

Supplementary Appendix

Supplementary Methods

Whole Genome Sequencing

Sequence quality was assessed for each sample using FastQC (Version 0.11.5) and summarized with MultiQC (Version 1.7)[1]. Reads were trimmed with Trimmomatic (Version 0.36) to remove low-quality bases, using a sliding window of 4 and minimum read quality (phred) score of 15 [2]. Contigs were assembled using SPAdes (Version 3.8.0)[3]. Kraken (Version 1.1 with minikraken 8Gb database) was used to confirm species identity from the contigs. Multi-locus sequence type (MLST) was identified *in silico* through the PubMLST database using the software package MLST [4,5]. For the minority of isolates without a sequence type (ST) identified from contigs, stringMLST[6] was performed from raw reads. Genes were annotated with Prokka (Version 1.13) [7]. A common core genome for each individual dataset was identified using Roary (Version 3.12.0)[8] and clusters (groups of genetically related isolates) were identified and categorized using RhierBAPS (Version 1.1.0)[9]. Genetic distances between isolates were calculated from unassembled reads using Mash distances. Mash distance is a rapid measure of genetic distance that is based off of the minhash algorithm and can be performed from raw reads[10]. Genetic trees based on these Mash distances were generated using Mashtree (Version 0.57) [10,11]. Prediction models were developed using R (Version 3.5.2). Sequence data will be uploaded and accessible through the National Center for Biotechnology Information's Sequence Read Archive.

ST Parametric Model Approach (Lineage-based)

For the entire dataset, we developed logistic regression models to predict antibiotic susceptibility, given as susceptible (1) or non-susceptible (0). The predictor variables consisted

of binary indicator variables of the presence of specific STs. Nine of the most common STs were considered as binary variables, with the final 10th variable representing the remaining STs with the lowest frequencies. For derivation and validation datasets, the area under the curve (AUC) for the receiver operating characteristic (ROC) curves was computed with 95% confidence intervals. For internal validation of individual derivation datasets, bootstrapping (1000 samples with replacement) was performed in order to estimate optimism-corrected (adjusted) AUCs. 0.6% of sampled isolates had STs that could not be identified (either because of assembly quality or because they were not present in the PubMLST database) and were classified with the group of 'remaining' STs for categorization.

Cluster Parametric Model Approach (Lineage-based)

Similar to the ST parametric approach, we also developed logistic regression models to predict antibiotic susceptibility but using clusters as the predictors. The predictors (clusters of genetically related isolates) were identified as described above, and consisted of binary indicator variables for the presence of a specific cluster ID. We limited the maximum number of identified clusters to 10, for comparability with the *ST Parametric Model Approach*, and to ensure model fit/convergence. For derivation and validation datasets, the area under the curve (AUC) for the receiver operating characteristic (ROC) curves was computed with 95% confidence intervals. Since the clustering could only be performed within an individual dataset, and was not comparable between datasets (unlike the ST approach), we did not perform external validation, but rather performed internal validation for each dataset. Bootstrapping (1000 samples with replacement) was performed in order to estimate optimism-corrected (adjusted) AUCs for each dataset.

ST Reference Database Approach (Lineage-based)

In addition to the above *ST and Cluster Parametric Models*, we also employed a reference database approach for each dataset. This consisted of creating an antibiotic susceptibility reference database, whereby we determined the prevalence of susceptibility to each antibiotic for each sequence type for a given dataset. Creating, in essence, an ST-specific antibiogram that could be used to predict susceptibility. Because of the inherent issue of overfitting with this approach, we only validated a given dataset (Dataset 1, 3, or 1 and 3) with a separate dataset (Dataset 2). Missing STs were combined together into a single category in the reference index. This reference database approach offers the advantage of providing antibiotic susceptibility phenotype predictions for rarer STs, which collectively make up a reasonable proportion of the total observed diversity in the clinical setting.

Best Genetic Match Reference Database Approach (Nearest neighbor)

We sought to test whether antibiotic susceptibility could be best predicted using the closest genetically matched isolate within the reference database, as opposed to a higher level lineage. To do this, we created a reference database of the best genetic match for each isolate and assigned predicted antibiotic susceptibility phenotype based on the phenotype of the best genetic match (shortest Mash distance). To evaluate whether this approach performed better when the reference database contained a closely related isolate than when it did not, we reported overall results as well as results considering only the top 25%'ile and bottom 75%'ile of similarity between the query strain and the closest strain in the reference database.

Model Performance

Test characteristics including sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were calculated, and in the case of continuous predictors, the optimal ROC threshold (by Youden's J statistic) was used (pROC package, R)[12]. Since

susceptibility to ertapenem in the validation dataset (Dataset 2) was 100%, we could not calculate model performance through external validation, however we have instead provided the internally validated performance of the model (for *ST and Cluster Parametric Models* approaches only) in the supplemental material (Supplementary Table 2). Moreover, we can still use the ertapenem models for predicting antibiotic selection in a *Sequential Decision-making Model* (see below).

Summary of Test Characteristics Across Models

We generated summary figures for the post-test probabilities of susceptibility for: (1) positive model predictions indicating a likely susceptible isolate (PPV [positive predictive value]); and (2) for negative model predictions indicating a likely resistant isolate (1-NPV [Negative predictive value]), for all antibiotics and models. For antibiotics with a low pre-test probability of susceptibility (e.g., ciprofloxacin and trimethoprim-sulfamethoxazole), the goal is to obtain post-test probabilities (for model results indicating susceptible) shifted to the right (and specifically to the right of syndrome-specific thresholds considered adequate by the prescriber). This would reflect an improved ability to predict susceptibility and the opportunity to use these agents empirically. For antibiotics with a generally high pre-test probability of susceptibility (e.g. ceftriaxone, gentamicin, ertapenem), the goal is to obtain post-test probabilities (for model results indicating resistant) shifted to the left (and specifically to the left of syndrome-specific thresholds), which would reflect an improvement in our ability to predict resistance and the opportunity to prevent inadequate prescribing. We assumed that 80% and 90% prevalence of susceptibility were important thresholds for low and high morbidity syndromes respectively [17–19].

Sequential Antibiotic Decision-Making Based Upon a Prespecified Antibiotic Cascade

To simulate the antibiotic decision-making process, we explored two different scenarios where an antibiotic was selected based upon model outputs, either favouring narrow spectrum agents (Narrow spectrum cascade: ciprofloxacin → trimethoprim-sulfamethoxazole → ceftriaxone → gentamicin → ertapenem) or favouring adequacy (Adequacy cascade: ertapenem → gentamicin → ceftriaxone → trimethoprim-sulfamethoxazole → ciprofloxacin). Here the narrow spectrum cascade is ordered from narrowest to broadest spectrum, and the adequacy cascade is ordered from highest susceptibility against *E.coli* to lowest, in this dataset. The particular order of these cascades could also be tailored to desired outcome, e.g. prioritizing the use of agents that are less expensive or less likely to cause *Clostridioides difficile* infection. To simulate the sequential selection process, the first model for a given antibiotic in a cascade was run, and if susceptibility was predicted, then this antibiotic was chosen. Otherwise the next antibiotic model was queried. For the models generating a probabilistic prediction, a threshold of 80% susceptibility was used to prompt recommendation of an antibiotic[13], though this could be adjusted[14]. We employed the *ST Reference Database Approach*, the *Best Genetic Match Reference Database*, and combinations of the two models in this sequential decision-making simulation. To test whether the narrow spectrum cascade improved prescribing, a spectrum score was generated similar to previously described [14]. We created a simple ordinal score (normalized by the maximum possible value) in the following fashion: we assigned the lowest spectrum antibiotic (ciprofloxacin) to a weight of 0, then the next lowest spectrum antibiotic trimethoprim-sulfamethoxazole to a weight of 1, then the next lowest spectrum antibiotic (ceftriaxone) to a weight of 2, and so on according to the narrowest to broadest spectrum cascade. With this score, the highest possible value (corresponding to 100% of patients receiving ertapenem

therapy) is 1 and the lowest possible value is 0 (corresponding to 100% of patients receiving ciprofloxacin). This score makes the arbitrary assumption of linear increases in the magnitude of spectrum as one moves along the spectrum cascade, but is illustrative for summarizing spectrum across treatment recommendation approaches.

Supplementary Tables

Supplementary Table 1. Susceptibility breakpoints ($\mu\text{g/mL}$) by antibiotic and dataset.

	Dataset 1	Dataset 2	Dataset 3
Ciprofloxacin	1	1	1
Ceftriaxone	1	ESBL Screen	1
Gentamicin	4	4	4
Trimethoprim-Sulfamethoxazole	40	40	40
Ertapenem	0.5	0.5	0.5

Supplementary Table 2. Receiver operating characteristic area under the curve (AUC) and test characteristics of predicting susceptibility to ertapenem, using internal validation and bootstrapping to correct for optimism. (A) *ST Parametric Model Approach* and (B) *Cluster Parametric Model Approach*.

(A)

Derivation/Validation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.99/0.99 (---)	0.98 (0.97-0.99)	1 (1-1)	1 (1-1)	0.11 (0.07-0.25)
Dataset 3	0.70/0.67 (0.63-0.78)	0.69 (0.22-0.93)	0.62 (0.31-1)	0.94 (0.91-1)	0.2 (0.13-0.4)
Dataset 1 and 3	0.73/0.70 (0.66-0.8)	0.35 (0.28-0.94)	0.98 (0.4-1)	1 (0.96-1)	0.08 (0.07-0.29)

(B)

Derivation/Validation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.98/0.98 (---)	0.96 (0.94-0.98)	1 (1-1)	1 (1-1)	0.06 (0.04-0.1)
Dataset 3	0.74/0.7 (0.67-0.81)	0.81 (0.3-0.91)	0.55 (0.38-1)	0.94 (0.92-1)	0.27 (0.144-0.4)

*Bootstrapped optimism adjusted AUCs are shown for internally validated sets as AUC/Adjusted AUC. Confidence intervals could not be calculated for Set 1 AUC.

Supplementary Table 3. Receiver operating characteristic area under the curve (AUC),

and test characteristics of predicting antibiotic susceptibility of Dataset 2 using an *ST Parametric Model Approach*. Ciprofloxacin (A), ceftriaxone (B), gentamicin (C), trimethoprim-sulfamethoxazole(D).

(A)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.89 (0.84-0.94)	0.88 (0.82-0.93)	0.82 (0.72-0.91)	0.91 (0.87-0.96)	0.76 (0.67-0.85)
Dataset 2*	0.91/0.90 (0.87-0.95)	0.9 (0.84-0.95)	0.82 (0.72-0.91)	0.92 (0.87-0.96)	0.8 (0.71-0.89)
Dataset 3	0.89 (0.85-0.94)	0.88 (0.82-0.93)	0.82 (0.72-0.93)	0.91 (0.87-0.96)	0.76 (0.67-0.85)
Dataset 1 and 3	0.89 (0.85-0.94)	0.88 (0.82-0.93)	0.82 (0.72-0.91)	0.91 (0.87-0.96)	0.76 (0.68-0.85)
Combined	0.9 (0.86-0.95)	0.89 (0.83-0.94)	0.82 (0.72-0.91)	0.92 (0.87-0.96)	0.78 (0.69-0.88)

(B)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.77 (0.67-0.86)	0.73 (0.42-0.85)	0.82 (0.59-1)	0.97 (0.93-1)	0.25 (0.19-0.4)
Dataset 2*	0.8/0.78 (0.73-0.88)	0.73 (0.39-0.87)	0.91 (0.59-1)	0.98 (0.93-1)	0.27 (0.19-0.44)
Dataset 3	0.78 (0.7-0.85)	0.54 (0.37-0.84)	0.95 (0.59-1)	0.98 (0.93-1)	0.23 (0.18-0.4)
Dataset 1 and 3	0.78 (0.7-0.85)	0.54 (0.36-0.85)	0.91 (0.59-1)	0.98 (0.93-1)	0.23 (0.18-0.4)
Combined	0.78 (0.7-0.85)	0.54 (0.35-0.84)	0.91 (0.59-1)	0.98 (0.93-1)	0.23 (0.18-0.39)

(C)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.7 (0.57-0.82)	0.72 (0.65-0.83)	0.71 (0.52-0.9)	0.95 (0.92-0.98)	0.26 (0.19-0.35)
Dataset 2*	0.75/0.69 (0.65-0.85)	0.73 (0.38-0.85)	0.71 (0.52-0.95)	0.95 (0.92-0.99)	0.27 (0.17-0.38)
Dataset 3	0.68 (0.56-0.8)	0.81 (0.32-0.88)	0.62 (0.38-1)	0.94 (0.91-1)	0.25 (0.15-0.4)
Dataset 1 and 3	0.72 (0.6-0.83)	0.72 (0.63-0.85)	0.71 (0.48-0.9)	0.95 (0.92-0.98)	0.26 (0.19-0.38)
Combined	0.72 (0.6-0.83)	0.72 (0.62-0.85)	0.71 (0.48-0.9)	0.95 (0.92-0.98)	0.26 (0.18-0.36)

(D)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.6 (0.51-0.69)	0.6 (0.4-0.77)	0.62 (0.4-0.85)	0.79 (0.74-0.87)	0.4 (0.33-0.49)
Dataset 2*	0.73/0.68 (0.65-0.81)	0.69 (0.52-0.82)	0.67 (0.46-0.85)	0.84 (0.78-0.9)	0.47 (0.38-0.59)
Dataset 3	0.61 (0.52-0.7)	0.86 (0.18-0.97)	0.38 (0.2-1)	0.77 (0.74-1)	0.53 (0.33-0.79)
Dataset 1 and 3	0.67 (0.59-0.76)	0.68 (0.36-0.78)	0.65 (0.5-0.9)	0.83 (0.77-0.9)	0.46 (0.36-0.55)
Combined	0.7 (0.62-0.78)	0.67 (0.31-0.83)	0.65 (0.48-0.92)	0.83 (0.78-0.92)	0.46 (0.35-0.56)

*Bootstrapped optimism adjusted AUCs are shown for internally validated sets as AUC/Adjusted AUC.

Supplementary Table 4. Internally validated receiver operating characteristic area under the curve (AUC), and test characteristics of predicting antibiotic susceptibility (internally) using a *Cluster Parametric Model Approach*. Ciprofloxacin (A), ceftriaxone (B), gentamicin (C), trimethoprim-sulfamethoxazole (D).

(A)

Derivation/Validation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.89/0.88 (0.86-0.92)	0.79 (0.65-0.87)	0.89 (0.79-0.99)	0.995 (0.91-0.99)	0.62 (0.51-0.71)
Dataset 2	0.9/0.89 (0.85-0.94)	0.81 (0.63-0.88)	0.88 (0.77-1)	0.93 (0.89-1)	0.68 (0.55-0.78)
Dataset 3	0.76/0.73 (0.71-0.82)	0.54 (0.43-0.69)	0.94 (0.76-0.97)	0.78 (0.56-0.89)	0.82 (0.79-0.85)

(B)

Derivation/Validation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.82/0.80 (0.77-0.87)	0.6 (0.65-0.75)	0.91 (0.8-0.98)	0.98 (0.96-0.99)	0.28 (0.25-0.33)
Dataset 2	0.77/0.73 (0.69-0.86)	0.62 (0.48-0.81)	0.86 (0.68-1)	0.97 (0.94-1)	0.25 (0.2-0.33)
Dataset 3	0.69/0.67 (0.64-0.73)	0.47 (0.39-0.55)	0.84 (0.76-0.92)	0.83 (0.77-0.89)	0.49 (0.46-0.53)

(C)

Derivation/Validation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.77/0.74 (0.71-0.83)	0.73 (0.55-0.79)	0.75 (0.64-0.89)	0.95 (0.93-0.97)	0.3 (0.23-0.36)
Dataset 2	0.76/0.7 (0.67-0.84)	0.62 (0.39-0.76)	0.86 (0.67-1)	0.97 (0.94-1)	0.22 (0.17-0.31)
Dataset 3	0.66/0.63 (0.6-0.71)	0.56 (0.44-0.66)	0.73 (0.62-0.84)	0.76 (0.7-0.82)	0.52 (0.48-0.57)

(D)

Derivation/Validation Set	AUC	Sensitivity	Specificity	PPV	NPV
Dataset 1	0.75/0.73 (0.7-0.8)	0.7 (0.63-0.76)	0.72 (0.64-0.82)	0.86 (0.83-0.9)	0.5 (0.44-0.55)
Dataset 2	0.67/0.62 (0.59-0.76)	0.59 (0.31-0.89)	0.73 (0.38-0.94)	0.84 (0.77-0.93)	0.4 (0.35-0.56)
Dataset 3	0.65/0.61 (0.6-0.71)	0.9 (0.81-0.97)	0.37 (0.26-0.46)	0.27 (0.25-0.3)	0.93 (0.89-0.98)

*Bootstrapped optimism adjusted AUCs are shown as AUC/Adjusted AUC.

Supplementary Table 5. Receiver operating characteristic area under the curve (AUC), and test characteristics of predicting antibiotic susceptibility of isolates in Set 2 using an *ST Reference Database Approach*. Ciprofloxacin (A), ceftriaxone (B), gentamicin (C), trimethoprim-sulfamethoxazole (D). N missing indicates the number of samples in the validation set that could not be matched to an ST in the reference database.

(A)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	N MISSING
Dataset 1	0.9 (0.84-0.95)	0.83 (0.74-0.9)	0.96 (0.9-1)	0.97 (0.93-1)	0.76 (0.68-0.84)	37
Dataset 3	0.85 (0.79-0.91)	0.74 (0.64-0.83)	0.96 (0.91-1)	0.97 (0.93-1)	0.69 (0.62-0.77)	30
Dataset 1 and 3	0.87 (0.82-0.93)	0.77 (0.78-0.86)	0.96 (0.91-1)	0.98 (0.94-1)	0.7 (0.63-0.79)	23
Combined	0.95 (0.92-0.98)	0.89 (0.83-0.95)	0.95 (0.89-1)	0.97 (0.94-1)	0.81 (0.73-0.9)	0

(B)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	N MISSING
Dataset 1	0.77 (0.66-0.88)	0.71 (0.64-0.79)	0.89 (0.72-1)	0.98 (0.95-1)	0.31 (0.25-0.39)	37
Dataset 3	0.67 (0.56-0.78)	0.64 (0.48-0.74)	0.86 (0.67-0.95)	0.96 (0.91-0.99)	0.28 (0.21-0.35)	30
Dataset 1 and 3	0.77 (0.7-0.94)	0.65 (0.53-0.74)	0.95 (0.9-1)	0.99 (0.98-1)	0.3 (0.25-0.37)	23
Combined	0.85 (0.8-0.91)	0.74 (0.61-0.8)	0.95 (0.91-1)	0.99 (0.98-1)	0.34 (0.27-0.41)	0

(C)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	N MISSING
Dataset 1	0.76 (0.66-0.86)	0.64 (0.53-0.8)	0.88 (0.65-1)	0.97 (0.94-1)	0.25 (0.2-0.35)	37
Dataset 3	0.73 (0.63-0.83)	0.57 (0.47-0.78)	0.89 (0.63-1)	0.97 (0.93-1)	0.24 (0.19-0.34)	30
Dataset 1 and 3	0.75 (0.65-0.85)	0.62 (0.49-0.79)	0.84 (0.63-1)	0.97 (0.93-1)	0.25 (0.19-0.35)	23
Combined	0.83 (0.75-0.91)	0.7 (0.6-0.83)	0.9 (0.71-1)	0.98 (0.95-1)	0.29 (0.23-0.39)	0

(D)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	N MISSING
Dataset 1	0.56 (0.46-0.66)	0.39 (0.22-0.63)	0.79 (0.55-0.95)	0.82 (0.73-0.94)	0.37 (0.31-0.44)	37
Dataset 3	0.61 (0.51-0.7)	0.33 (0.21-0.82)	0.91 (0.43-1)	0.88 (0.73-1)	0.4 (0.36-0.55)	30
Dataset 1 and 3	0.63 (0.54-0.72)	0.49 (0.38-0.68)	0.85 (0.64-0.94)	0.88 (0.78-0.95)	0.42 (0.36-0.49)	23
Combined	0.8 (0.74-0.87)	0.65 (0.53-0.86)	0.85 (0.61-0.94)	0.91 (0.84-0.97)	0.5 (0.43-0.64)	0

Supplementary Table 6. Receiver operating characteristic area under the curve (AUC), and test characteristics of predicting antibiotic susceptibility of Set 2 using a *Best Genetic Match Reference Database* (shortest Mash distance). Models were also stratified by the quartile with the shortest mash distances (Top). Ciprofloxacin (A), ceftriaxone (B), gentamicin (C), trimethoprim-sulfamethoxazole (D). ‘Sens prevalence’ corresponds to the prevalence of susceptibility to the given antibiotic for a specific subset.

(A)

All Isolates

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.91 (0.87-0.96)	0.95 (0.91-0.98)	0.88 (0.79-0.95)	0.94 (0.9-0.98)	0.89 (0.82-0.96)	0.63
Dataset 3	0.83 (0.89-0.88)	0.68 (0.6-0.77)	0.98 (0.95-1)	0.99 (0.96-1)	0.6 (0.54-0.66)	0.63
Dataset 1 and 3	0.92 (0.88-0.96)	0.88 (0.82-0.93)	0.96 (0.91-1)	0.98 (0.95-1)	0.79 (0.71-0.87)	0.63

25% Top Match (Shortest Mash distances)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.97 (0.92-1)	1 (1-1)	0.95 (0.84-1)	0.96 (0.89-1)	1 (1-1)	0.55
Dataset 3	0.85 (0.74-0.96)	0.71 (0.47-0.94)	1 (1-1)	1 (1-1)	0.84 (0.75-0.96)	0.36
Dataset 1 and 3	0.94 (0.88-1)	0.88 (0.77-1)	1 (1-1)	1 (1-1)	0.86 (0.75-1)	0.56

75% Bottom Match (Longest Mash distances)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.89 (0.82-0.95)	0.94 (0.88-0.98)	0.84 (0.73-0.95)	0.94 (0.89-0.98)	0.84 (0.73-0.94)	0.66
Dataset 3	0.83 (0.77-0.88)	0.69 (0.6-0.77)	0.97 (0.9-1)	0.99 (0.96-1)	0.48 (0.41-0.56)	0.72
Dataset 1 and 3	0.91 (0.86-0.96)	0.87 (0.8-0.94)	0.95 (0.87-1)	0.98 (0.94-1)	0.75 (0.65-0.86)	0.65

(B)

All Isolates

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.58 (0.49-0.67)	0.94 (0.89-0.97)	0.23 (0.05-0.4)	0.9 (0.88-0.92)	0.33 (0.12-0.6)	0.88
Dataset 3	0.68 (0.57-0.79)	0.82 (0.75-0.88)	0.55 (0.36-0.73)	0.93 (0.9-0.96)	0.3 (0.19-0.41)	0.88
Dataset 1 and 3	0.72 (0.61-0.83)	0.89 (0.84-0.94)	0.55 (0.32-0.73)	0.93 (0.9-0.96)	0.42 (0.27-0.57)	0.88

25% Top Match (Shortest Mash distances)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.57 (0.4-0.74)	0.97 (0.92-1)	0.17 (0-0.5)	0.88 (0.85-0.93)	0.5 (0-1)	0.86
Dataset 3	0.81 (0.67-0.96)	0.82 (0.68-0.94)	0.8 (0.5-1)	0.94 (0.85-1)	0.57 (0.39-0.8)	0.77
Dataset 1 and 3	0.88 (0.71-1)	0.92 (0.82-1)	0.83 (0.5-1)	0.97 (0.92-1)	0.63 (0.42-1)	0.86

75% Bottom Match (Longest Mash distances)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.59 (0.48-0.71)	0.92 (0.87-0.97)	0.27 (0.07-0.53)	0.91 (0.88-0.94)	0.31 (0.09-0.56)	0.89
Dataset 3	0.58 (0.43-0.72)	0.82 (0.74-0.88)	0.33 (0.08-0.58)	0.93 (0.9-0.95)	0.15 (0.04-0.28)	0.91
Dataset 1 and 3	0.66 (0.53-0.79)	0.88 (0.82-0.93)	0.44 (0.19-0.69)	0.92 (0.89-0.95)	0.33 (0.17-0.5)	0.88

(C)

All Isolates

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.65 (0.54-0.76)	0.87 (0.82-0.92)	0.43 (0.24-0.62)	0.92 (0.89-0.95)	0.31 (0.18-0.45)	0.88
Dataset 3	0.65 (0.54-0.77)	0.78 (0.72-0.85)	0.52 (0.29-0.72)	0.93 (0.89-0.96)	0.25 (0.15-0.35)	0.88
Dataset 1 and 3	0.66 (0.55-0.78)	0.85 (0.79-0.91)	0.48 (0.29-0.67)	0.92 (0.9-0.95)	0.31 (0.19-0.44)	0.88

25% Top Match (Shortest Mash distances)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.86 (0.66-1)	0.92 (0.85-1)	0.8 (0.4-1)	0.97 (0.92-1)	0.57 (0.33-1)	0.89
Dataset 3	0.76 (0.58-0.94)	0.68 (0.53-0.82)	0.83 (0.5-1)	0.96 (0.89-1)	0.29 (0.18-0.46)	0.88
Dataset 1 and 3	0.82 (0.62-1)	0.85 (0.74-0.95)	0.8 (0.4-1)	0.97 (0.91-1)	0.4 (0.22-0.71)	0.89

75% Bottom Match (Longest Mash distances)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.59 (0.47-0.71)	0.86 (0.79-0.92)	0.31 (0.06-0.56)	0.9 (0.87-0.93)	0.24 (0.07-0.42)	0.87
Dataset 3	0.61 (0.47-0.74)	0.81 (0.74-0.88)	0.4 (0.2-0.6)	0.91 (0.88-0.94)	0.21 (0.09-0.34)	0.88
Dataset 1 and 3	0.61 (0.49-0.74)	0.85 (0.78-0.91)	0.38 (0.19-0.63)	0.91 (0.88-0.94)	0.26 (0.13-0.41)	0.87

(D)

All Isolates

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.63 (0.56-0.71)	0.77 (0.7-0.84)	0.5 (0.37-0.63)	0.79 (0.74-0.84)	0.47 (0.37-0.58)	0.71
Dataset 3	0.53 (0.46-0.6)	0.29 (0.21-0.37)	0.77 (0.66-0.88)	0.75 (0.64-0.86)	0.31 (0.27-0.35)	0.71
Dataset 1 and 3	0.61 (0.53-0.69)	0.59 (0.5-0.67)	63 (0.5-0.75)	0.79 (0.73-0.85)	0.39 (0.32-0.46)	0.71

25% Top Match (Shortest Mash distance)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.57 (0.39-0.76)	0.65 (0.5-0.79)	0.5 (0.2-0.8)	0.81 (0.71-0.92)	0.29 (0.14-0.47)	0.77
Dataset 3	0.54 (0.43-0.66)	0.21 (0.07-0.39)	0.88 (0.69-1)	0.77 (0.43-1)	0.39 (0.32-0.46)	0.64
Dataset 1 and 3	0.73 (0.6-0.86)	0.61 (0.45-0.77)	0.85 (0.61-1)	0.8 (0.79-1)	0.48 (0.37-0.61)	0.64

75% Bottom Match (Longest Mash distance)

Derivation Set	AUC	Sensitivity	Specificity	PPV	NPV	Sens Prevalence
Dataset 1	0.66 (0.57-0.75)	0.82 (0.74-0.9)	0.5 (0.36-0.67)	0.78 (0.73-0.84)	0.57 (0.44-0.71)	0.7
Dataset 3	0.51 (0.42-0.6)	0.31 (0.22-0.4)	0.71 (0.54-0.86)	0.75 (0.63-0.86)	0.27 (0.22-0.32)	0.73
Dataset 1 and 3	0.58 (0.48-0.67)	0.59 (0.49-0.69)	0.56 (0.41-0.72)	0.76 (0.69-0.84)	0.37 (0.28-0.45)	0.7

Supplementary Table 7. Proportion of antimicrobial use by antibiotic, adequacy, oral administration option, and spectrum score, for Dataset 2.

Model	CIP	SXT	CRO	GEN	ETP	Adequate	Oral	Spectrum Score	Missing Recommendation (%)
Ciprofloxacin Empiric	1	0	0	0	0	0.63	1	0	0
TMP/SMX Empiric	0	1	0	0	0	0.71	1	0.25	0
Ceftriaxone Empiric	0	0	1	0	0	0.88	0	0.5	0
Gentamicin Empiric	0	0	0	1	0	0.88	0	0.75	0
Ertapenem Empiric	0	0	0	0	1	1	0	1	0
Narrow Spectrum Cascade*									
ST Reference Database	0.47	0	0.17	0	0.36	0.97	0.47	0.445	0.13
Best Match - ALL	0.6	0.1	0.2	0.07	0.03	0.89	0.7	0.2075	0
Best Match - Top***	0.51	0.04	0.27	0.16	0.02	0.98	0.55	0.285	0.75
Combined Best Match/ST REF	0.49	0.01	0.17	0.05	0.28	0.97	0.5	0.405	0.12
Adequacy Cascade**									
ST Reference Database	0.05	0	0	0	0.95	0.99	0.05	0.95	0.13
Best Match - ALL	0.01	0	0.01	0.01	0.97	0.98	0.01	0.9825	0
Best Match - Top***	0	0	0	0	1	1	0	1	0.75
Combined Best Match/ST REF	0.05	0	0	0	0.95	0.99	0.05	0.95	0.12

*Narrow spectrum cascade: Cipro->Septra->Ceftriaxone->Gentamicin

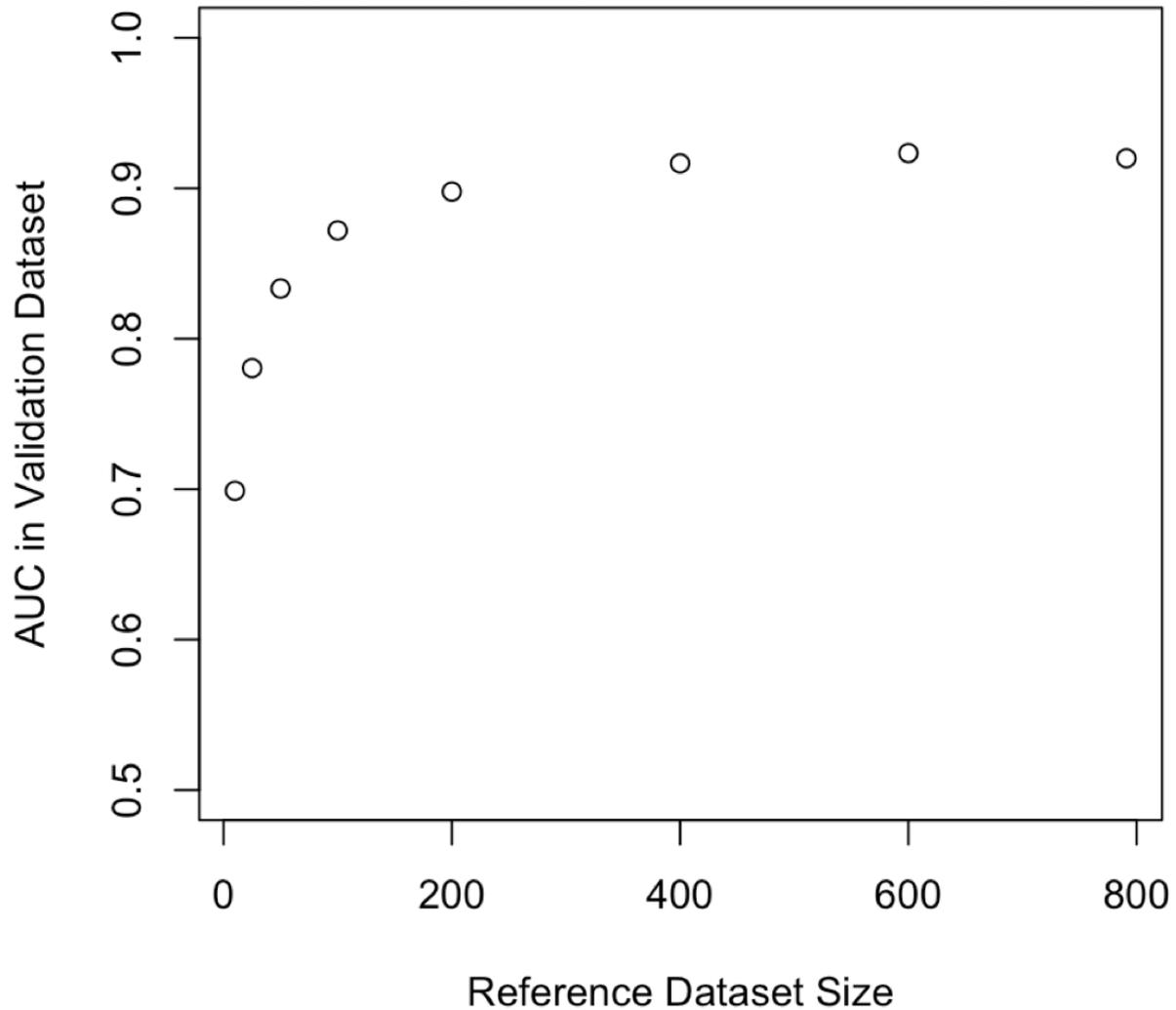
**Adequacy cascade: Ertapenem->Gentamicin->Ceftriaxone->Septra->Ciprofloxacin

***TOP 25%'tile of Best Match, and 80% treatment threshold used for continuous predictors.

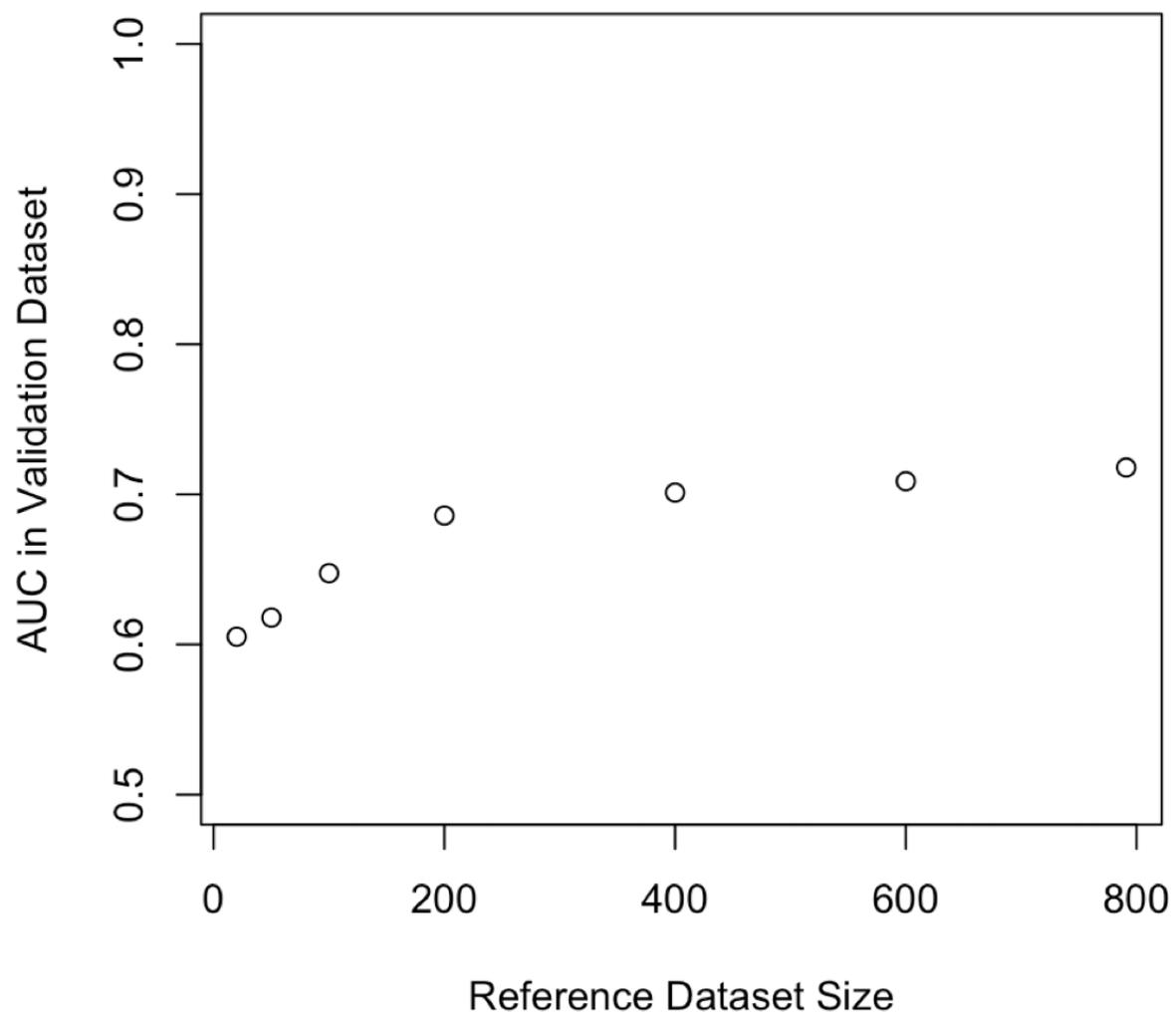
Supplementary Figures

Supplementary Figure 1. Average receiver operating characteristic area under the curve (AUC), from 250 bootstraps, as a function of randomly selected without replacement derivation databases of varying sizes tested in Dataset 2. Ciprofloxacin (A), ceftriaxone (B), gentamicin (C), trimethoprim-sulfamethoxazole (D).

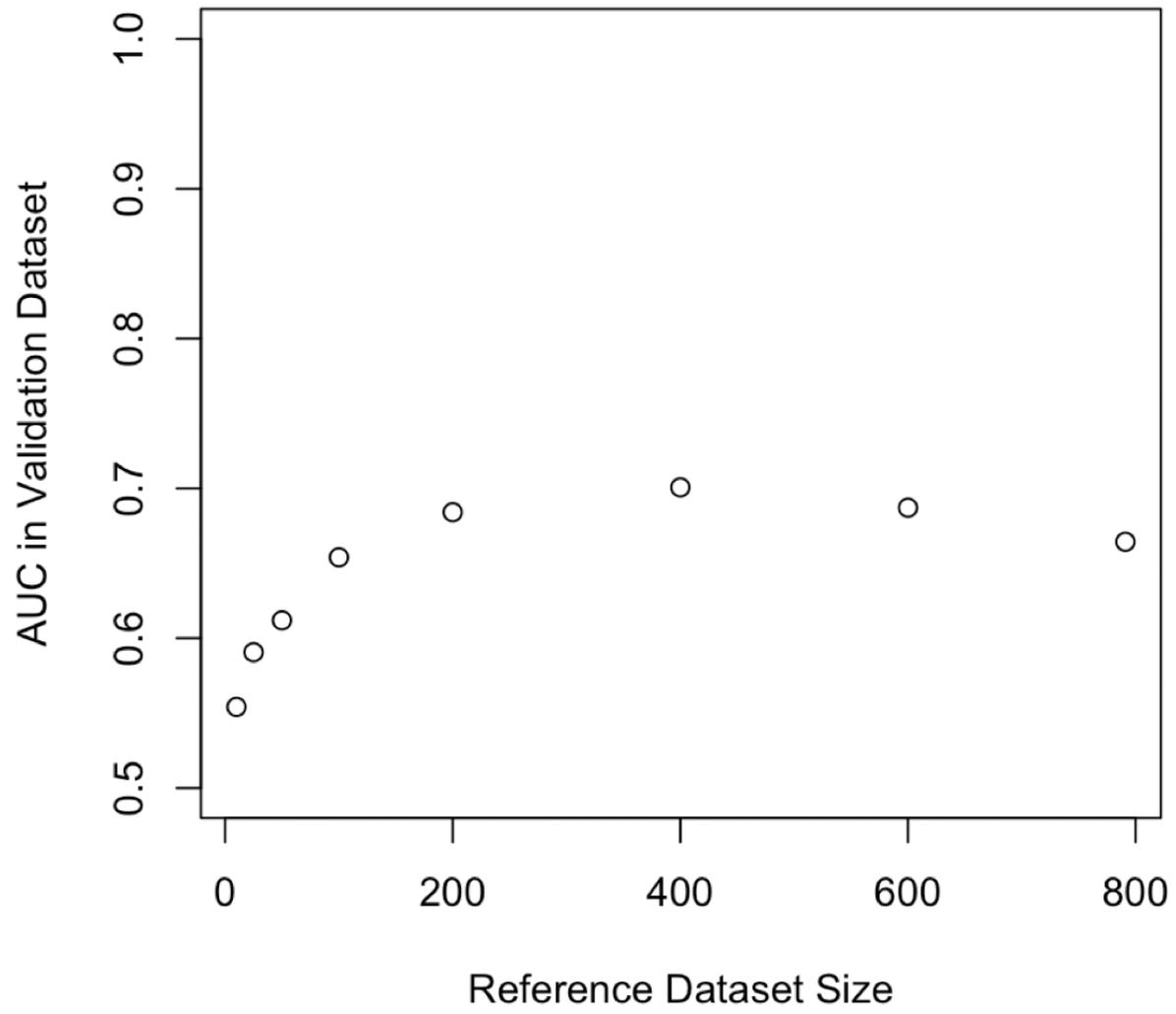
(A)



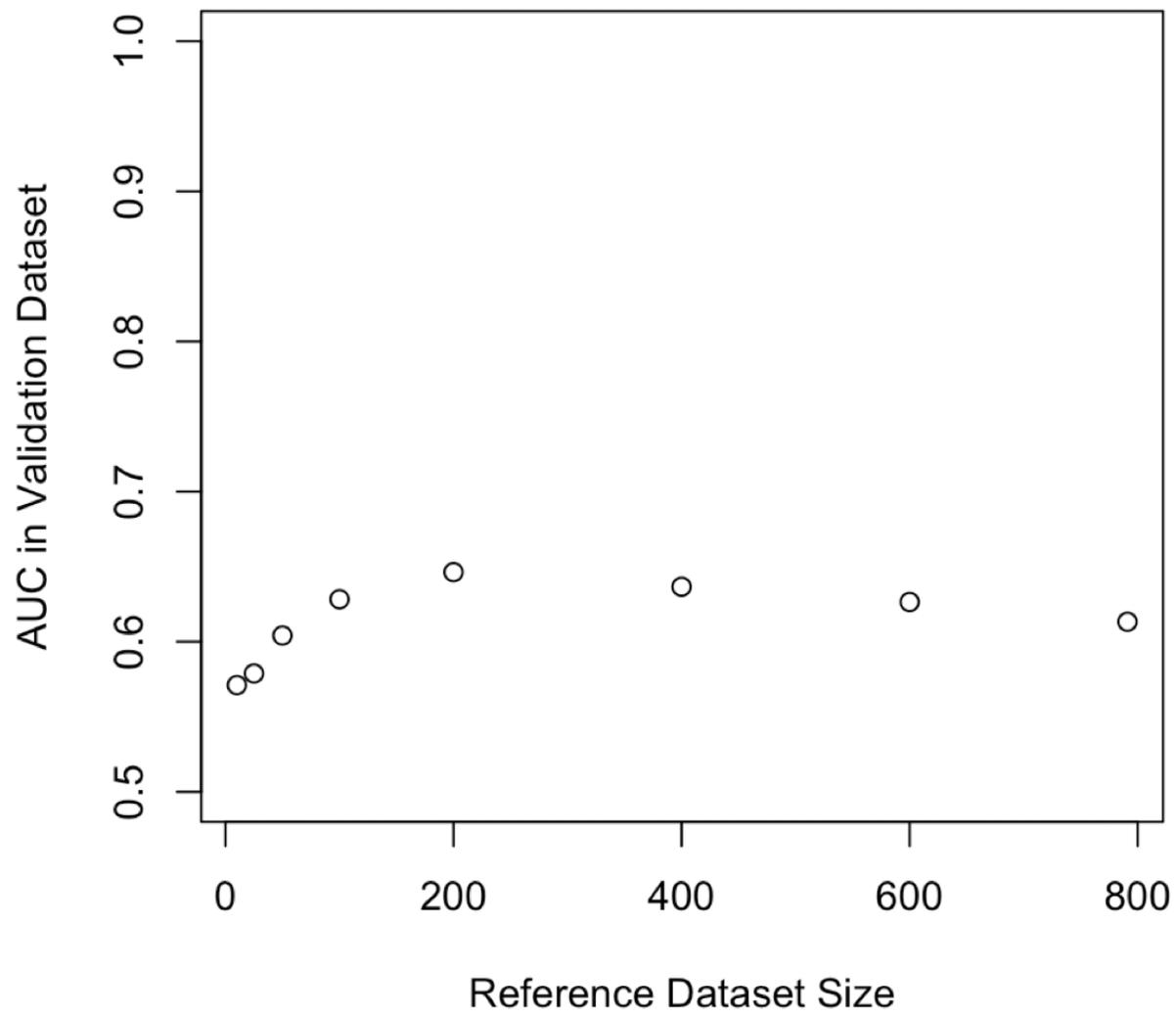
(B)



(C)



(D)



Supplementary References

1. Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016; 32:3047–3048. Available at: <http://dx.doi.org/10.1093/bioinformatics/btw354>.
2. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**; 30:2114–2120.
3. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **2012**; 19:455–477.
4. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **2010**; 11:595.
5. Seemann T. MLST. Available at: <https://github.com/tseemann/mlst>. Accessed 13 August 2019.
6. Gupta A, Jordan IK, Rishishwar L. stringMLST: a fast k-mer based tool for multilocus sequence typing. *Bioinformatics* **2017**; 33:119–121.
7. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **2014**; 30:2068–2069.
8. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**; 31:3691–3693.
9. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Research. 2018; 3:93. Available at: <http://dx.doi.org/10.12688/wellcomeopenres.14694.1>.
10. Ondov BD, Treangen TJ, Melsted P, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* **2016**; 17:132.
11. Katz L. Mashtree. Available at: <https://github.com/lkatz/mashtree>. Accessed 13 August 2019.
12. Robin X. Package 'pROC'. Available at: <https://cran.r-project.org/web/packages/pROC/pROC.pdf>. Accessed 14 August 2019.
13. Cressman AM, MacFadden DR, Verma AA, Razak F, Daneman N. Empiric Antibiotic Treatment Thresholds for Serious Bacterial Infections: A Scenario-Based Survey Study. *Clin Infect Dis* **2018**; Available at: <http://dx.doi.org/10.1093/cid/ciy1031>.
14. MacFadden DR, Coburn B, Shah N, et al. Decision-support models for empiric antibiotic selection in Gram-negative bloodstream infections. *Clin Microbiol Infect* **2019**; 25:108.e1–108.e7.