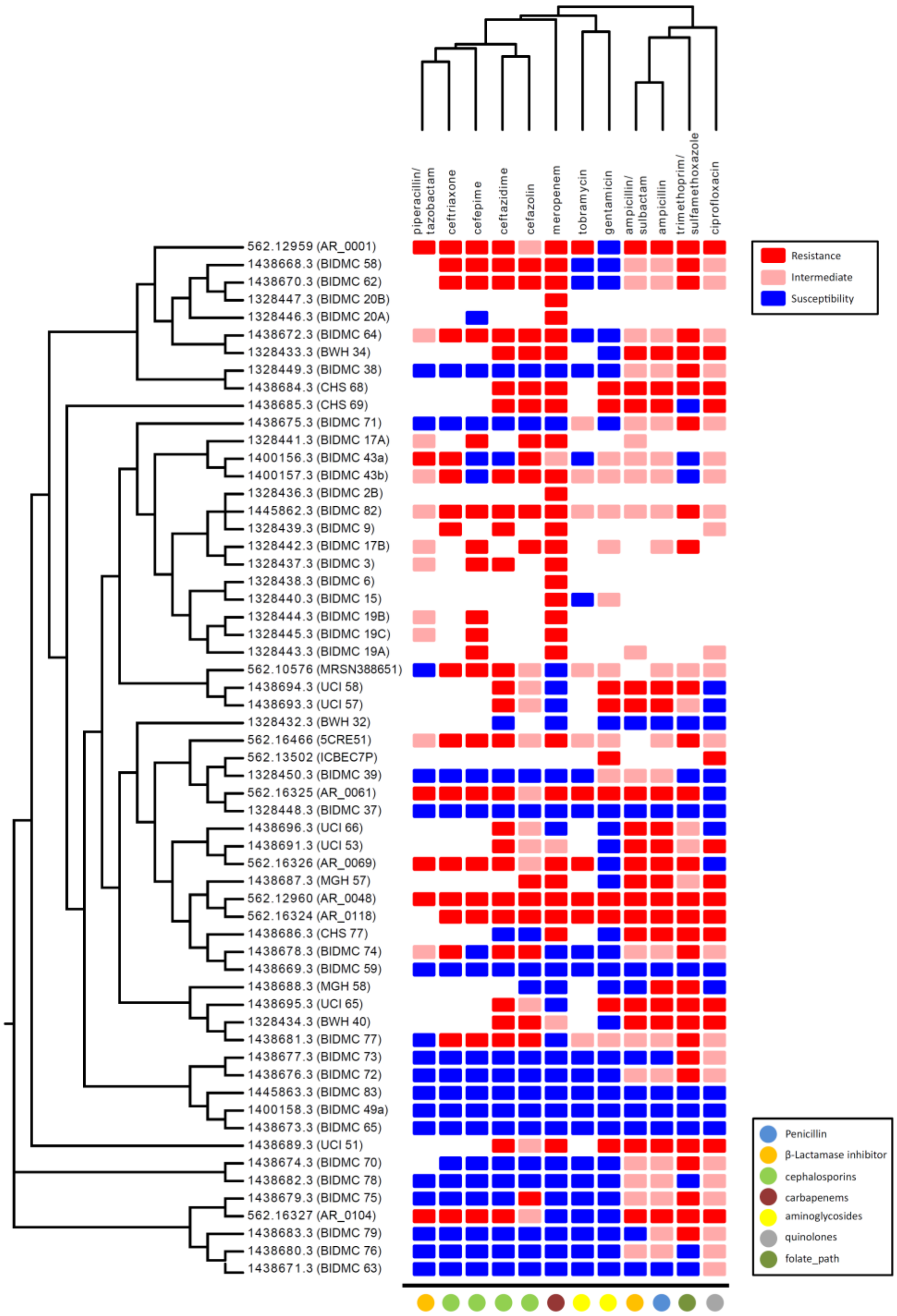
# A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the *Escherichia coli* strains

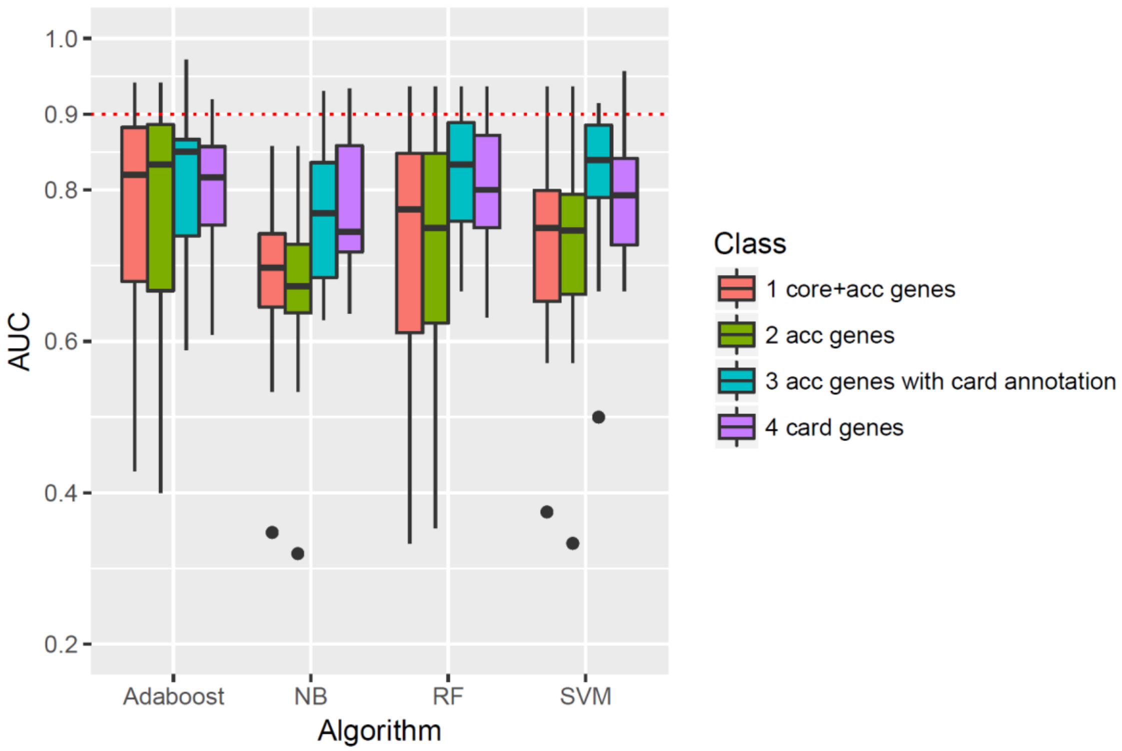
Hsuan-Lin Her and Yu-Wei Wu

This supplementary information includes:

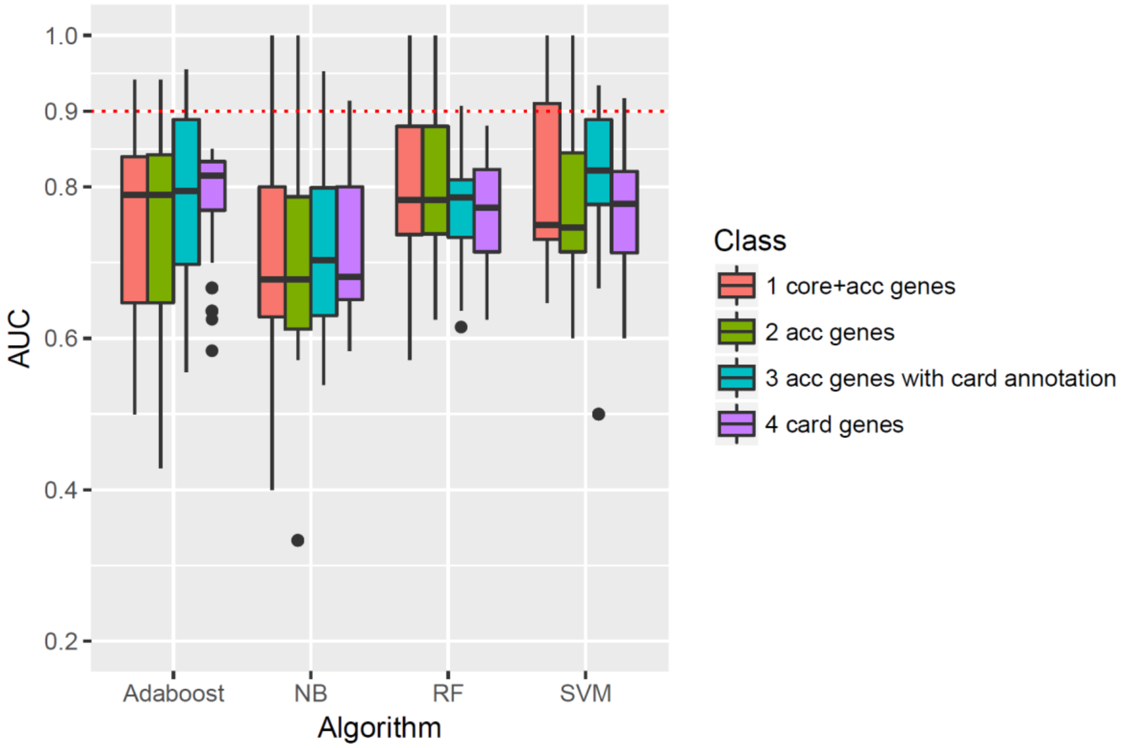
- Supplementary Figure S1-S7
- Supplementary Table S1-S13



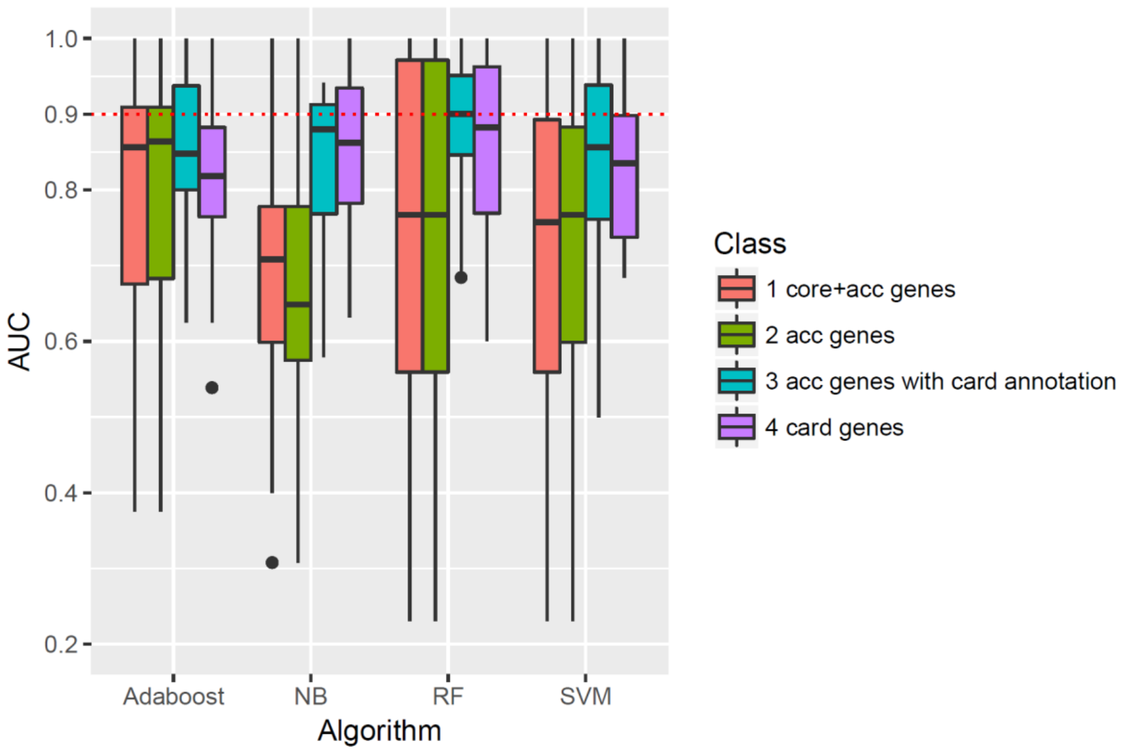
Supplementary Figure S1. Antibiotic resistance (AMR) activities of 59 *Escherichia coli* strains. A phylogenetic tree of *E. coli* strains is shown on the left. Red and blue colors in the heatmap respectively indicate resistance and susceptibility to corresponding antibiotics. Pink color indicates either intermediate, non-susceptible, or non-resistance. Classes of drugs are listed at the bottom of the heatmap using different colors. The hierarchical clustering tree of the drugs estimated from the resistance profiles is shown at the top of the figure.



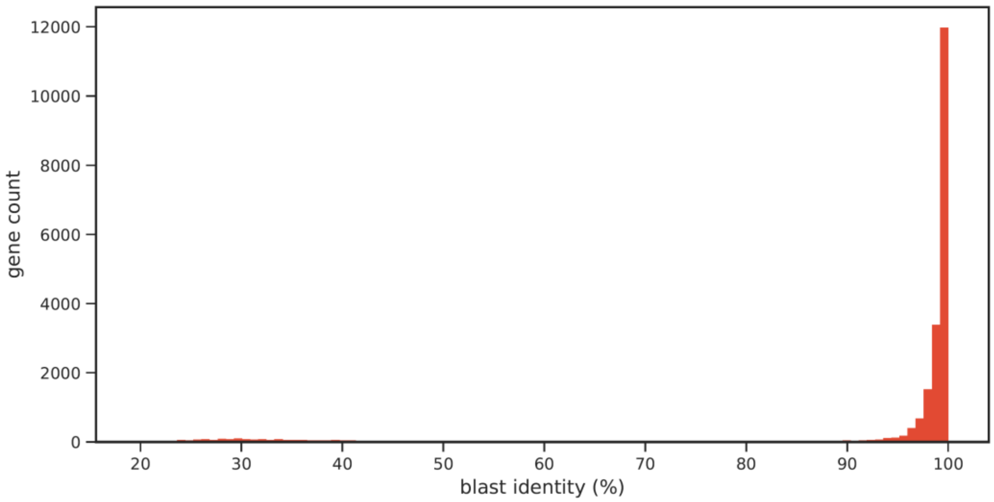
Supplement Figure S2. $F_1$ score of the AMR precision performances based on the presence/absence patterns of 1) all core + accessory gene clusters; 2) all accessory gene clusters; 3) accessory gene clusters with CARD annotations; and 4) all CARD gene clusters. The boxplots indicate the distribution of the prediction accuracies for the 12 selected drugs (see Methods and Results in the main text). The four panels indicate four different machine learning algorithms (Adaboost, Naïve Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM)) used in the prediction process.



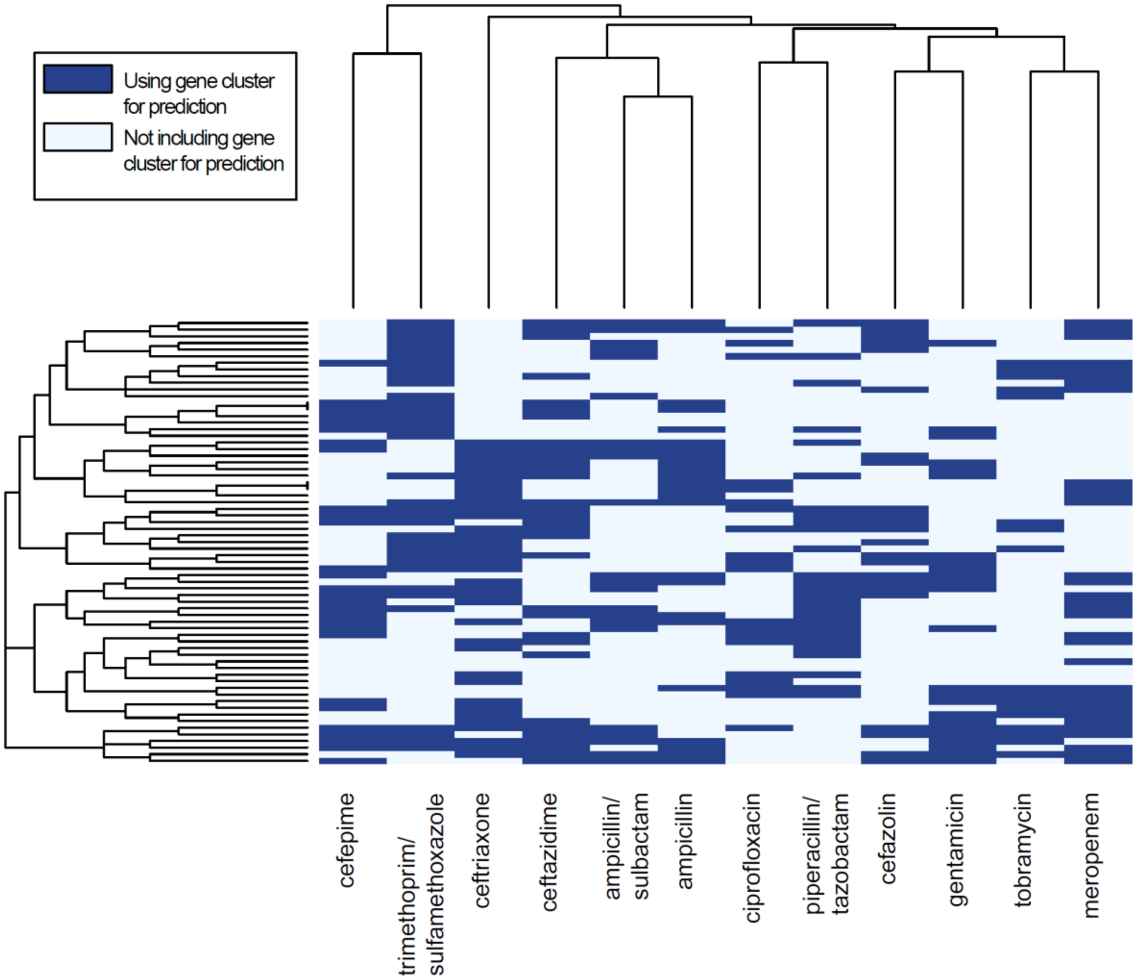
Supplement Figure S3. Precision of the AMR precision performances based on the presence/absence patterns of 1) all core + accessory gene clusters; 2) all accessory gene clusters; 3) accessory gene clusters with CARD annotations; and 4) all CARD gene clusters. The boxplots indicate the distribution of the prediction accuracies for the 12 selected drugs (see Methods and Results in the main text). The four panels indicate four different machine learning algorithms (Adaboost, Naïve Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM)) used in the prediction process.



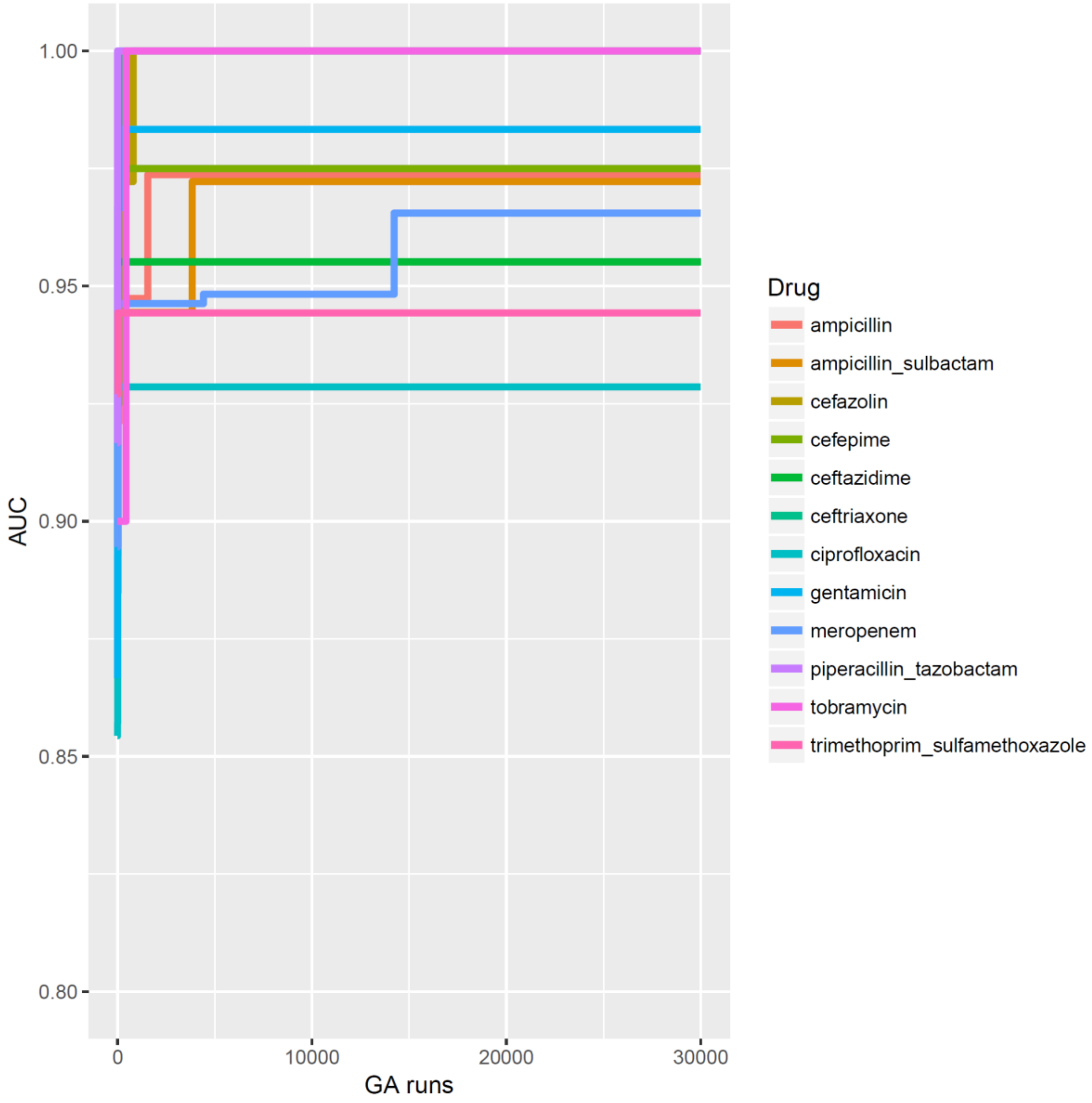
Supplement Figure S4. Recall of the AMR precision performances based on the presence/absence patterns of 1) all core + accessory gene clusters; 2) all accessory gene clusters; 3) accessory gene clusters with CARD annotations; and 4) all CARD gene clusters. The boxplots indicate the distribution of the prediction accuracies for the 12 selected drugs (see Methods and Results in the main text). The four panels indicate four different machine learning algorithms (Adaboost, Naïve Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM)) used in the prediction process.



Supplementary Figure S5. BLAST identity histogram between best-hit genes among genes of five *E. coli* strains with complete genomes (BIDMC 71, 17A, UCI 58, AR_0118, and MRSN388634)



Supplementary Figure S6. Heatmap for gene cluster usage patterns identified by the genetic algorithm. Lighter colors indicate "not using" the gene clusters for prediction purpose for corresponding drugs, and Darker colors indicate "using" the gene clusters.



Supplementary Figure S7. AUC measured for each drugs over the genetic algorithm runs. X-axis indicates the round of GA runs and y-axis represents the AUC measured from the fitness function (which is the AUC estimated from the Support Vector Machine leave-one-out evaluation using the selected gene clusters defined in each genetic algorithm "genome" (GA-genome). The AUC improvements of the 12 drugs were displayed by lines with different colors.

Supplementary Table S1. The genome IDs and assembly statistics of the E. coli genomes used in this study.

| Genome ID[1] | Genome Name | Genome status | Contig # | N50 | Genome Size | GC Content | Isolation Source | Isolation Country | Geographic Location | Host Name |
|---|---|---|---|---|---|---|---|---|---|---|
| 1328432.3 | *Escherichia coli* BWH 32 | WGS | 45 | 419572 | 5087927 | 50.5 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328433.3 | *Escherichia coli* BWH 34 | WGS | 5 | 5076304 | 5410353 | 50.6 | Wound | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328434.3 | *Escherichia coli* BWH 40 | WGS | 22 | 514200 | 5288563 | 50.7 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328436.3 | *Escherichia coli* BIDMC 2B | WGS | 11 | 3766414 | 5503383 | 50.4 | Blood culture | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328437.3 | *Escherichia coli* BIDMC 3 | WGS | 18 | 544288 | 5479839 | 50.4 | SWAB, abdominal incision | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328438.3 | *Escherichia coli* BIDMC 6 | WGS | 15 | 1439476 | 5496130 | 50.4 | bile | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328439.3 | *Escherichia coli* BIDMC 9 | WGS | 49 | 310831 | 5413189 | 50.3 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328440.3 | *Escherichia coli* BIDMC 15 | WGS | 38 | 445693 | 5457007 | 50.3 | SWAB, abdominal incision | United States | Boston, MA, USA | Human, Homo sapiens |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1328441.3 | *Escherichia coli* BIDMC 17A | WGS | 15 | 4063444 | 5509549 | 50.4 | Blood culture | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328442.3 | *Escherichia coli* BIDMC 17B | WGS | 13 | 1677481 | 5508202 | 50.4 | peritoneal fluid | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328443.3 | *Escherichia coli* BIDMC 19A | WGS | 15 | 851919 | 5467890 | 50.4 | bronchio-alveolar lavage | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328444.3 | *Escherichia coli* BIDMC 19B | WGS | 16 | 836780 | 5486042 | 50.4 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328445.3 | *Escherichia coli* BIDMC 19C | WGS | 5 | 5097434 | 5523477 | 50.4 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328446.3 | *Escherichia coli* BIDMC 20A | WGS | 30 | 503255 | 5343200 | 50.7 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328447.3 | *Escherichia coli* BIDMC 20B | WGS | 17 | 675905 | 5347131 | 50.7 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328448.3 | *Escherichia coli* BIDMC 37 | WGS | 17 | 831510 | 4653786 | 50.8 | Blood | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328449.3 | *Escherichia coli* BIDMC 38 | WGS | 30 | 470547 | 5249867 | 50.8 | Blood | United States | Boston, MA, USA | Human, Homo sapiens |
| 1328450.3 | *Escherichia coli* BIDMC 39 | WGS | 28 | 1105895 | 4855466 | 50.9 | Blood | United States | Boston, MA, USA | Human, Homo sapiens |
| 1400156.3 | *Escherichia coli* BIDMC 43a | WGS | 13 | 1014274 | 5388459 | 50.4 | Blood | United | Boston, MA, | Human, Homo |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | culture | States | USA | sapiens |
| 1400157.3 | *Escherichia coli* BIDMC 43b | WGS | 27 | 522898 | 5353135 | 50.3 | Blood culture | United States | Boston, MA, USA | Human, Homo sapiens |
| 1400158.3 | *Escherichia coli* BIDMC 49a | WGS | 33 | 361440 | 5051406 | 50.7 | Blood culture | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438668.3 | *Escherichia coli* BIDMC 58 | WGS | 13 | 794957 | 5373651 | 50.7 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438669.3 | *Escherichia coli* BIDMC 59 | WGS | 2 | 472021 | 4660068 | 50.8 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438670.3 | *Escherichia coli* BIDMC 62 | WGS | 12 | 705442 | 5384699 | 50.7 | bronchioal veolar lavage | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438671.3 | *Escherichia coli* BIDMC 63 | WGS | 6 | 789640 | 5085719 | 50.8 | SWAB, abdominal incision | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438672.3 | *Escherichia coli* BIDMC 64 | WGS | 13 | 816168 | 5345955 | 50.7 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438673.3 | *Escherichia coli* BIDMC 65 | WGS | 9 | 272156 | 5077115 | 50.6 | abssess | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438674.3 | *Escherichia coli* BIDMC 70 | WGS | 5 | 2755006 | 5131875 | 50.8 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438675.3 | *Escherichia coli* BIDMC 71 | WGS | 18 | 359503 | 5310744 | 50.6 | Urine | United | Boston, MA, | Human, Homo |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | States | USA | sapiens |
| 1438676.3 | *Escherichia coli* BIDMC 72 | WGS | 11 | 208212 | 5035467 | 50.6 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438677.3 | *Escherichia coli* BIDMC 73 | WGS | 11 | 299922 | 5039445 | 50.6 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438678.3 | *Escherichia coli* BIDMC 74 | WGS | 13 | 431375 | 4973995 | 50.7 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438679.3 | *Escherichia coli* BIDMC 75 | WGS | 15 | 343128 | 5068691 | 50.8 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438680.3 | *Escherichia coli* BIDMC 76 | WGS | 10 | 325867 | 5012972 | 50.8 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438681.3 | *Escherichia coli* BIDMC 77 | WGS | 16 | 480328 | 5224265 | 50.6 | Blood culture | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438682.3 | *Escherichia coli* BIDMC 78 | WGS | 7 | 854500 | 5144669 | 50.8 | Blood culture | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438683.3 | *Escherichia coli* BIDMC 79 | WGS | 14 | 450616 | 5122875 | 50.8 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438684.3 | *Escherichia coli* CHS 68 | WGS | 15 | 704233 | 5227634 | 50.8 | Urine | United States | North Carolina, USA | Human, Homo sapiens |
| 1438685.3 | *Escherichia coli* CHS 69 | WGS | 12 | 231547 | 5250696 | 50.7 | Urine | United States | North Carolina, USA | Human, Homo sapiens |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1438686.3 | *Escherichia coli* CHS 77 | WGS | 7 | 519460 | 5000065 | 50.7 | Urine | United States | North Carolina, USA | Human, Homo sapiens |
| 1438687.3 | *Escherichia coli* MGH 57 | WGS | 14 | 317350 | 5029585 | 50.7 | Wound | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438688.3 | *Escherichia coli* MGH 58 | WGS | 7 | 245580 | 5341964 | 50.8 | Urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1438689.3 | *Escherichia coli* UCI 51 | WGS | 16 | 488690 | 5282810 | 50.9 | urine | United States | California, USA | Human, Homo sapiens |
| 1438691.3 | *Escherichia coli* UCI 53 | WGS | 13 | 372907 | 4928679 | 50.5 | bile | United States | California, USA | Human, Homo sapiens |
| 1438693.3 | *Escherichia coli* UCI 57 | WGS | 8 | 575594 | 5036718 | 50.7 | urine | United States | California, USA | Human, Homo sapiens |
| 1438694.3 | *Escherichia coli* UCI 58 | WGS | 12 | 539623 | 5071451 | 50.6 | urine | United States | California, USA | Human, Homo sapiens |
| 1438695.3 | *Escherichia coli* UCI 65 | WGS | 12 | 341304 | 5414383 | 50.5 | urine | United States | California, USA | Human, Homo sapiens |
| 1438696.3 | *Escherichia coli* UCI 66 | WGS | 13 | 363067 | 5227535 | 50.6 | urine | United States | California, USA | Human, Homo sapiens |
| 1445862.3 | *Escherichia coli* BIDMC 82 | WGS | 145 | 114668 | 5415668 | 50.3 | urine | United States | Boston, MA, USA | Human, Homo sapiens |
| 1445863.3 | *Escherichia coli* BIDMC 83 | WGS | 97 | 162884 | 5090477 | 50.5 | urine | United States | Boston, MA, USA | Human, Homo sapiens |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 562.10576 | *Escherichia coli* strain MRSN388634 | WGS | 84 | 208790 | 5211994 | 50.33 | urine | United States | USA | Human, Homo sapiens |
| 562.12959 | *Escherichia coli* strain AR_0001 | WGS | 4 | 4466630 | 5468742 | 50.62 | N/A[2] | N/A | N/A | N/A |
| 562.12960 | *Escherichia coli* strain AR_0048 | WGS | 10 | 4952051 | 5353781 | 50.51 | N/A | N/A | N/A | N/A |
| 562.13502 | *Escherichia coli* strain ICBEC7P | WGS | 232 | 83916 | 5307281 | 50.26 | epidermis | N/A | Brazil: Santos | Spheniscus magellanicus |
| 562.16324 | *Escherichia coli* strain AR_0118 | Complete | 4 | 5018461 | 5365068 | 50.65 | N/A | N/A | N/A | N/A |
| 562.16325 | *Escherichia coli* strain AR_0061 | Complete | 3 | 4595238 | 4734219 | 50.79 | N/A | N/A | N/A | N/A |
| 562.16326 | *Escherichia coli* strain AR_0069 | Complete | 3 | 4658583 | 4826754 | 50.8 | N/A | N/A | N/A | N/A |
| 562.16327 | *Escherichia coli* strain AR_0104 | Complete | 4 | 4992088 | 5262025 | 50.76 | N/A | N/A | N/A | N/A |
| 562.16466 | *Escherichia coli* strain 5CRE51 | Complete | 3 | 5041366 | 5256426 | 50.62 | urine | N/A | Taiwan: Tainan | Human, Homo sapiens |

---

[1] PATRIC genome ID, which is also the NCBI taxon ID.

[2] N/A indicates that no information is available on PATRIC database.

Supplementary Table S2. Re-annotated *E. coli* AMR activities against different antibiotic drugs.

| E. coli strain | BWH 32 | BWH 34 | BWH 40 | BIDMC 2B | BIDMC 3 | BIDMC 6 | BIDMC 9 | BIDMC 15 | BIDMC 17A | BIDMC 17B |
|---|---|---|---|---|---|---|---|---|---|---|
| Taxon ID | 1328432 | 1328433 | 1328434 | 1328436 | 1328437 | 1328438 | 1328439 | 1328440 | 1328441 | 1328442 |
| meropenem | S[1] | R[1] | I[1] | R | R | R | R | R | R | R |
| gentamicin | S | R | R |  |  |  | Non-S[1] |  |  |  |
| ciprofloxacin | S | S | S |  |  |  |  | Non-S |  | Non-S |
| trimethoprim/sulfamethoxazole |  | R | R |  |  |  |  |  | R | R |
| cefazolin | S | R | R |  |  |  |  |  |  | Non-S |
| ampicillin | S | R | R |  |  |  |  |  |  | R |
| ampicillin/sulbactam | S | R | R |  | R |  | R |  |  |  |
| ceftazidime | S | R | R |  |  |  |  |  | Non-S |  |
| cefepime |  |  |  |  | R |  |  |  | R | R |
| piperacillin/tazobactam |  |  |  |  |  |  | R |  |  |  |
| ceftriaxone |  |  |  |  | I |  |  |  | Non-S | Non-S |
| tobramycin |  |  |  |  |  |  |  | S |  |  |
| nitrofurantoin | S | S | S |  |  |  |  |  |  |  |
| amikacin | S | S | S |  |  |  |  |  |  |  |
| imipenem | S | S | S |  |  |  |  |  |  |  |
| tigecycline | S | R | R |  |  |  |  |  |  |  |
| ertapenem | S |  |  |  |  |  |  |  |  |  |
| levofloxacin | S | R | R |  |  |  |  |  |  |  |
| cefotaxime | S | R | R |  |  |  |  |  |  |  |
| cefoxitin | S | R | R |  |  |  |  |  |  |  |
| tetracycline |  |  |  |  |  |  |  | S |  | S |
| cephalothin | S | S | R |  |  |  |  |  |  |  |
| colistin |  |  |  |  |  |  |  |  |  |  |
| aztreonam |  |  |  |  |  |  |  |  |  |  |
| amoxicillin/clavulanic acid |  |  |  |  |  |  |  |  |  |  |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| doripenem | | | | | | | | | | |
| doxycycline | | | | | | | | | | |
| polymyxin B | | | | | | | | | | |
| minocycline | | | | | | | | | | |
| fosfomycin | | | | | | | | | | |
| norfloxacin | | | | | | | | | | |
| chloramphenicol | | | | | | | | | | |
| ticarcillin-clavulanic acid | | | | | | | Non-S | | | |
| trimethoprim | | | | | | | | | | |
| cefotaxime/clavulanic acid | | | | | | | | | | |
| cephalexin | | | | | | | | | | |
| piperacillin | | | | | | | | | | |

| E. coli strain | BIDMC 19A | BIDMC 19B | BIDMC 19C | BIDMC 20A | BIDMC 20B | BIDMC 37 | BIDMC 38 | BIDMC 39 | BIDMC 43a | BIDMC 43b |
|---|---|---|---|---|---|---|---|---|---|---|
| Taxon ID | 1328443 | 1328444 | 1328445 | 1328446 | 1328447 | 1328448 | 1328449 | 1328450 | 1400156 | 1400157 |
| meropenem | R | R | R | R | R | S | S | S | I | R |
| gentamicin | Non-S | | | | | S | Non-S | S | Non-S | Non-S |
| ciprofloxacin | | | | | | S | S | Non-S | Non-S | Non-S |
| trimethoprim/sulfamethoxazole | | | | | | S | S | S | R | R |
| cefazolin | | | | | | S | Non-S | Non-S | Non-S | Non-S |
| ampicillin | | | | | | S | R | S | S | S |
| ampicillin/sulbactam | | | | | | S | S | S | S | R |
| ceftazidime | Non-S | | | | | S | I | Non-S | Non-S | Non-S |
| cefepime | R | R | R | S | | S | S | S | S | S |
| piperacillin/tazobactam | | | | | | S | S | S | R | R |
| ceftriaxone | | Non-S | Non-S | | | S | S | S | R | I |
| tobramycin | | | | | | S | S | S | S | I |
| nitrofurantoin | | | | | | | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| amikacin | | S | | | | | | | S | S |
| imipenem | | | | | | | | | | |
| tigecycline | | | | | | | | | R | I |
| ertapenem | | | | | | | | | I | I |
| levofloxacin | | | | | | | | | | |
| cefotaxime | | | | | | | | | | |
| cefoxitin | | | | | | | | | | |
| tetracycline | | | | | | | | | S | |
| cephalothin | | | | | | | | | | |
| colistin | | | | | | | | | | Non-R[1] |
| aztreonam | | | | | | | | | R | R |
| amoxicillin/clavulanic acid | | | | | | | | | | |
| doripenem | | | | | | | | | I | S |
| doxycycline | | | | | | | | | S | S |
| polymyxin B | | | | | | | | | | |
| minocycline | | | | | | | | | S | S |
| fosfomycin | | | | | | | | | | |
| norfloxacin | | | | | | | | | | |
| chloramphenicol | | | | | | | | | | |
| ticarcillin-clavulanic acid | | | | | | | | | | |
| trimethoprim | | | | | | | | | | |
| cefotaxime/clavulanic acid | | | | | | | | | | |
| cephalexin | | | | | | | | | | |
| piperacillin | | | | | | | | | | |

| E. coli strain | BIDMC 49a | BIDMC 58 | BIDMC 59 | BIDMC 62 | BIDMC 63 | BIDMC 64 | BIDMC 65 | BIDMC 70 | BIDMC 71 | BIDMC 72 |
|---|---|---|---|---|---|---|---|---|---|---|
| Taxon ID | 1400158 | 1438668 | 1438669 | 1438670 | 1438671 | 1438672 | 1438673 | 1438674 | 1438675 | 1438676 |
| meropenem | S | R | S | R | S | R | S | S | S | S |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| gentamicin | S | Non-S | S | Non-S | Non-S | Non-S | S | Non-S | Non-S | Non-S |
| ciprofloxacin | S | S | S | S | S | S | S | S | S | S |
| trimethoprim/sulfamethoxazole | S | R | S | R | S | R | S | S | S | S |
| cefazolin | S | Non-S | S | Non-S | S | Non-S | S | Non-S | Non-S | Non-S |
| ampicillin | S | R | S | R | S | R | S | R | R | R |
| ampicillin/sulbactam | S | R | S | R | S | R | S | S | S | S |
| ceftazidime | S | Non-S | S | Non-S | S | Non-S | S | I | Non-S | I |
| cefepime | S | R | S | R | S | R | S | S | S | S |
| piperacillin/tazobactam | S | R | S | R | S | R | S | S | S | S |
| ceftriaxone | S | | S | | S | Non-S | S | | S | S |
| tobramycin | S | S | S | S | S | S | S | S | Non-S | S |
| nitrofurantoin | | S | S | | | S | | S | I | S |
| amikacin | | | | | | | | | | |
| imipenem | | | | | | | | | | |
| tigecycline | | | | | | | | | | |
| ertapenem | | | | | | | | | | |
| levofloxacin | | | | | | | | | | |
| cefotaxime | | | | | | | | | | |
| cefoxitin | | | | | | | | | | |
| tetracycline | | Non-S | | S | | Non-S | | | | |
| cephalothin | | | | | | | | | | |
| colistin | | S | | S | | S | | | | |
| aztreonam | | | | | | | | | | |
| amoxicillin/clavulanic acid | | | | | | | | | | |
| doripenem | | | | | | | | | | |
| doxycycline | | R | | R | | S | | | | |
| polymyxin B | | | | | | | | | | |
| minocycline | | R | | R | | S | | | | |
| fosfomycin | | | | | | | | | | |

| |
|---|
| norfloxacin |
| chloramphenicol |
| ticarcillin-clavulanic acid |
| trimethoprim |
| cefotaxime/clavulanic acid |
| cephalexin |
| piperacillin |

| E. coli strain | BIDMC 73 | BIDMC 74 | BIDMC 75 | BIDMC 76 | BIDMC 77 | BIDMC 78 | BIDMC 79 | CHS 68 | CHS 69 | CHS 77 |
|---|---|---|---|---|---|---|---|---|---|---|
| Taxon ID | 1438677 | 1438678 | 1438679 | 1438680 | 1438681 | 1438682 | 1438683 | 1438684 | 1438685 | 1438686 |
| meropenem | S | S | S | S | S | S | S | R | R | R |
| gentamicin | Non-S | Non-S | Non-S | Non-S | Non-S | Non-S | Non-S | R | R | R |
| ciprofloxacin | S | S | S | S | Non-S | S | S | R | R | S |
| trimethoprim/sulfamethoxazole | S | R | R | S | R | S | S | R | R | S |
| cefazolin | S | Non-S | Non-S | Non-S | Non-S | Non-S | Non-S | R | R | R |
| ampicillin | R | R | R | S | R | S | R | R | S | R |
| ampicillin/sulbactam | S | R | S | S | R | S | S | R | R | S |
| ceftazidime | S | Non-S | Non-S | I | Non-S | I | S | R | R | R |
| cefepime | S | S | S | S | R | S | S | | | |
| piperacillin/tazobactam | S | R | S | S | R | S | S | | | |
| ceftriaxone | S | I | S | S | S | S | S | | | |
| tobramycin | S | S | S | S | Non-S | S | S | | | |
| nitrofurantoin | S | S | S | S | | | S | R | R | R |
| amikacin | | | | | S | | | R | S | S |
| imipenem | | | | | | | | S | | R |
| tigecycline | | | | | | | | R | R | R |
| ertapenem | | | | | | | | R | R | R |
| levofloxacin | | | | | | | | R | R | R |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| cefotaxime | | | | | | | | R | R | I |
| cefoxitin | | | | | | | | R | R | R |
| tetracycline | | | | | | | | | | |
| cephalothin | | | | | | | | R | R | R |
| colistin | | | | | | | | | | |
| aztreonam | | | | | | | | | | |
| amoxicillin/clavulanic acid | | | | | | | | | | |
| doripenem | | | | | | | | | | |
| doxycycline | | | | | | | | | | |
| polymyxin B | | | | | | | | | | |
| minocycline | | | | | | | | | | |
| fosfomycin | | | | | | | | | | |
| norfloxacin | | | | | | | | | | |
| chloramphenicol | | | | | | | | | | |
| ticarcillin-clavulanic acid | | | | | | | | | | |
| trimethoprim | | | | | | | | | | |
| cefotaxime/clavulanic acid | | | | | | | | | | |
| cephalexin | | | | | | | | | | |
| piperacillin | | | | | | | | | | |

| E. coli strain | MGH 57 | MGH 58 | UCI 51 | UCI 53 | UCI 57 | UCI 58 | UCI 65 | UCI 66 | BIDMC 82 | BIDMC 83 |
|---|---|---|---|---|---|---|---|---|---|---|
| Taxon ID | 1438687 | 1438688 | 1438689 | 1438691 | 1438693 | 1438694 | 1438695 | 1438696 | 1445862 | 1445863 |
| meropenem | R | S | R | I | S | S | S | S | R | S |
| gentamicin | R | S | R | R | S | S | R | S | Non-S | S |
| ciprofloxacin | S | S | R | S | R | R | R | S | Non-S | S |
| trimethoprim/sulfamethoxazole | R | S | Non-S | Non-S | Non-S | Non-S | Non-S | Non-S | R | S |
| cefazolin | R | R | R | R | R | R | R | R | Non-S | S |
| ampicillin | Non-R | R | R | Non-R | Non-R | R | R | Non-R | R | S |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ampicillin/sulbactam | | | R | R | R | R | R | R | R | S |
| ceftazidime | R | S | R | R | R | R | R | R | Non-S | S |
| cefepime | | | | | | | | | R | S |
| piperacillin/tazobactam | | | | | | | | | R | S |
| ceftriaxone | | | | | | | | | Non-S | S |
| tobramycin | | | | | | | | | I | S |
| nitrofurantoin | | S | S | S | I | I | R | S | S | S |
| amikacin | S | S | S | S | S | S | S | S | S | |
| imipenem | | | | | | | | | | |
| tigecycline | I | S | I | I | I | S | I | S | | |
| ertapenem | R | S | Non-S | Non-S | Non-S | S | Non-S | S | | |
| levofloxacin | R | S | R | R | S | S | R | S | | |
| cefotaxime | R | S | R | R | R | R | R | R | | |
| cefoxitin | | | R | R | R | R | R | R | | |
| tetracycline | | | | | | | | | | |
| cephalothin | | | | | | | | | | |
| colistin | | | | | | | | | | |
| aztreonam | | | | | | | | | | |
| amoxicillin/clavulanic acid | | | | | | | | | | |
| doripenem | | | | | | | | | | |
| doxycycline | | | | | | | | | S | |
| polymyxin B | | | | | | | | | | |
| minocycline | | | | | | | | | S | |
| fosfomycin | | | | | | | | | | |
| norfloxacin | | | | | | | | | | |
| chloramphenicol | | | | | | | | | | |
| ticarcillin-clavulanic acid | | | | | | | | | | |
| trimethoprim | | | | | | | | | | |
| cefotaxime/clavulanic acid | | | | | | | | | | |

| |
|---|
| cephalexin |
| piperacillin |

| E. coli strain | MRSN388642 | AR_0001 | AR_0048 | ICBEC7P | AR_0118 | AR_0061 | AR_0069 | AR_0104 | 5CRE51 |
|---|---|---|---|---|---|---|---|---|---|
| Taxon ID | 562.10576 | 562.12959 | 562.12960 | 562.13502 | 562.16324 | 562.16325 | 562.16326 | 562.16327 | 562.16466 |
| meropenem | S | R | R |  | R | R | R | S | R |
| gentamicin | Non-S | R | R | R | R | S | S | R | Non-S |
| ciprofloxacin | Non-S | S | R | R | R | R | S | S | Non-S |
| trimethoprim/sulfamethoxazole | Non-S | Non-S | R |  | R | Non-S | Non-S | Non-S | Non-S |
| cefazolin | Non-S | R | R |  | R | R | R | R | Non-S |
| ampicillin | Non-S | R | R |  | R | R | R | R | R |
| ampicillin/sulbactam | R | R | R |  | R | R | R | R | R |
| ceftazidime |  | R | R |  | R | R | R | R |  |
| cefepime | R | R | R |  | R | R | R | R | R |
| piperacillin/tazobactam | R | R | R |  | R | R | R | R | R |
| ceftriaxone | S | R | R |  |  | R | R | R | Non-S |
| tobramycin | Non-S | R | R |  | R | R | R | S | Non-S |
| nitrofurantoin | S |  |  |  |  |  |  |  | S |
| amikacin | S | S | R |  | R | Non-R | S | S | S |
| imipenem |  | S | S |  | Non-R | S | S | S | Non-R |
| tigecycline | S | R | R |  | R | R | R | R | R |
| ertapenem | S | R | R | S | R | R | R | Non-R | R |
| levofloxacin | Non-S | R | R |  | R | S | S | R |  |
| cefotaxime |  | R | R | R | R | R | R | R |  |
| cefoxitin |  | Non-S | Non-S | S | R | Non-R | Non-S | Non-S | R |
| tetracycline | Non-S | R | S | R | S | R | R | R | R |
| cephalothin |  |  |  |  |  |  |  |  |  |
| colistin | R | S | S |  | S | S | S | S | R |
| aztreonam | R | R | R |  | R | R | Non-R | R | R |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| amoxicillin/clavulanic acid | R | R | | R | R | R | R | R |
| doripenem | R | R | | | R | R | S | |
| doxycycline | | | | | | | | |
| polymyxin B | | | | Non-R | Non-R | Non-R | Non-R | |
| minocycline | | | | | | | | |
| fosfomycin | | | | | | | | Non-S |
| norfloxacin | | | | | | | | R |
| chloramphenicol | | | | | | | | Non-S |
| ticarcillin-clavulanic acid | | | | | | | | |
| trimethoprim | | | R | | | | | |
| cefotaxime/clavulanic acid | | | | | | | | Non-S |
| cephalexin | | | | | | | | R |
| piperacillin | | | | | | | | R |

---

[1] S: susceptibility; R: resistance; Non-S: non-susceptibility; Non-R: non-resistance; I: intermediate.

Supplementary Table S3. Number of entries of each antibiotic drug in the PATRIC database. Shaded drugs are those used in the prediction and evaluation processes.

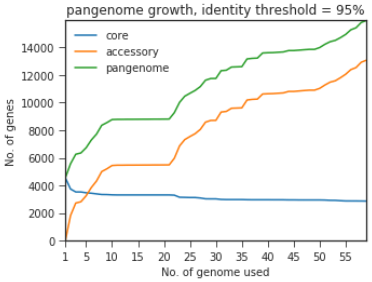
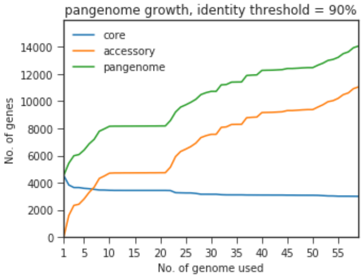
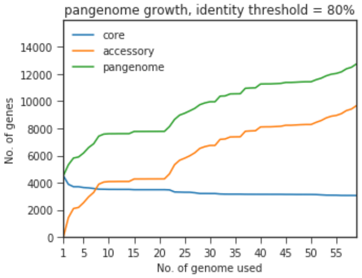
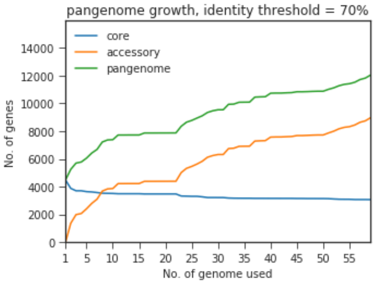
| Antibiotic drugs | Drug class | Entry number measured for *E. coli* in PATRIC | Resistant / Susceptible strains # | I / non-R / non-S strain #[1] |
|---|---|---|---|---|
| meropenem | carbapenems | 58 | 29 / 26[2] | 3 |
| ciprofloxacin | quinolones | 49 | 14 / 13 | 22 |
| gentamicin | aminoglycosides | 49 | 10 / 30 | 9 |
| ampicillin | penicillins | 47 | 19 / 8 | 20 |
| cefazolin | cephalosporins | 47 | 18 / 17 | 12 |
| trimethoprim/ sulfamethoxazole | folate pathway | 47 | 29 / 13 | 5 |
| ampicillin/ sulbactam | beta lactam inhibitor | 46 | 18 / 10 | 18 |
| ceftazidime | cephalosporins | 46 | 27 / 19 | 0 |
| cefepime | cephalosporins | 39 | 19 / 20 | 0 |
| ceftriaxone | cephalosporins | 33 | 17 / 16 | 0 |
| piperacillin/ tazobactam | beta lactam inhibitor | 33 | 6 / 17 | 10 |
| tobramycin | aminoglycosides | 33 | 5 / 22 | 6 |
| nitrofurantoin | nitrofurans | 28 | 4 / 21 | 3 |
| amikacin | aminoglycosides | 27 | 3 / 23 | 1 |
| imipenem | carbapenems | 24 | 13 / 5 | 6 |
| ertapenem | carbapenems | 23 | 10 / 6 | 7 |
| cefotaxime | cephalosporins | 21 | 18 / 2 | 1 |
| levofloxacin | quinolones | 21 | 13 / 7 | 1 |
| cefoxitin | cephalosporins | 20 | 13 / 2 | 5 |
| tetracycline | tetracyclines | 15 | 6 / 6 | 3 |

| | | | | |
|---|---|---|---|---|
| colistin | lipopeptides | 12 | 2 / 9 | 1 |
| tigecycline | tetracyclines | 12 | 1 / 9 | 2 |
| aztreonam | lipopeptides | 10 | 9 / 0 | 1 |
| amoxicillin/ clavulanic acid | beta lactam inhibitor | 7 | 7 / 0 | 0 |
| doripenem | carbapenems | 7 | 4 / 2 | 1 |
| cephalothin | cephalosporins | 6 | 4 / 2 | 0 |
| doxycycline | tetracyclines | 6 | 2 / 4 | 0 |
| minocycline | tetracyclines | 6 | 2 / 4 | 0 |
| polymyxin B | lipopeptides | 4 | 0 / 0 | 4 |
| cefotaxime/ clavulanic acid | beta lactam inhibitor | 1 | 1 / 0 | 0 |
| cephalexin | cephalosporins | 1 | 1 / 0 | 0 |
| chloramphenicol | phenicols | 1 | 1 / 0 | 0 |
| fosfomycin | fosfomycins | 1 | 0 / 0 | 1 |
| norfloxacin | quinolones | 1 | 0 / 0 | 1 |
| piperacillin | penicillins | 1 | 0 / 0 | 1 |
| ticarcillin-clavulanic acid | beta lactam inhibitor | 1 | 1 / 0 | 0 |
| trimethoprim | folate pathway | 1 | 0 / 0 | 1 |

[1] I stands for “Intermediate,” non-R stands for “non-resistant,” and non-S stands for “non-susceptible.” See Methods for details.

[2] The former number indicates strains annotated as “resistant” while the latter indicates “susceptible.”

Supplementary Table S4. Number of gene clusters for core, accessory, and the whole pan-genome using different clustering identity cutoffs. The plots of the growth curves of the core, accessory, and pan-genome are also shown.

| Identity threshold | 95% | 90% | 80% | 70% |
|---|---|---|---|---|
| Core gene clusters | 2874 | 3007 | 3068 | 3074 |
| Accessory gene clusters | 13076 | 11054 | 9862 | 8971 |
| Pan-genome size | 15950 | 14361 | 12930 | 12045 |
| Pan-genome growth plot |  |  |  |  |

Supplementary Table S5. Distribution of COG categories in the core- and accessory-genomes.

| COG | Core-genome | Accessory-genome |
|---|---:|---:|
| J | 6.29% | 0.68% |
| A | 0.04% | 0.03% |
| K | 6.21% | 8.08% |
| L | 4.34% | 14.76% |
| B | 0.00% | 0.00% |
| D | 1.01% | 0.91% |
| Y | 0.00% | 0.00% |
| V | 1.16% | 2.65% |
| T | 4.49% | 2.23% |
| M | 5.87% | 10.68% |
| N | 1.42% | 5.34% |
| Z | 0.00% | 0.00% |
| W | 0.00% | 0.34% |
| U | 1.61% | 8.46% |
| O | 4.64% | 1.63% |
| C | 8.64% | 3.10% |
| G | 8.53% | 7.16% |
| E | 10.21% | 2.91% |
| F | 2.96% | 0.92% |
| H | 4.38% | 1.27% |
| I | 2.51% | 1.15% |
| P | 7.07% | 3.99% |
| Q | 1.08% | 1.22% |
| R | 7.15% | 9.62% |
| S | 10.40% | 12.86% |

Supplementary Table S6. The 68 accessory genes with CARD annotations.

| Gene cluster# | AROs[1] | Best ARO hit |
|---|---|---|
| Cluster 1202 | ARO:3003577 | PmrE |
| Cluster 1513 | ARO:3003576 | PmrC |
| Cluster 1549 | ARO:3003689 | MCR-1 |
| Cluster 1564 | ARO:3004139 | MCR-3 |
| Cluster 1961 | ARO:3003550 | mdtP |
| Cluster 2387 | ARO:3003889 | Escherichia coli GlpT with mutation conferring resistance to fosfomycin |
| Cluster 2855 | ARO:3001214 | mdtM |
| Cluster 2965 | ARO:3003839 | Mrx |
| Cluster 3112 | ARO:3002705 | floR |
| Cluster 3154 | ARO:3000166, ARO:3000165 | tet(A) |
| Cluster 3361 | ARO:3000027 | emrA |
| Cluster 3397 | ARO:3003577 | PmrE |
| Cluster 3408 | ARO:3003577 | PmrE |
| Cluster 3412 | ARO:3003577 | PmrE |
| Cluster 3521 | ARO:3002071, ARO:3002117, ARO:3002085, ARO:3002053, ARO:3002049, ARO:3002034, ARO:3002095, ARO:3002044, ARO:3002130, ARO:3002015, ARO:3002097, ARO:3002113, ARO:3002050, ARO:3002041, ARO:3002128, ARO:3002013, ARO:3002120, ARO:3002047, ARO:3002029, ARO:3002036, ARO:3002122, ARO:3002054, ARO:3002091, ARO:3002057, ARO:3002074, ARO:3002033, ARO:3002115, ARO:3002072, ARO:3002037, ARO:3002023, ARO:3002075, ARO:3002124, ARO:3002039, | CMY-2 |

| | | |
|---|---|---|
| | ARO:3002024, ARO:3002045, ARO:3002088, ARO:3002058, ARO:3002084, ARO:3002027, ARO:3002031, ARO:3002123, ARO:3002061, ARO:3002129, ARO:3002112, ARO:3002055, ARO:3002043, ARO:3002067, ARO:3002126, ARO:3002038, ARO:3002092, ARO:3002065, ARO:3002099, ARO:3002077, ARO:3002026, ARO:3002100, ARO:3002056, ARO:3002025, ARO:3001857, ARO:3002093, ARO:3002040, ARO:3002028, ARO:3002046, ARO:3002083, ARO:3002064, ARO:3002089, ARO:3002096, ARO:3002078, ARO:3002103, ARO:3002018, ARO:3002094, ARO:3002069, ARO:3002062, ARO:3002125, ARO:3002059, ARO:3002087, ARO:3002070, ARO:3002035, ARO:3002082, ARO:3001856, ARO:3002066, ARO:3002090, ARO:3002042, ARO:3002048, ARO:3002116, ARO:3002127, ARO:3002052, ARO:3002108, ARO:3002076, ARO:3002017, ARO:3002068, ARO:3002079, ARO:3002060, ARO:3002111, ARO:3002016, ARO:3002106, ARO:3002032, ARO:3002051, ARO:3002086, ARO:3002114, ARO:3002081, ARO:3002080, ARO:3002107, ARO:3002073, ARO:3002098 | |
| Cluster 4155 | ARO:3002826 | EreA2 |
| Cluster 4265 | ARO:3003548 | mdtN |
| Cluster 4326 | ARO:3002655 | APH(4)-Ia |
| Cluster 4764 | ARO:3002660 | APH(6)-Id |
| Cluster 5292 | ARO:3000316 | mphA |
| Cluster 5317 | ARO:3002540 | AAC(3)-VIa |
| Cluster 5512 | ARO:3002320, ARO:3002325, ARO:3002324, ARO:3002317, ARO:3002329, ARO:3002326, ARO:3002327, ARO:3002318, ARO:3002313, ARO:3002321, ARO:3002322, ARO:3002314, ARO:3002315, ARO:3002319, ARO:3002312, | KPC-3 |

| | | |
|---|---|---|
| | ARO:3002316, ARO:3003180, ARO:3002323 | |
| Cluster 5572 | ARO:3001988, ARO:3001944, ARO:3001973, ARO:3001946, ARO:3001963, ARO:3001971, ARO:3001972, ARO:3001881, ARO:3001947, ARO:3001981, ARO:3001886, ARO:3001942, ARO:3001950, ARO:3001908, ARO:3001961, ARO:3001926, ARO:3001900, ARO:3001966, ARO:3001953, ARO:3001911, ARO:3001879, ARO:3001982, ARO:3001992, ARO:3001928, ARO:3002495, ARO:3001872, ARO:3001984, ARO:3001958, ARO:3002006, ARO:3001877, ARO:3001876, ARO:3001883, ARO:3001909, ARO:3003168, ARO:3001965, ARO:3001871, ARO:3001939, ARO:3001957, ARO:3001994, ARO:3001970, ARO:3001910, ARO:3001945, ARO:3001912, ARO:3001902, ARO:3001925, ARO:3002005, ARO:3001980, ARO:3001907, ARO:3001880, ARO:3001889, ARO:3001924, ARO:3001989, ARO:3001913, ARO:3001985 | CTX-M-14 |
| Cluster 5574 | ARO:3001917, ARO:3001941, ARO:3001930, ARO:3002002, ARO:3001929, ARO:3001977, ARO:3003163, ARO:3001866, ARO:3001916, ARO:3001976, ARO:3001896, ARO:3001878, ARO:3001939, ARO:3001982, ARO:3001922, ARO:3001943, ARO:3001948, ARO:3001884, ARO:3001969, ARO:3001874, ARO:3001914, ARO:3001932, ARO:3003165, ARO:3001962, ARO:3001927, ARO:3001967, ARO:3001894, ARO:3001885, ARO:3001991, ARO:3001940, ARO:3001890, ARO:3003164, ARO:3001994, ARO:3003166, ARO:3001960, ARO:3001904, ARO:3001899, ARO:3001997, ARO:3001873, ARO:3001925, ARO:3001968, ARO:3001895, ARO:3001875, ARO:3001892, ARO:3001921, ARO:3001995, ARO:3001956, ARO:3001898, ARO:3001919, ARO:3001923, ARO:3001891, ARO:3001915, ARO:3001864, ARO:3001974, ARO:3002000, ARO:3001933 | CTX-M-55 |

| | | |
|---|---|---|
| Cluster 5575 | ARO:3001396, ARO:3001399, ARO:3001425, ARO:3001793, ARO:3001781, ARO:3001427, ARO:3001806 | OXA-1 |
| Cluster 5686 | ARO:3001145, ARO:3001187, ARO:3002456, ARO:3001157, ARO:3003156, ARO:3001337, ARO:3001155, ARO:3001340, ARO:3001190, ARO:3001094, ARO:3001107, ARO:3002467, ARO:3001084, ARO:3001195, ARO:3001123, ARO:3001087, ARO:3001199, ARO:3001179, ARO:3002465, ARO:3002468, ARO:3001090, ARO:3001086, ARO:3001116, ARO:3001066, ARO:3001194, ARO:3001126, ARO:3001351, ARO:3001158, ARO:3001152, ARO:3001106, ARO:3001178, ARO:3001347, ARO:3001112, ARO:3001111, ARO:3001115, ARO:3001072, ARO:3001102, ARO:3001361, ARO:3001105, ARO:3001085, ARO:3002450, ARO:3001153, ARO:3001159, ARO:3001108, ARO:3001345, ARO:3001201, ARO:3001114, ARO:3001132, ARO:3001127, ARO:3001099, ARO:3001139, ARO:3002455, ARO:3001089, ARO:3002474, ARO:3001176, ARO:3003152, ARO:3001167, ARO:3003155, ARO:3001203, ARO:3001119, ARO:3001183, ARO:3001118, ARO:3001189, ARO:3001141, ARO:3001196, ARO:3002438, ARO:3001357, ARO:3001135, ARO:3001140, ARO:3001098, ARO:3001177, ARO:3001130, ARO:3001175, ARO:3003154, ARO:3001137, ARO:3001191, ARO:3001088, ARO:3002466, ARO:3001067, ARO:3001103, ARO:3001096, ARO:3001100, ARO:3002470, ARO:3001075, ARO:3001336, ARO:3001184, ARO:3001202, ARO:3001117, ARO:3001364, ARO:3001060, ARO:3001185, ARO:3001338, ARO:3001131, ARO:3001188, ARO:3001362, ARO:3001350, ARO:3001071, ARO:3001136, ARO:3002469, ARO:3001148, ARO:3001149, ARO:3001068, ARO:3001156, ARO:3001168, ARO:3001200, ARO:3001113, ARO:3001160, ARO:3001133, ARO:3001144, ARO:3001120, ARO:3001344, | SHV-160 |

| | | |
|---|---|---|
| | ARO:3001356, ARO:3001073, ARO:3001198, ARO:3001197, ARO:3001138, ARO:3001181, ARO:3001182, ARO:3001076, ARO:3001061, ARO:3001109, ARO:3003153, ARO:3002437, ARO:3001125, ARO:3001154, ARO:3001151, ARO:3001093, ARO:3001129, ARO:3001091, ARO:3001204, ARO:3001134, ARO:3001186, ARO:3001124, ARO:3001092, ARO:3001150, ARO:3001104, ARO:3001059, ARO:3001083, ARO:3002439, ARO:3002464, ARO:3001082, ARO:3001147, ARO:3001062, ARO:3001146, ARO:3001128, ARO:3001101, ARO:3001352, ARO:3001064, ARO:3002471, ARO:3001193, ARO:3001074, ARO:3001121, ARO:3001070, ARO:3001122, ARO:3001095, ARO:3001192 | |
| Cluster 5687 | ARO:3000950, ARO:3000942, ARO:3000910, ARO:3000968, ARO:3000924, ARO:3001055, ARO:3001390, ARO:3000888, ARO:3001020, ARO:3000961, ARO:3000893, ARO:3000884, ARO:3001018, ARO:3001028, ARO:3000887, ARO:3000883, ARO:3000879, ARO:3001030, ARO:3000989, ARO:3001006, ARO:3001045, ARO:3000947, ARO:3000922, ARO:3001037, ARO:3001392, ARO:3000952, ARO:3001003, ARO:3000900, ARO:3000987, ARO:3000890, ARO:3000903, ARO:3001394, ARO:3000958, ARO:3001000, ARO:3000891, ARO:3000876, ARO:3001057, ARO:3001373, ARO:3001026, ARO:3001388, ARO:3000995, ARO:3001014, ARO:3000918, ARO:3000882, ARO:3000946, ARO:3000949, ARO:3000975, ARO:3000904, ARO:3000978, ARO:3000878, ARO:3001391, ARO:3000892, ARO:3001046, ARO:3001048, ARO:3001005, ARO:3000956, ARO:3000941, ARO:3000988, ARO:3000934, ARO:3000999, ARO:3000969, ARO:3000929, ARO:3001054, ARO:3001012, ARO:3001393, ARO:3001385, ARO:3000912, ARO:3001007, ARO:3001378, ARO:3000960, ARO:3000899, ARO:3001056, ARO:3000986, ARO:3003158, ARO:3000990, | TEM-1 |

| | | |
|---|---|---|
| | ARO:3001022, ARO:3001025, ARO:3000936, ARO:3003157, ARO:3000972, ARO:3001021, ARO:3000954, ARO:3000916, ARO:3000993, ARO:3001035, ARO:3000938, ARO:3000963, ARO:3000970, ARO:3000951, ARO:3000935, ARO:3000894, ARO:3001053, ARO:3000914, ARO:3000962, ARO:3000939, ARO:3000982, ARO:3000981, ARO:3000873, ARO:3000896, ARO:3000953, ARO:3001386, ARO:3000985, ARO:3001050, ARO:3000957, ARO:3000974, ARO:3001369, ARO:3001032, ARO:3001375, ARO:3000898, ARO:3000996, ARO:3001034, ARO:3001015, ARO:3001033, ARO:3001376, ARO:3001016, ARO:3001004, ARO:3001024, ARO:3000948, ARO:3000874, ARO:3000998, ARO:3001041, ARO:3000928, ARO:3000965, ARO:3001374, ARO:3001058, ARO:3000880, ARO:3000959, ARO:3000994, ARO:3000976, ARO:3001051, ARO:3000979, ARO:3000984, ARO:3000875, ARO:3000923, ARO:3001023, ARO:3001001, ARO:3001042, ARO:3001017, ARO:3000955, ARO:3000937, ARO:3001019, ARO:3001002, ARO:3000964, ARO:3000931, ARO:3000967, ARO:3001043, ARO:3000977, ARO:3000980, ARO:3001384, ARO:3000921, ARO:3000943, ARO:3000886, ARO:3000983, ARO:3001029, ARO:3001382, ARO:3000997, ARO:3001049, ARO:3000917, ARO:3000971, ARO:3001052, ARO:3000926, ARO:3000911, ARO:3000944, ARO:3001047, ARO:3001013, ARO:3001383, ARO:3000973 | |
| Cluster 5700 | ARO:3002534, ARO:3002535, ARO:3002533 | AAC(3)-IIa |
| Cluster 5871 | ARO:3000410 | sul1 |
| Cluster 5926 | ARO:3000412 | sul2 |
| Cluster 5972 | ARO:3001799, ARO:3001487, ARO:3001410, ARO:3001430, ARO:3001810, ARO:3001800, ARO:3001428, ARO:3001426, ARO:3001397, ARO:3001398, ARO:3001802, ARO:3001416, | OXA-2 |

| | | |
|---|---|---|
| | ARO:3001606 | |
| Cluster 6005 | ARO:3000508 | gadX |
| Cluster 6010 | ARO:3001404 | OXA-9 |
| Cluster 6105 | ARO:3002641 | APH(3')-Ia |
| Cluster 6126 | ARO:3000590, ARO:3002357, ARO:3000589, ARO:3002359, ARO:3002356, ARO:3002360, ARO:3002355, ARO:3002354, ARO:3003183, ARO:3002362, ARO:3003182, ARO:3000467, ARO:3004093, ARO:3002358 | NDM-1 |
| Cluster 6148 | ARO:3002534, ARO:3002535, ARO:3002533 | AAC(3)-IIb |
| Cluster 6195 | ARO:3002639 | APH(3")-Ib |
| Cluster 6303 | ARO:3000413 | sul3 |
| Cluster 6324 | ARO:3002605, ARO:3002604 | aadA5 |
| Cluster 6352 | ARO:3002602, ARO:3002615, ARO:3002620, ARO:3002621, ARO:3002619, ARO:3002608, ARO:3002611, ARO:3002603, ARO:3003197, ARO:3002612, ARO:3002617, ARO:3002618, ARO:3002601, ARO:3002613 | aadA |
| Cluster 6441 | ARO:3002602, ARO:3002603, ARO:3002621, ARO:3002619, ARO:3002608, ARO:3002611, ARO:3002615, ARO:3003197, ARO:3002612, ARO:3002617, ARO:3002618, ARO:3002601, ARO:3002620 | aadA2 |
| Cluster 6471 | ARO:3002539 | AAC(3)-IV |
| Cluster 6932 | ARO:3000375 | ErmB |
| Cluster 7009 | ARO:3003838 | gadW |
| Cluster 7012 | ARO:3003838 | gadW |
| Cluster 7543 | ARO:3003841 | kdpE |
| Cluster 7731 | ARO:3002683, ARO:3002670, ARO:3002689, | catI |

|  |  |  |
|---|---|---|
|  | ARO:3002670 |  |
| Cluster 8051 | ARO:3002676 | catB3 |
| Cluster 8131 | ARO:3002548, ARO:3002577, ARO:3002592, ARO:3002576, ARO:3002581, ARO:3002546, ARO:3002579, ARO:3002591, ARO:3003676, ARO:3002580, ARO:3002547, ARO:3002582, ARO:3002586, ARO:3002578, ARO:3002594, ARO:3002596 | AAC(6')-Ib3 |
| Cluster 8328 | ARO:3002548, ARO:3002577, ARO:3002592, ARO:3002576, ARO:3002581, ARO:3002546, ARO:3002579, ARO:3002591, ARO:3003676, ARO:3002580, ARO:3002547, ARO:3002582, ARO:3002586, ARO:3002578, ARO:3002594, ARO:3002596 | AAC(6')-Ib' |
| Cluster 8425 | ARO:3002860 | dfrA17 |
| Cluster 8607 | ARO:3002548, ARO:3002577, ARO:3002592, ARO:3002576, ARO:3002581, ARO:3002580, ARO:3002591, ARO:3003676, ARO:3002579, ARO:3002547, ARO:3002582, ARO:3002586, ARO:3002578, ARO:3002594, ARO:3002596 | AAC(6')-IIc |
| Cluster 9305 | ARO:3000516 | emrR |
| Cluster 936 | ARO:3003549 | mdtO |
| Cluster 9427 | ARO:3002895 | SAT-1 |
| Cluster 9594 | ARO:3002863 | dfrA8 |
| Cluster 9729 | ARO:3002858 | dfrA12 |
| Cluster 10085 | ARO:3002861, ARO:3002859 | dfrA14 |
| Cluster 10099 | ARO:3002861 | dfrA5 |
| Cluster 10111 | ARO:3003013, ARO:3002854 | dfrA15 |
| Cluster 10952 | ARO:3000168 | tet(D) |
| Cluster 10956 | ARO:3003209, ARO:3002804, ARO:3002872, | FosA3 |

| | | |
|---|---|---|
| | ARO:3003210 | |
| Cluster 12174 | ARO:3004039 | Escherichia coli emrE |
| Cluster 12195 | ARO:3003836 | qacH |
| Cluster 14077 | ARO:3002831 | vgaC |
| Cluster 14078 | ARO:3002831 | vgaC |
| Cluster 14090 | ARO:3002831 | vgaC |
| Cluster 14095 | ARO:3002831 | vgaC |

[1] RGI may assign multiple CARD hits to some of the genes. The best hit reported by RGI is also listed in the table.

Supplementary Table S7. AUC measurements for predicting drug resistance profiles using 1) core + accessory gene clusters; 2) accessory gene clusters; 3) accessory gene clusters with CARD annotations; and 4) CARD gene clusters using different algorithms (Adaboost, Naïve Bayes, Random Forest, and SVM).

Adaboost

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.66 (0.04)[1] | 0.64 (0.05) | 0.73 (0.03) | 0.79 (0.02) |
| ampicillin\ sulbactam | 0.69 (0.04) | 0.70 (0.04) | 0.76 (0.00) | 0.73 (0.01) |
| piperacillin\ tazobactam | 0.79 (0.04) | 0.80 (0.06) | 0.93 (0.05) | 0.75 (0.00) |
| cefazolin | 0.82 (0.01) | 0.82 (0.01) | 0.88 (0.01) | 0.86 (0.00) |
| ceftazidime | 0.84 (0.01) | 0.83 (0.01) | 0.82 (0.00) | 0.82 (0.00) |
| ceftriaxone | 0.92 (0.03) | 0.90 (0.02) | 0.85 (0.00) | 0.84 (0.02) |
| cefepime | 0.85 (0.01) | 0.86 (0.01) | 0.88 (0.02) | 0.85 (0.01) |
| meropenem | 0.72 (0.02) | 0.72 (0.02) | 0.86 (0.00) | 0.83 (0.02) |
| gentamicin | 0.72 (0.03) | 0.69 (0.06) | 0.77 (0.02) | 0.67 (0.00) |
| tobramycin | 0.69 (0.06) | 0.71 (0.04) | 0.77 (0.09) | 0.54 (0.02) |
| ciprofloxacin | 0.71 (0.04) | 0.72 (0.02) | 0.75 (0.03) | 0.69 (0.02) |
| trimethoprim\ sulfamethoxazole | 0.66 (0.01) | 0.65 (0.02) | 0.82 (0.00) | 0.84 (0.01) |

Naïve Bayes

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.75 (0.00) | 0.75 (0.00) | 0.78 (0.00) | 0.81 (0.00) |
| ampicillin\ sulbactam | 0.67 (0.00) | 0.67 (0.00) | 0.71 (0.00) | 0.71 (0.00) |
| piperacillin\ tazobactam | 0.47 (0.00) | 0.49 (0.00) | 0.72 (0.00) | 0.67 (0.00) |
| cefazolin | 0.63 (0.00) | 0.63 (0.00) | 0.80 (0.00) | 0.77 (0.00) |
| ceftazidime | 0.69 (0.00) | 0.65 (0.00) | 0.70 (0.00) | 0.72 (0.00) |
| ceftriaxone | 0.73 (0.00) | 0.73 (0.00) | 0.82 (0.00) | 0.73 (0.00) |

| | | | | |
|---|---|---|---|---|
| cefepime | 0.64 (0.00) | 0.64 (0.00) | 0.69 (0.00) | 0.69 (0.00) |
| meropenem | 0.72 (0.00) | 0.70 (0.00) | 0.79 (0.00) | 0.71 (0.00) |
| gentamicin | 0.58 (0.00) | 0.55 (0.00) | 0.60 (0.00) | 0.62 (0.00) |
| tobramycin | 0.71 (0.00) | 0.71 (0.00) | 0.85 (0.00) | 0.78 (0.00) |
| ciprofloxacin | 0.71 (0.00) | 0.71 (0.00) | 0.67 (0.00) | 0.85 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.55 (0.00) | 0.52 (0.00) | 0.84 (0.00) | 0.82 (0.00) |

Random Forest

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.67 (0.06) | 0.66 (0.06) | 0.60 (0.08) | 0.59 (0.07) |
| ampicillin\ sulbactam | 0.67 (0.06) | 0.70 (0.06) | 0.65 (0.07) | 0.59 (0.08) |
| piperacillin\ tazobactam | 0.50 (0.07) | 0.53 (0.06) | 0.52 (0.04) | 0.63 (0.08) |
| cefazolin | 0.68 (0.06) | 0.64 (0.05) | 0.73 (0.05) | 0.73 (0.03) |
| ceftazidime | 0.62 (0.04) | 0.67 (0.06) | 0.72 (0.04) | 0.65 (0.03) |
| ceftriaxone | 0.77 (0.05) | 0.78 (0.04) | 0.74 (0.08) | 0.65 (0.08) |
| cefepime | 0.70 (0.05) | 0.70 (0.05) | 0.69 (0.05) | 0.66 (0.05) |
| meropenem | 0.76 (0.04) | 0.79 (0.03) | 0.76 (0.04) | 0.74 (0.05) |
| gentamicin | 0.56 (0.03) | 0.56 (0.04) | 0.57 (0.05) | 0.51 (0.03) |
| tobramycin | 0.66 (0.05) | 0.68 (0.05) | 0.65 (0.08) | 0.67 (0.04) |
| ciprofloxacin | 0.65 (0.08) | 0.65 (0.07) | 0.71 (0.03) | 0.82 (0.04) |
| trimethoprim\ sulfamethoxazole | 0.52 (0.02) | 0.53 (0.05) | 0.75 (0.07) | 0.67 (0.04) |

SVM

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.75 (0.00) | 0.75 (0.00) | 0.64 (0.00) | 0.77 (0.00) |
| ampicillin\ sulbactam | 0.70 (0.00) | 0.70 (0.00) | 0.74 (0.00) | 0.79 (0.00) |

| | | | | |
|---|---|---|---|---|
| piperacillin\ tazobactam | 0.44 (0.00) | 0.44 (0.00) | 0.67 (0.00) | 0.67 (0.00) |
| cefazolin | 0.74 (0.00) | 0.74 (0.00) | 0.92 (0.00) | 0.80 (0.00) |
| ceftazidime | 0.68 (0.00) | 0.70 (0.00) | 0.85 (0.00) | 0.81 (0.00) |
| ceftriaxone | 0.76 (0.00) | 0.76 (0.00) | 0.88 (0.00) | 0.82 (0.00) |
| cefepime | 0.74 (0.00) | 0.77 (0.00) | 0.85 (0.00) | 0.66 (0.00) |
| meropenem | 0.82 (0.00) | 0.80 (0.00) | 0.84 (0.00) | 0.79 (0.00) |
| gentamicin | 0.60 (0.00) | 0.70 (0.00) | 0.78 (0.00) | 0.73 (0.00) |
| tobramycin | 0.70 (0.00) | 0.70 (0.00) | 0.68 (0.00) | 0.80 (0.00) |
| ciprofloxacin | 0.78 (0.00) | 0.74 (0.00) | 0.71 (0.00) | 0.82 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.62 (0.00) | 0.58 (0.00) | 0.87 (0.00) | 0.76 (0.00) |

---

[1] The number outside the parentheses indicates the averaged measurement from 10 repeated experiments, and the number within parentheses represents standard deviation from the 10 repeated estimations. For example, 0.66 (0.04) indicates that the averaged AUC measurements from 10 repeated estimations is 0.66 with standard deviation 0.04.

Supplementary Table S8. $F_1$ score (which is the harmonic mean of precision and recall) measured for predicting drug resistance profiles using 1) core + accessory gene clusters; 2) accessory gene clusters; 3) accessory gene clusters with CARD annotations; and 4) CARD gene clusters using different algorithms (Adaboost, Naïve Bayes, Random Forest, and SVM).

Adaboost

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.51 (0.07)[1] | 0.49 (0.08) | 0.61 (0.04) | 0.70 (0.03) |
| ampicillin\sulbactam | 0.59 (0.05) | 0.61 (0.05) | 0.70 (0.00) | 0.66 (0.01) |
| piperacillin\tazobactam | 0.90 (0.02) | 0.91 (0.03) | 0.97 (0.02) | 0.92 (0.00) |
| cefazolin | 0.82 (0.02) | 0.83 (0.01) | 0.88 (0.01) | 0.86 (0.00) |
| ceftazidime | 0.81 (0.01) | 0.80 (0.01) | 0.79 (0.00) | 0.79 (0.00) |
| ceftriaxone | 0.92 (0.03) | 0.90 (0.02) | 0.86 (0.00) | 0.84 (0.02) |
| cefepime | 0.86 (0.01) | 0.87 (0.01) | 0.89 (0.02) | 0.86 (0.01) |
| meropenem | 0.71 (0.02) | 0.71 (0.02) | 0.85 (0.00) | 0.82 (0.02) |
| gentamicin | 0.89 (0.02) | 0.89 (0.03) | 0.85 (0.01) | 0.83 (0.00) |
| tobramycin | 0.90 (0.03) | 0.90 (0.02) | 0.93 (0.02) | 0.85 (0.03) |
| ciprofloxacin | 0.74 (0.04) | 0.74 (0.02) | 0.74 (0.04) | 0.63 (0.01) |
| trimethoprim\sulfamethoxazole | 0.52 (0.02) | 0.52 (0.03) | 0.73 (0.00) | 0.78 (0.02) |

Naïve Bayes

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.67 (0.00) | 0.67 (0.00) | 0.67 (0.00) | 0.70 (0.00) |
| ampicillin\sulbactam | 0.53 (0.00) | 0.53 (0.00) | 0.64 (0.00) | 0.64 (0.00) |
| piperacillin\tazobactam | 0.74 (0.00) | 0.69 (0.00) | 0.89 (0.00) | 0.89 (0.00) |
| cefazolin | 0.67 (0.00) | 0.67 (0.00) | 0.79 (0.00) | 0.75 (0.00) |
| ceftazidime | 0.63 (0.00) | 0.60 (0.00) | 0.63 (0.00) | 0.67 (0.00) |

| | | | | |
|---|---|---|---|---|
| ceftriaxone | 0.73 (0.00) | 0.73 (0.00) | 0.83 (0.00) | 0.77 (0.00) |
| cefepime | 0.65 (0.00) | 0.65 (0.00) | 0.73 (0.00) | 0.74 (0.00) |
| meropenem | 0.74 (0.00) | 0.73 (0.00) | 0.79 (0.00) | 0.72 (0.00) |
| gentamicin | 0.73 (0.00) | 0.68 (0.00) | 0.84 (0.00) | 0.86 (0.00) |
| tobramycin | 0.86 (0.00) | 0.86 (0.00) | 0.93 (0.00) | 0.93 (0.00) |
| ciprofloxacin | 0.76 (0.00) | 0.76 (0.00) | 0.69 (0.00) | 0.86 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.35 (0.00) | 0.32 (0.00) | 0.75 (0.00) | 0.73 (0.00) |

Random Forest

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.51 (0.11) | 0.50 (0.11) | 0.38 (0.14) | 0.36 (0.15) |
| ampicillin\ sulbactam | 0.50 (0.14) | 0.57 (0.11) | 0.52 (0.11) | 0.38 (0.18) |
| piperacillin\ tazobactam | 0.83 (0.03) | 0.85 (0.02) | 0.85 (0.01) | 0.88 (0.02) |
| cefazolin | 0.68 (0.06) | 0.62 (0.05) | 0.74 (0.04) | 0.71 (0.03) |
| ceftazidime | 0.50 (0.05) | 0.57 (0.08) | 0.64 (0.06) | 0.55 (0.04) |
| ceftriaxone | 0.76 (0.05) | 0.77 (0.03) | 0.75 (0.07) | 0.67 (0.07) |
| cefepime | 0.73 (0.05) | 0.73 (0.04) | 0.72 (0.04) | 0.70 (0.04) |
| meropenem | 0.77 (0.05) | 0.79 (0.03) | 0.76 (0.04) | 0.75 (0.05) |
| gentamicin | 0.87 (0.01) | 0.86 (0.02) | 0.86 (0.02) | 0.85 (0.02) |
| tobramycin | 0.92 (0.01) | 0.92 (0.02) | 0.92 (0.02) | 0.93 (0.01) |
| ciprofloxacin | 0.67 (0.08) | 0.67 (0.07) | 0.73 (0.03) | 0.82 (0.04) |
| trimethoprim\ sulfamethoxazole | 0.12 (0.09) | 0.14 (0.14) | 0.65 (0.10) | 0.52 (0.08) |

SVM

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.67 (0.00) | 0.67 (0.00) | 0.50 (0.00) | 0.67 (0.00) |

| | | | | |
|---|---|---|---|---|
| ampicillin\ sulbactam | 0.57 (0.00) | 0.57 (0.00) | 0.67 (0.00) | 0.74 (0.00) |
| piperacillin\ tazobactam | 0.79 (0.00) | 0.79 (0.00) | 0.89 (0.00) | 0.89 (0.00) |
| cefazolin | 0.74 (0.00) | 0.74 (0.00) | 0.91 (0.00) | 0.80 (0.00) |
| ceftazidime | 0.61 (0.00) | 0.65 (0.00) | 0.82 (0.00) | 0.76 (0.00) |
| ceftriaxone | 0.75 (0.00) | 0.75 (0.00) | 0.88 (0.00) | 0.82 (0.00) |
| cefepime | 0.75 (0.00) | 0.78 (0.00) | 0.84 (0.00) | 0.70 (0.00) |
| meropenem | 0.83 (0.00) | 0.81 (0.00) | 0.84 (0.00) | 0.79 (0.00) |
| gentamicin | 0.88 (0.00) | 0.91 (0.00) | 0.88 (0.00) | 0.91 (0.00) |
| tobramycin | 0.94 (0.00) | 0.94 (0.00) | 0.91 (0.00) | 0.96 (0.00) |
| ciprofloxacin | 0.79 (0.00) | 0.74 (0.00) | 0.71 (0.00) | 0.81 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.38 (0.00) | 0.33 (0.00) | 0.81 (0.00) | 0.67 (0.00) |

---

[1] The number outside the parentheses indicates the averaged measurement from 10 repeated experiments, and the number within parentheses represents standard deviation from the 10 repeated estimations.

Supplementary Table S9. Precision measurements for predicting drug resistance profiles using 1) core + accessory gene clusters; 2) accessory gene clusters; 3) accessory gene clusters with CARD annotations; and 4) CARD gene clusters using different algorithms (Adaboost, Naïve Bayes, Random Forest, and SVM).

Adaboost

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.55 (0.06)[1] | 0.51 (0.07) | 0.57 (0.02) | 0.66 (0.01) |
| ampicillin\ sulbactam | 0.66 (0.07) | 0.70 (0.08) | 0.62 (0.00) | 0.62 (0.03) |
| piperacillin\ tazobactam | 0.89 (0.02) | 0.89 (0.03) | 0.95 (0.03) | 0.85 (0.00) |
| cefazolin | 0.79 (0.01) | 0.79 (0.00) | 0.83 (0.02) | 0.83 (0.00) |
| ceftazidime | 0.81 (0.02) | 0.80 (0.02) | 0.79 (0.00) | 0.79 (0.00) |
| ceftriaxone | 0.87 (0.03) | 0.86 (0.02) | 0.79 (0.00) | 0.81 (0.02) |
| cefepime | 0.82 (0.02) | 0.84 (0.02) | 0.89 (0.03) | 0.82 (0.01) |
| meropenem | 0.70 (0.01) | 0.70 (0.01) | 0.82 (0.00) | 0.81 (0.02) |
| gentamicin | 0.85 (0.01) | 0.83 (0.03) | 0.90 (0.02) | 0.83 (0.00) |
| tobramycin | 0.89 (0.02) | 0.89 (0.02) | 0.91 (0.04) | 0.83 (0.01) |
| ciprofloxacin | 0.65 (0.02) | 0.65 (0.01) | 0.72 (0.02) | 0.75 (0.04) |
| trimethoprim\ sulfamethoxazole | 0.56 (0.03) | 0.55 (0.03) | 0.65 (0.00) | 0.79 (0.03) |

Naïve Bayes

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 1.00 (0.00) | 1.00 (0.00) | 0.54 (0.00) | 0.58 (0.00) |
| ampicillin\ sulbactam | 0.80 (0.00) | 0.80 (0.00) | 0.58 (0.00) | 0.58 (0.00) |
| piperacillin\ tazobactam | 0.72 (0.00) | 0.73 (0.00) | 0.84 (0.00) | 0.81 (0.00) |
| cefazolin | 0.59 (0.00) | 0.59 (0.00) | 0.81 (0.00) | 0.80 (0.00) |
| ceftazidime | 0.63 (0.00) | 0.57 (0.00) | 0.69 (0.00) | 0.71 (0.00) |
| ceftriaxone | 0.71 (0.00) | 0.71 (0.00) | 0.75 (0.00) | 0.65 (0.00) |

| | | | | |
|---|---|---|---|---|
| cefepime | 0.65 (0.00) | 0.65 (0.00) | 0.67 (0.00) | 0.65 (0.00) |
| meropenem | 0.64 (0.00) | 0.62 (0.00) | 0.72 (0.00) | 0.66 (0.00) |
| gentamicin | 0.80 (0.00) | 0.78 (0.00) | 0.79 (0.00) | 0.80 (0.00) |
| tobramycin | 0.90 (0.00) | 0.90 (0.00) | 0.95 (0.00) | 0.91 (0.00) |
| ciprofloxacin | 0.62 (0.00) | 0.62 (0.00) | 0.63 (0.00) | 0.80 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.40 (0.00) | 0.33 (0.00) | 0.63 (0.00) | 0.65 (0.00) |

Random Forest

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.88 (0.16) | 0.88 (0.15) | 0.54 (0.21) | 0.53 (0.16) |
| ampicillin\ sulbactam | 0.96 (0.13) | 0.96 (0.13) | 0.64 (0.13) | 0.57 (0.24) |
| piperacillin\ tazobactam | 0.74 (0.03) | 0.75 (0.02) | 0.75 (0.02) | 0.79 (0.04) |
| cefazolin | 0.65 (0.06) | 0.64 (0.05) | 0.71 (0.06) | 0.73 (0.06) |
| ceftazidime | 0.62 (0.07) | 0.71 (0.09) | 0.75 (0.03) | 0.65 (0.05) |
| ceftriaxone | 0.77 (0.08) | 0.79 (0.06) | 0.71 (0.07) | 0.61 (0.07) |
| cefepime | 0.69 (0.04) | 0.69 (0.06) | 0.67 (0.04) | 0.64 (0.04) |
| meropenem | 0.70 (0.04) | 0.72 (0.02) | 0.71 (0.03) | 0.68 (0.04) |
| gentamicin | 0.77 (0.01) | 0.77 (0.02) | 0.78 (0.02) | 0.75 (0.01) |
| tobramycin | 0.87 (0.02) | 0.88 (0.02) | 0.86 (0.03) | 0.87 (0.01) |
| ciprofloxacin | 0.62 (0.07) | 0.62 (0.06) | 0.66 (0.03) | 0.79 (0.05) |
| trimethoprim\ sulfamethoxazole | 0.43 (0.38) | 0.48 (0.45) | 0.72 (0.10) | 0.68 (0.11) |

SVM

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 1.00 (0.00) | 1.00 (0.00) | 0.50 (0.00) | 0.60 (0.00) |
| ampicillin\ sulbactam | 1.00 (0.00) | 1.00 (0.00) | 0.75 (0.00) | 0.78 (0.00) |

| | | | | |
|---|---|---|---|---|
| piperacillin\ tazobactam | 0.71 (0.00) | 0.71 (0.00) | 0.81 (0.00) | 0.81 (0.00) |
| cefazolin | 0.72 (0.00) | 0.72 (0.00) | 0.89 (0.00) | 0.78 (0.00) |
| ceftazidime | 0.65 (0.00) | 0.67 (0.00) | 0.93 (0.00) | 0.87 (0.00) |
| ceftriaxone | 0.75 (0.00) | 0.75 (0.00) | 0.83 (0.00) | 0.78 (0.00) |
| cefepime | 0.75 (0.00) | 0.76 (0.00) | 0.89 (0.00) | 0.65 (0.00) |
| meropenem | 0.75 (0.00) | 0.74 (0.00) | 0.79 (0.00) | 0.73 (0.00) |
| gentamicin | 0.79 (0.00) | 0.83 (0.00) | 0.90 (0.00) | 0.85 (0.00) |
| tobramycin | 0.88 (0.00) | 0.88 (0.00) | 0.88 (0.00) | 0.92 (0.00) |
| ciprofloxacin | 0.73 (0.00) | 0.71 (0.00) | 0.67 (0.00) | 0.79 (0.00) |
| trimethoprim\ sulfamethoxazole | 1.00 (0.00) | 0.60 (0.00) | 0.79 (0.00) | 0.64 (0.00) |

[1] The number outside the parentheses indicates the averaged measurement from 10 repeated experiments, and the number within parentheses represents standard deviation from the 10 repeated estimations.

Supplementary Table S10. Recall measurements for predicting drug resistance profiles using 1) core + accessory gene clusters; 2) accessory gene clusters; 3) accessory gene clusters with CARD annotations; and 4) CARD gene clusters using different algorithms (Adaboost, Naïve Bayes, Random Forest, and SVM).

Adaboost

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.47 (0.08)[1] | 0.47 (0.10) | 0.66 (0.06) | 0.74 (0.04) |
| ampicillin\ sulbactam | 0.53 (0.05) | 0.54 (0.05) | 0.80 (0.00) | 0.70 (0.00) |
| piperacillin\ tazobactam | 0.92 (0.03) | 0.93 (0.04) | 1.00 (0.00) | 1.00 (0.00) |
| cefazolin | 0.86 (0.03) | 0.87 (0.02) | 0.94 (0.00) | 0.88 (0.00) |
| ceftazidime | 0.80 (0.02) | 0.81 (0.03) | 0.79 (0.00) | 0.79 (0.00) |
| ceftriaxone | 0.97 (0.03) | 0.96 (0.03) | 0.94 (0.00) | 0.86 (0.03) |
| cefepime | 0.90 (0.00) | 0.90 (0.00) | 0.89 (0.02) | 0.90 (0.00) |
| meropenem | 0.72 (0.03) | 0.72 (0.03) | 0.88 (0.00) | 0.84 (0.02) |
| gentamicin | 0.95 (0.04) | 0.96 (0.03) | 0.80 (0.02) | 0.83 (0.00) |
| tobramycin | 0.91 (0.04) | 0.90 (0.05) | 0.95 (0.02) | 0.88 (0.05) |
| ciprofloxacin | 0.85 (0.06) | 0.86 (0.03) | 0.77 (0.06) | 0.54 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.49 (0.04) | 0.48 (0.04) | 0.85 (0.00) | 0.77 (0.00) |

Naïve Bayes

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.50 (0.00) | 0.50 (0.00) | 0.88 (0.00) | 0.88 (0.00) |
| ampicillin\ sulbactam | 0.40 (0.00) | 0.40 (0.00) | 0.70 (0.00) | 0.70 (0.00) |
| piperacillin\ tazobactam | 0.76 (0.00) | 0.65 (0.00) | 0.94 (0.00) | 1.00 (0.00) |
| cefazolin | 0.76 (0.00) | 0.76 (0.00) | 0.76 (0.00) | 0.71 (0.00) |
| ceftazidime | 0.63 (0.00) | 0.63 (0.00) | 0.58 (0.00) | 0.63 (0.00) |
| ceftriaxone | 0.75 (0.00) | 0.75 (0.00) | 0.94 (0.00) | 0.94 (0.00) |

| | | | | |
|---|---|---|---|---|
| cefepime | 0.65 (0.00) | 0.65 (0.00) | 0.80 (0.00) | 0.85 (0.00) |
| meropenem | 0.88 (0.00) | 0.88 (0.00) | 0.88 (0.00) | 0.81 (0.00) |
| gentamicin | 0.67 (0.00) | 0.60 (0.00) | 0.90 (0.00) | 0.93 (0.00) |
| tobramycin | 0.82 (0.00) | 0.82 (0.00) | 0.91 (0.00) | 0.95 (0.00) |
| ciprofloxacin | 1.00 (0.00) | 1.00 (0.00) | 0.77 (0.00) | 0.92 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.31 (0.00) | 0.31 (0.00) | 0.92 (0.00) | 0.85 (0.00) |

Random Forest

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.36 (0.09) | 0.35 (0.10) | 0.31 (0.15) | 0.28 (0.13) |
| ampicillin\ sulbactam | 0.35 (0.12) | 0.41 (0.10) | 0.44 (0.11) | 0.29 (0.15) |
| piperacillin\ tazobactam | 0.95 (0.05) | 0.98 (0.04) | 0.99 (0.02) | 1.00 (0.00) |
| cefazolin | 0.71 (0.07) | 0.62 (0.08) | 0.78 (0.05) | 0.71 (0.04) |
| ceftazidime | 0.43 (0.06) | 0.47 (0.09) | 0.57 (0.09) | 0.48 (0.05) |
| ceftriaxone | 0.75 (0.05) | 0.76 (0.03) | 0.80 (0.09) | 0.74 (0.07) |
| cefepime | 0.78 (0.07) | 0.78 (0.05) | 0.77 (0.05) | 0.77 (0.06) |
| meropenem | 0.85 (0.07) | 0.89 (0.05) | 0.83 (0.06) | 0.83 (0.08) |
| gentamicin | 0.99 (0.02) | 0.97 (0.04) | 0.96 (0.03) | 0.98 (0.03) |
| tobramycin | 0.99 (0.02) | 0.96 (0.04) | 0.98 (0.02) | 0.99 (0.02) |
| ciprofloxacin | 0.72 (0.10) | 0.72 (0.09) | 0.82 (0.04) | 0.87 (0.06) |
| trimethoprim\ sulfamethoxazole | 0.07 (0.06) | 0.08 (0.09) | 0.60 (0.11) | 0.43 (0.07) |

SVM

| Drugs | core + accessory gene clusters | accessory gene clusters | accessory gene clusters with CARD annotation | CARD gene clusters |
|---|---|---|---|---|
| ampicillin | 0.50 (0.00) | 0.50 (0.00) | 0.50 (0.00) | 0.75 (0.00) |
| ampicillin\ sulbactam | 0.40 (0.00) | 0.40 (0.00) | 0.60 (0.00) | 0.70 (0.00) |

| | | | | |
|---|---|---|---|---|
| piperacillin\tazobactam | 0.88 (0.00) | 0.88 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| cefazolin | 0.76 (0.00) | 0.76 (0.00) | 0.94 (0.00) | 0.82 (0.00) |
| ceftazidime | 0.58 (0.00) | 0.63 (0.00) | 0.74 (0.00) | 0.68 (0.00) |
| ceftriaxone | 0.75 (0.00) | 0.75 (0.00) | 0.94 (0.00) | 0.88 (0.00) |
| cefepime | 0.75 (0.00) | 0.80 (0.00) | 0.80 (0.00) | 0.75 (0.00) |
| meropenem | 0.92 (0.00) | 0.88 (0.00) | 0.88 (0.00) | 0.85 (0.00) |
| gentamicin | 1.00 (0.00) | 1.00 (0.00) | 0.87 (0.00) | 0.97 (0.00) |
| tobramycin | 1.00 (0.00) | 1.00 (0.00) | 0.95 (0.00) | 1.00 (0.00) |
| ciprofloxacin | 0.85 (0.00) | 0.77 (0.00) | 0.77 (0.00) | 0.85 (0.00) |
| trimethoprim\sulfamethoxazole | 0.23 (0.00) | 0.23 (0.00) | 0.85 (0.00) | 0.69 (0.00) |

[1] The number outside the parentheses indicates the averaged measurement from 10 repeated experiments, and the number within parentheses represents standard deviation from the 10 repeated estimations.

Supplementary Table S11. AUC measurements for predicting drug resistance profiles by SVM algorithm (with radial basis function kernel) using leave-one-out approach with either the 68 accessory gene clusters with CARD annotations (acc/card) or GA-selected gene clusters.

| | 68 acc/card geneclusters | GA-selected gene clusters |
|---|---|---|
| Ampicillin | 0.64 (0.00)[1] | 0.97 (0.00) |
| Gentamicin | 0.74 (0.00) | 0.98 (0.00) |
| Trimethoprim/Sulfamethoxazole | 0.67 (0.00) | 0.94 (0.00) |
| Ciprofloxacin | 0.92 (0.00) | 0.93 (0.00) |
| Ampicillin/Sulbactam | 0.85 (0.00) | 0.97 (0.00) |
| Cefazolin | 0.88 (0.00) | 1.0 (0.00) |
| Cefepime | 0.85 (0.00) | 0.98 (0.00) |
| Ceftazidime | 0.84 (0.00) | 0.96 (0.00) |
| Ceftriaxone | 0.78 (0.00) | 1.0 (0.00) |
| Meropenem | 0.68 (0.00) | 0.95 (0.00) |
| Piperacillin/Tazobactam | 0.71 (0.00) | 1.0 (0.00) |
| Tobramycin | 0.87 (0.00) | 1.0 (0.00) |

[1] The number outside the parentheses indicates the averaged measurement from 10 repeated experiments, and the number within parentheses represents standard deviation from the 10 repeated estimations.

Supplementary Table S12. Gene clusters selected by the genetic algorithm for different antibiotic drugs.

| Gene cluster | Mer[1] | Cipro[1] | Gen[1] | Amp[1] | Cefa[1] | Trime_Sulf[1] | Amp_Sul[1] | Cefta[1] | Cefe[1] | Ceftr[1] | Pip_Taz[1] | Tob[1] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cluster 1202 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Cluster 1513 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| Cluster 1549 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Cluster 1564 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| Cluster 1961 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| Cluster 2387 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| Cluster 2855 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Cluster 2965 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Cluster 3112 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Cluster 3154 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Cluster 3361 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| Cluster 3397 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| Cluster 3408 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cluster 3412 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| Cluster 3521 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| Cluster 4155 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| Cluster 4265 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 |
| Cluster 4326 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cluster 4764 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cluster 5292 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cluster 5317 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| Cluster 5512 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| Cluster 5572 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| Cluster 5574 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cluster 5575 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| Cluster 5686 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| Cluster 5687 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| Cluster 5700 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Cluster 5871 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Cluster 5926 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Cluster 5972 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cluster 6005 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Cluster 6010 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| Cluster 6105 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Cluster 6126 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| Cluster 6148 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| Cluster 6195 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cluster 6303 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Cluster 6324 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 |
| Cluster 6352 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Cluster 6441 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| Cluster 6932 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| Cluster 7009 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Cluster 7012 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| Cluster 7543 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Cluster 7731 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| Cluster 8051 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| Cluster 8131 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Cluster 8328 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| Cluster 8425 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cluster 8607 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cluster 9305 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| Cluster 936 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| Cluster 9427 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| Cluster 9594 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| Cluster 9729 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| Cluster 10085 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| Cluster 10099 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cluster 10111 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| Cluster 10952 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Cluster 10956 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| Cluster 12174 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Cluster 12195 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Cluster 14077 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| Cluster 14078 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| Cluster 14090 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Cluster 14095 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| Total number | 32 | 20 | 24 | 24 | 23 | 35 | 22 | 32 | 28 | 36 | 27 | 17 |

[1] Mer: meropenem; Cipro: ciprofloxacin; Gen: gentamicin; Amp: ampicillin; Cefa: Cefazolin; Trime_Sulf: trimethoprim/sulfamethoxazole; Amp_Sul: ampicillin/sulbactam; Cefta: ceftazidime; Cefe: cefepime; Ceftr: ceftriaxone; Pip_Taz: piperacillin/tazobactam; Tob: tobramycin

Supplementary Table S13. Number of gene clusters identified by Scoary (with the pan-genome built using Roary) to be associated with the AMR phenotypes ($P < 0.05$). The best hit CARD annotations of the genes are also listed.

| Drug | Identified gene#[1] | Annotations of CARD gene clusters |
|---|---|---|
| ampiciilin | 285 | *CMY-2*, *AAC(6)-Ib3*, *CTX-M-15* |
| ampicillin/sulbactam | 364 | *CTX-M-15*, *CMY-2*, *AAC(6)-Ib3*, *sul1* |
| piperacillin/tazobactam | 183 | *AAC(6)-Ib3*, *sul1*, *APH(3)-Ia* |
| cefazolin | 631 | *APH(6)-Id*, *sul1*, *AAC(6)-IIc*, *AAC(3)-IIb*, *EreA2*, *MCR-3*, *SHV-160*, *KPC-3*, *PmrE*, *sul2* |
| ceftazidime | 357 | *KPC-3*, *Escherichia coli UhpT* with mutation conferring resistance to fosfomycin, *CTX-M-15*, *APH(6)-Id*, *CMY-2* |
| cefepime | 844 | *KPC-3*, *APH(6)-Id*, *Escherichia coli UhpT* with mutation conferring resistance to fosfomycin, *mdtO*, *Escherichia coli GlpT* with mutation conferring resistance to fosfomycin, *NDM-1*, *AAC(6)-Ib3* |
| ceftriaxone | 841 | *APH(6)-Id*, *KPC-3*, *Escherichia coli UhpT* with mutation conferring resistance to fosfomycin, *mdtO*, *sul2* |
| meropenem | 1237 | *KPC-3*, *APH(6)-Id*, *sul1*, *AAC(6)-IIc*, *AAC(3)-IIb*, *EreA2*, *MCR-3*, *AcrF*, *SHV-* |

| | | |
|---|---|---|
| | | *160*, *sul2*, *mdtM* |
| gentamicin | 539 | *aadA*, *CTX-M-15*, *AAC(3)-IIa*, *AAC(6)-Ib3*, *TEM-1*, *CMY-2*, *OXA-2*, *AAC(3)-IIa* |
| tobramycin | 783 | *AAC(6)-Ib3*, *sul1*, *CMY-2*, *NDM-1*, *gadX*, *OXA-2* |
| ciprofloxacin | 335 | *CTX-M-15* |
| trimethoprim/sulfamethoxazole | 271 | *dfrA17*, *aadA5*, *APH(3)-Ib*, *Mrx*, *mphA*, *sul2*, *KPC-3*, *PmrE*, *tet(A)*, *vgaC*, *sul1* |

---

[1] These are the number of gene clusters identified by Scoary to be assocated with the AMR phenotypes (P < 0.05).

Supplementary Table S14. Evaluation of SVM prediction performances of antimicrobial resistance activities using gene sets discovered by Scoary.

| Drugs | Gene cluster sets discovered by Scoary[1] | Gene cluster sets with CARD annotations discovered by Scoary[2] |
|---|---|---|
| ampicillin | 0.75 (0.00) | 0.79 (0.00) |
| ampicillin\ sulbactam | 0.77 (0.00) | 0.89 (0.00) |
| piperacillin\ tazobactam | 0.75 (0.00) | 0.64 (0.00) |
| cefazolin | 0.83 (0.00) | 0.54 (0.00) |
| ceftazidime | 0.83 (0.00) | 0.92 (0.00) |
| ceftriaxone | 0.85 (0.00) | 0.88 (0.00) |
| cefepime | 0.74 (0.00) | 0.87 (0.00) |
| meropenem | 0.79 (0.00) | 0.75 (0.00) |
| gentamicin | 0.85 (0.00) | 0.68 (0.00) |
| tobramycin | 0.80 (0.00) | 0.92 (0.00) |
| ciprofloxacin | 0.93 (0.00) | 0.75 (0.00) |
| trimethoprim\ sulfamethoxazole | 0.76 (0.00) | 0.87 (0.00) |

[1] The number outside the parentheses indicates the averaged measurement from 10 repeated experiments, and the number within parentheses represents standard deviation from the 10 repeated estimations.

[1] Gene cluster sets with CARD annotations indicate that only gene clusters with any RGI/CARD annotations are included in the evaluation process.