phenotype (Carbenicillin)	0.1180	8.48×10^{-5}
Phenotype (4 conditions) – replicate 1	*Phenotype (4 conditions) – replicate 2*	*0.9255*	*3.85×10^{-8}*
Phenotype is the growth rate time courses.			
Phenotype (10,000x)	Phenotype (Phage)	-0.1084	0.9877
Phenotype (10,000x)	Phenotype (100x)	0.0638	0.0729

Phenotype (10,000x)	Phenotype (Carbenicillin)	0.1079	$9x10^{-3}$
Phenotype (Phage)	Phenotype (100x)	0.1449	$1.4x10^{-3}$
Phenotype (Phage)	Phenotype (Carbenicillin)	-0.0595	0.9026
Phenotype (100x)	Phenotype (Carbenicillin)	0.1318	$1x10^{-3}$

Supplementary Table 2.3

Results from Mantel test describe correlation between phylogenetic distance and phenotypic distance (average of 12 replicates) for the environmental isolates. Here, we demonstrate these correlations with two types of features: (1) time derivative of growth curves and (2) growth rate.

Matrix 1 Phylogeny	**Matrix 2 Phenotype**	**Spearman correlation (ρ)**	**p-value**
Phenotype is the time derivative of growth curves.			
Actinomycetales	10,000x	0.064	0.13
Micrococcales	10,000x	0.14	0.29
Bacillales	10,000x	0.11	$2.7x10^{-7}$
Pseudomonadales	10,000x	-0.03	0.62
Enterobacteriales	10,000x	-0.12	0.65
Lactobacillales	10,000x	0.31	0.26
Streptomycetales	10,000x	0	0.50
Phenotype is the growth rate time courses.			
Actinomycetales	10,000x	-0.0076	0.53
Micrococcales	10,000x	0.57	0.014
Bacillales	10,000x	0.12	$2.6x10^{-7}$
Pseudomonadales	10,000x	0.09	0.10
Enterobacteriales	10,000x	-0.18	0.79
Lactobacillales	10,000x	0.31	0.32
Streptomycetales	10,000x	0	0.50

3. Whole genome sequencing of isolate libraries

The whole genome sequence of all isolates in our library are publicly available **(Methods)**.[3] Here, we compile these sequences for two purposes (1) to make a phylogenetic tree and (2) for antibiotic resistance prediction using known mutations or genes conferring resistance. We further describe the method for making the phylogenetic tree in the main text. The workflow for the derivation of the phylogenetic tree **(Supplementary Figure 3.1** for clinical isolates**)** is described in **Supplementary Figure 3.2** (clinical isolates) and **Supplementary Figure 3.3** (environmental isolates). The results of the phylogenetic trees for both libraries were used as the labels for the corresponding growth dynamics in **Supplementary Section 1**. We used three distinct approaches to predict the antimicrobial resistance for the 244 isolates using the available WGS: (1) a compilation of mechanisms from a literature search, (2) CARD, and (3) ResFinder.

3.1 Phylogenetic tree: Supplementary Figures 3.1-3.3, Supplementary Table 3.1

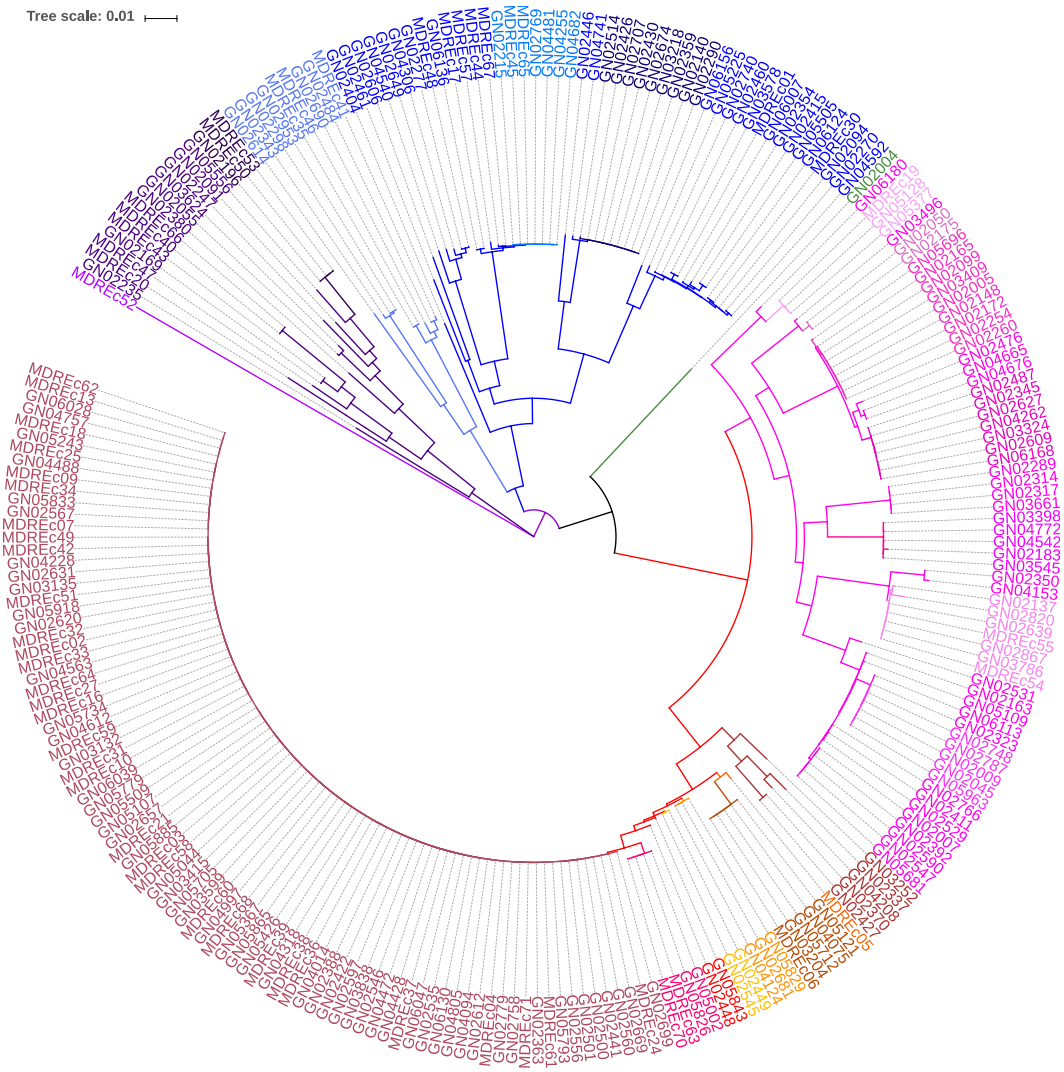
Supplementary Figure 3.1
A visualization of the phylogenetic tree.

Using the methodology described in the **Methods**, we generate a phylogenetic tree for the 244 clinical isolate library.

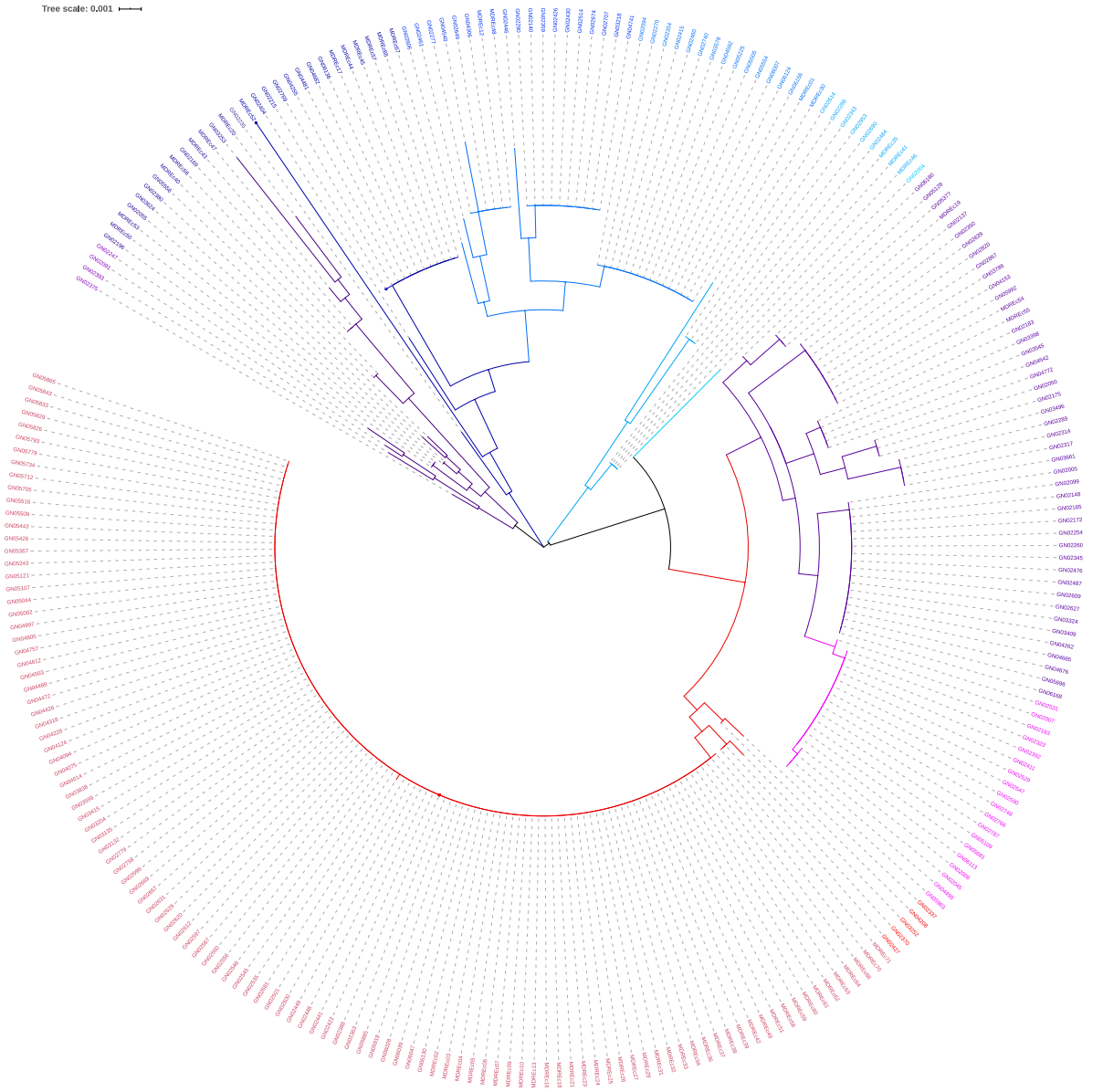
(a) With the SNP-based approach to strain definition, there are 203 unique strains in the library of 244 clinical isolates. The majority of the clusters (188 isolates) consisted of a single isolate, the largest cluster consists of 20 isolates, 12 clusters consist of sets of only two or three isolates, and 2 clusters consist of sets of four or five isolates.

(b) With the MLST-based strain definition, there are 41 unique strains in the library of 244 clinical isolates. The majority of the library resided within the largest cluster consisting of 104 isolates followed by clusters of size 19, 15, and 14. And in contrast to the SNP approach, only 21 clusters consisted of a single isolate.

a.

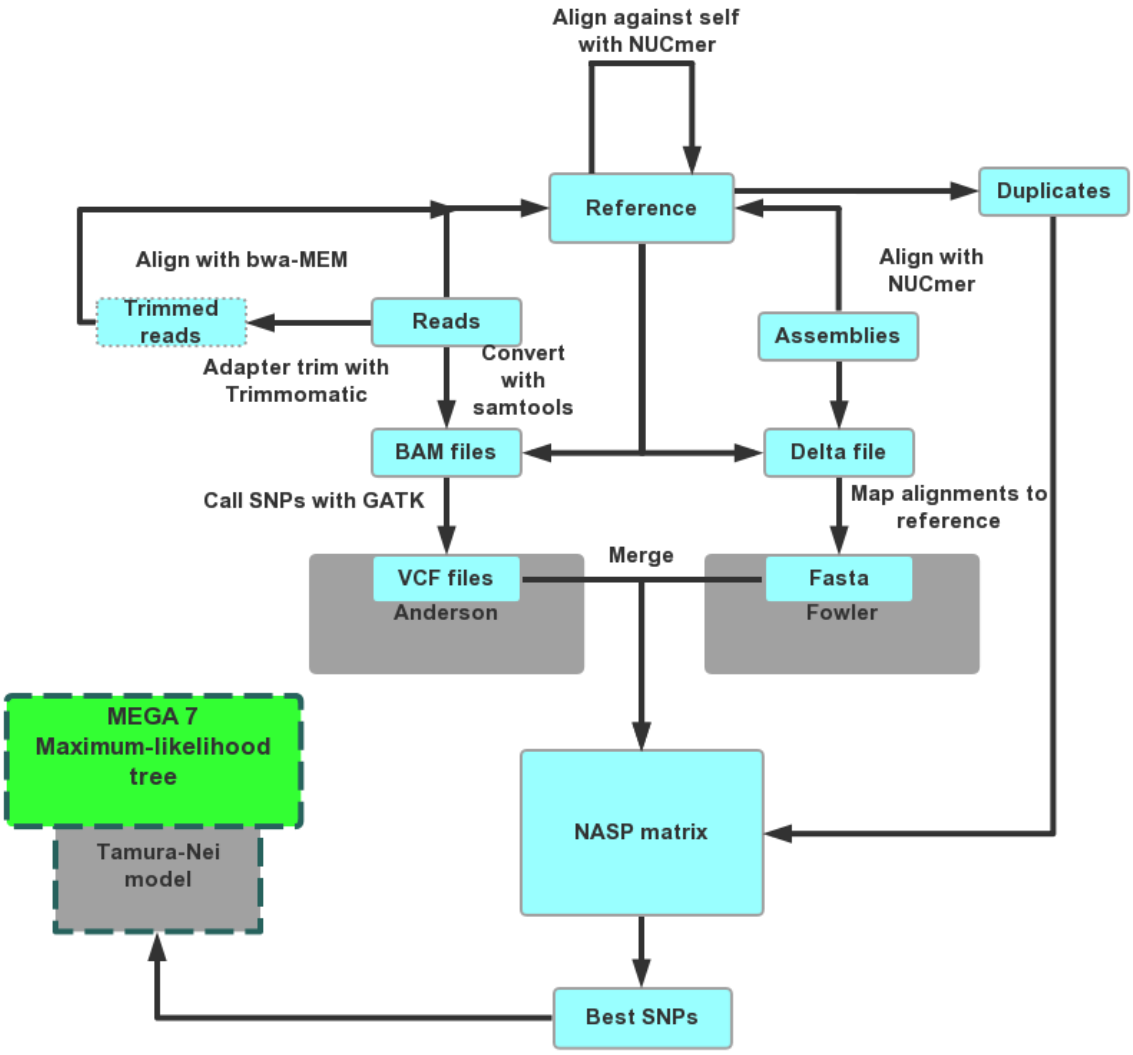


b.



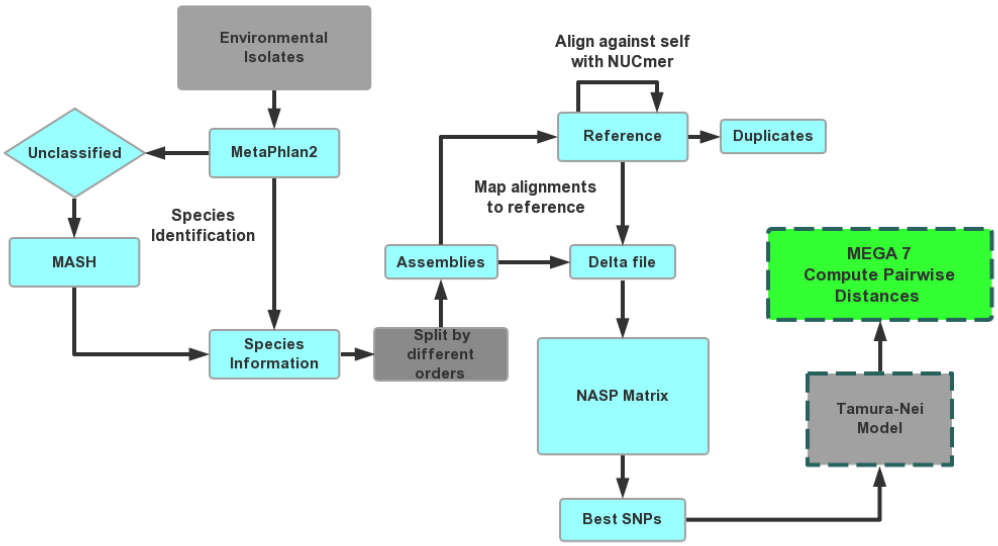
Supplementary Figure 3.2
Analysis pipeline for phylogenetic tree.

The illustration describes the methodology for generating the phylogenetic tree from the clinical isolates **(See methods for a detailed description)**.



Supplementary Figure 3.3
Analysis pipeline for phylogenetic tree.

The illustration describes the methodology for generating the phylogenetic tree from the environmental isolates **(See methods for a detailed description)**.



Supplementary Table 3.1

The order list and reference genome ID of phylogenetic analysis

Order List	Reference ID
Actinomycetales	FS151
Flavobacteriales	FS15
Sphingobacteriales	FS18
Corynebacteriales	FS511
Micrococcales	FS239
Bacillales	FS108
Lactobacillales	FS623
Rhizobiales	FS161
Rhodospirillales	FS07
Sphingomonadales	FS586
Streptomycetales	FS156
Burkholderiales	FS335
Neisseriales	FS494
Enterobacteriales	FS123
Pseudomonadales	FS106
Xanthomonadales	FS157
Unclassified	NA
Eukaryota	FS211

3.2 WGS-based antibiotic resistance prediction: Supplementary Tables 3.2-3.3

To apply the current capability of genomic sequences to predict antibiotic resistances, we used three approaches. The first was to compile a comprehensive list of known resistance genes for 4 antibiotics for *Enterobacteriaceae* based on gene summaries found in the literature as well as a pubmed search.[4-13] The second was to use the Comprehensive Antibiotic Resistance Database (CARD), and the third was to use ResFinder; both are publicly available databases containing antimicrobial resistance genes.[14,15] We compared these three methods for two purposes (1) to represent different approaches to identifying antimicrobial resistances and (2) to utilize common sources of antimicrobial resistance genes or gene products. In the case of Ampicillin-Sulbactam, Trimethoprim-Sulfamethoxazole, and Gentamicin, known genes conferring resistance were identified based on a threshold of 95% gene similarity and 50% gene length. For Ciprofloxacin, in addition to gene identification based on the described threshold, specific amino acid substitutions/mutations were searched for (**Supplementary Table 3.2**).[16,17] **Supplementary Table 3.3** describes the results based on 95% gene similarity and 50% gene length with the compiled literature search and the two curated databases, CARD and ResFinder.

Ciprofloxacin mutations of interest: We looked into mutations described in **Supplementary Table 3.2** where the mutation resulted in an amino acid change.[16,17] For an isolate to be predicted as resistant according to the WGS, we looked for one of two conditions to be fulfilled: (1) the presence of AcrA (945112) and AcrB (945108) and at least 2 mutations in gyrA (946614) and (2) the presence of AcrA (945112) and AcrB (945108) and at least 1 mutation in gyrA (946614) and at least 1 mutation in parC (947499).

MATLAB files in package (https://github.com/youlab/strain_prediction_CZ)**:** predictResistance_WGS.m (main file), resistance_card.m, resistance_literature.m, resistance_resfinder.m, resistance_roc_WGS.m (main file)

Supplementary Table 3.2

A summary of the genes/mutations and corresponding gene ID's associated with CIP resistance.

Gene ID (pubmed)	Descriptions of mutation
915402	mutations in amino acids 426-464
945108	G288D substitution
946614	mutations in amino acids 67-106
947499	mutations in amino acids 47-133
947501	mutations in amino acids 420-458

Supplementary Table 3.3

A summary of the predicted resistance profiles for the 244 isolates using a database of known resistance genes/mutations compiled from the literature and two publicly available resources, ResFinder and CARD.[4-15]

Antibiotic (class)	**Metric**	**Ampicillin-sulbactam (SAM)** (β-lactam + β-lactamase inhibitor)	**Gentamicin (GM)** (aminoglycoside)	**Trimethoprim-sulfamethoxazole (SXT)** (sulfonamide combination)	**Ciprofloxacin (CIP)** (fluoroquinolone)
Literature	**Sensitivity (TPR)**	**93.33%**	**100%**	**92.14%**	**83.56%**
	Specificity (TNR)	**42.20%**	**47.32%**	**77.88%**	**95.92%**
	Accuracy	**70.49%**	**55.74%**	**86.07%**	**88.52%**
ResFinder[15]	Sensitivity (TPR)	88.89%	94.87%	92.14%	22.60%
	Specificity (TNR)	45.87%	47.80%	78.85%	100%
	Accuracy	69.67%	55.33%	86.48%	53.69%
CARD[14]	**Sensitivity (TPR)**	**89.63%**	**100%**	**83.57%**	**100%**
	Specificity (TNR)	**44.04%**	**1.46%**	**87.50%**	**0%**
	Accuracy	**69.26%**	**17.21%**	**85.25%**	**59.84%**

References:

1 Kulp, A. J. *et al.* Genome-Wide Assessment of Outer Membrane Vesicle Production in Escherichia coli. *PLoS One* **10**, e0139200, doi:10.1371/journal.pone.0139200 (2015).

2 Glerean, E. *et al.* Reorganization of functionally connected brain subnetworks in high-functioning autism. *Hum Brain Mapp* **37**, 1066-1079, doi:10.1002/hbm.23084 (2016).

3 Kanamori, H. *et al.* Genomic Analysis of Multidrug-Resistant Escherichia coli from North Carolina Community Hospitals: Ongoing Circulation of CTX-M-Producing ST131-H30Rx and ST131-H30R1 Strains. *Antimicrob Agents Chemother* **61**, doi:10.1128/AAC.00912-17 (2017).

4 Cherif, T., Saidani, M., Decre, D., Boutiba-Ben Boubaker, I. & Arlet, G. Cooccurrence of Multiple AmpC beta-Lactamases in Escherichia coli, Klebsiella pneumoniae, and Proteus mirabilis in Tunisia. *Antimicrob Agents Chemother* **60**, 44-51, doi:10.1128/AAC.00828-15 (2016).

5 Doi, Y. & Arakawa, Y. 16S ribosomal RNA methylation: emerging resistance mechanism against aminoglycosides. *Clin Infect Dis* **45**, 88-94, doi:10.1086/518605 (2007).

6 McDermott, P. F. *et al.* Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella. *Antimicrob Agents Chemother* **60**, 5515-5520, doi:10.1128/AAC.01030-16 (2016).

7 Mendonca, N., Leitao, J., Manageiro, V., Ferreira, E. & Canica, M. Spread of extended-spectrum beta-lactamase CTX-M-producing escherichia coli clinical isolates in community and nosocomial environments in Portugal. *Antimicrob Agents Chemother* **51**, 1946-1955, doi:10.1128/AAC.01412-06 (2007).

8 Ramirez, M. S. & Tolmasky, M. E. Aminoglycoside modifying enzymes. *Drug Resist Updat* **13**, 151-171, doi:10.1016/j.drup.2010.08.003 (2010).

9 Shin, H. W. *et al.* Characterization of trimethoprim-sulfamethoxazole resistance genes and their relatedness to class 1 integron and insertion sequence common region in gram-negative bacilli. *J Microbiol Biotechnol* **25**, 137-142 (2015).

10 Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. *J Antimicrob Chemother* **68**, 2234-2244, doi:10.1093/jac/dkt180 (2013).

11 Sugumar, M., Kumar, K. M., Manoharan, A., Anbarasu, A. & Ramaiah, S. Detection of OXA-1 beta-lactamase gene of Klebsiella pneumoniae from blood stream infections (BSI) by conventional PCR and in-silico analysis to understand the mechanism of OXA mediated resistance. *PLoS One* **9**, e91800, doi:10.1371/journal.pone.0091800 (2014).

12 Tyson, G. H. *et al.* WGS accurately predicts antimicrobial resistance in Escherichia coli. *J Antimicrob Chemother* **70**, 2763-2769, doi:10.1093/jac/dkv186 (2015).

13 Waltner-Toews, R. I. *et al.* Clinical characteristics of bloodstream infections due to ampicillin-sulbactam-resistant, non-extended- spectrum-beta-lactamase-producing Escherichia coli and the role of TEM-1 hyperproduction. *Antimicrob Agents Chemother* **55**, 495-501, doi:10.1128/AAC.00797-10 (2011).

14 McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* **57**, 3348-3357, doi:10.1128/AAC.00419-13 (2013).

15 Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* **67**, 2640-2644, doi:10.1093/jac/dks261 (2012).
16 Blair, J. M. *et al.* AcrB drug-binding pocket substitution confers clinically relevant resistance and altered substrate specificity. *Proc Natl Acad Sci U S A* **112**, 3511-3516, doi:10.1073/pnas.1419939112 (2015).
17 Jacoby, G. A. Mechanisms of resistance to quinolones. *Clin Infect Dis* **41 Suppl 2**, S120-126, doi:10.1086/428052 (2005).