Supplementary material

PlasmidCall: predicting whether resistance genes in short-read assemblies lie on plasmids or chromosomes, tested in 432 carbapenem-resistant Enterobacterales

Vahhab Piranfar

Contents: Supplementary Methods S1-S3; Supplementary Results S1-S4; Tables S1-S3; References. Citations to Tables 1 and 2 and Figures 1-3 refer to the main article. Reference numbers refer to the reference list of this Supplement.

Supplementary Methods

S1. Study design, cohort, component classifiers and reference standard

Study design

PlasmidCall v1.2-General is the principal model: a logistic-regression meta-classifier that classifies short-read contigs as plasmid or chromosome from the calls of twelve existing plasmid classifiers, the component classifiers. Its precision was the primary endpoint, with coverage co-primary. Four related gradient-boosting models, GB-Contig, GB-Panel, GB-Calls and GB-Full, here called the development models, constitute the model-development component of the study: they combine component-classifier outputs with contig, assembly-graph and isolate information in different input blocks, and their prespecified comparisons form an ablation ladder. They are alternative configurations of one estimator, not successive versions of PlasmidCall. PlasmidCall v1.1, an earlier PlasmidCall model, and a router that passes each contig to v1.1 or v1.2-General were included as secondary reference methods. The fitted models had their parameters and thresholds set on development data, whereas the router used its predefined routing rule. Without refitting or threshold changes, the models and the router were evaluated in an independent evaluation cohort against reference labels from hybrid assemblies of the same isolates, and the component classifiers, which supply the models' inputs, were evaluated alongside them as comparators.

Independent evaluation cohort

All methods were evaluated in the independent evaluation cohort, the principal evaluation dataset of this study, which comprised every BioSample in NCBI BioProject PRJNA1133668 with both an Illumina and an Oxford Nanopore Technologies (ONT) run: 432 of 434 BioSamples. No isolate was excluded for any other reason. The isolates are carbapenem-resistant Enterobacterales collected at a New York City hospital in 2009-2018 and sequenced on Illumina MiSeq and ONT MinION instruments[1].

By the depositors' species labels, the cohort comprises *Klebsiella* (313, including 288 *K. pneumoniae*), *Enterobacter* (66), *Escherichia coli* (17), *Citrobacter* (16), *Serratia marcescens* (16), *Proteus mirabilis* (2), *Morganella morganii* (1) and *Raoultella ornithinolytica* (1). These labels define the genus strata. Each isolate contributed one Illumina run; for SAMN08720235, the only isolate with two, SRR29750977 was used and SRR29751248 was not.

No BioSample or run accession is shared with any isolate used to train or tune an evaluated model.

Short-read assembly and component classifiers

Reads were obtained from the European Nucleotide Archive. Illumina pairs of every cohort isolate were capped at an expected depth of 100× by subsampling with SeqKit v2.8.2[2], the same cap applied to 151 of the development models' 189 training isolates; the other 38, like v1.2-General's training isolates, were assembled at native depth. Depth was estimated from sequenced bases and the NCBI species-level genome size. Each isolate was assembled from these reads alone with Unicycler v0.5.1[3] (SPAdes v3.15.5[4]) at default settings. Contigs of at least 1 kb were eligible for all contig-level analyses.

Twelve existing plasmid classifiers, the component classifiers, ran in pinned containers with pinned databases (Table S1). Their outputs are the inputs of the models described below, and each classifier was also evaluated as a comparator. All isolates were run under the configuration that generated the training inputs of v1.2-General and the development models.

A failed run was repeated up to twice. A single parser then mapped outputs to seven states:

- plasmid;
- chromosome;
- unknown, unclassified and repeat, which are abstentions;
- FAILED, after three failed attempts;
- MISSING, when the expected output was absent.

Only the plasmid state is a plasmid call. FAILED and MISSING are never converted into calls.

Resistance determinant annotation

Resistance determinants were annotated with AMRFinderPlus v4.2.7 (database 2026-08-07.1)[5]. A contig was ARG-bearing if it carried at least one determinant of type and subtype AMR, in the core scope, with a drug class other than efflux. Whether a contig was ARG-bearing, and its number of such determinants, were also inputs to development models GB-Contig, GB-Calls and GB-Full.

Reference standard

Reference labels came from hybrid assemblies of the same isolates. Unicycler v0.5.1 built one hybrid assembly per isolate from all of its ONT reads and the same depth-capped Illumina reads. One isolate's two ONT runs were concatenated in accession order.

Hybrid sequences were labelled by four rules applied in order:

1. The longest sequence is chromosome.

2. Any other sequence of at least 1 Mb is chromosome.

3. Any other closed circular sequence shorter than 1 Mb is plasmid.

4. Any other sequence in the chromosome's assembly-graph component is chromosome.

Remaining sequences were unlabelled. The rules were set after inspecting the hybrid assemblies of two cohort isolates (SAMN40477277 and SAMN40477344), before any prediction had been compared with a reference label. No plasmid database contributed to any label.

Each eligible contig was aligned with minimap2 v2.28-r1209[6] (asm5 preset) to the hybrid sequences of its own isolate. Circular sequences were doubled so that alignments across the origin remained whole. Alignments of at least 95% identity were merged per target.

A contig took its best target's label if three conditions held: the target was labelled, it covered at least 80% of the contig, and every target within 0.10 of that coverage carried the same label. Otherwise the contig was unresolved. Unresolved contigs were excluded from every denominator and never counted as negatives.

All predictions were fixed before any long read was retrieved, and the reference standard was built without access to them.

Database exposure

Exposure was measured against the PLSDB-derived database distributed with PLASMe[7,8]: 33,125 plasmids of 1-350 kb (doi:10.5281/zenodo.8046934). Each eligible contig was aligned with BLAST+ v2.13.0 megablast[9] (E-value $\leq 1 \times 10^{-20}$, identity ≥80%, at most 50 targets). A contig was database-exposed if the fraction of its positions covered by the union of alignments of at least 99% identity was at least 0.95. All other contigs, including those with no alignment, were database-absent. Because the database contains no plasmid longer than 350 kb, contigs from larger plasmids could be database-exposed only through sequence shared with the smaller plasmids in the database, so exposure partly reflects plasmid size.

Exposure was computed without reference labels or predictions, and it measures representation in one database, not novelty. It was used:

- to stratify results;
- to check the reference standard, by comparing exposure between plasmid- and chromosome-labelled contigs.

S2. Model specifications and thresholds

PlasmidCall v1.2-General

PlasmidCall v1.2-General, the principal model, is a logistic-regression meta-classifier whose only inputs are the twelve component classifiers' calls, encoded as 102 features (Table S2).

Valid calls exclude abstentions and FAILED and MISSING states. Features are median-imputed and standardised before L2-regularised fitting (C = 1). When a classifier returns a state never observed for it in training, its eight classifier-specific features are set to their training means and contribute nothing to the score.

PlasmidCall v1.2-General was fitted on all 3,811 truth-labelled contigs (626 plasmid, 130 ARG-bearing) of 39 *E. coli* isolates; training labels came from public complete genome assemblies of the same BioSamples, with replicon roles taken from NCBI assembly reports. Contigs scoring at least 0.9285 are called plasmid. This threshold is the smallest score at which leave-one-isolate-out development precision reached 0.970 with at least ten plasmid calls. 501 development contigs were called plasmid at this threshold. v1.2-General was not refitted or recalibrated for the evaluation cohort, and its threshold was not changed.

Secondary reference methods: PlasmidCall v1.1 and the router

PlasmidCall v1.1 is a histogram gradient-boosting classifier that uses similar call features plus the number of abstaining classifiers, contig length and the assembly's number of contigs of at least 1 kb. It was fitted on the 1,460 resistance-gene-bearing contigs (280 plasmid) of a published benchmark of 250 genomes from seven genera[10]. Its thresholds are 0.9524 (standard) and 0.9605 (high-confidence).

The router sends ARG-bearing contigs to v1.1 and all others to v1.2-General, whose specified domain excludes ARG-bearing contigs. Contigs whose annotation failed receive no routed call. v1.1 was not refitted to these data, and neither its thresholds nor the routing rule were changed.

Four gradient-boosting models

The four development models, GB-Contig, GB-Panel, GB-Calls and GB-Full, were built to test alternative combinations of classifier-derived and contextual information. They share one estimator: histogram gradient boosting with 1,000 iterations, no early stopping, otherwise default settings and a fixed seed.

The four models differ only in their input blocks and form a prespecified ablation ladder (Table S2).

The ablation ladder (Supplementary Methods S3, Endpoints and comparisons) evaluates three contrasts between them in the cohort. GB-Contig→GB-Calls adds only the ten further classifiers. GB-Calls→GB-Full adds only scores, neighbour votes and isolate context. GB-Panel→GB-Calls is a block exchange, not a single addition: it adds the contig block and removes scores and neighbour votes.

Training and thresholds of the gradient-boosting models

The four models share one training set: 13,195 truth-labelled contigs (2,889 plasmid) from 189 isolates, labelled in the same way as those of v1.2-General. Each balanced threshold is the score that maximises the smaller of recall and precision over pooled out-of-fold development scores, with ties resolved to the lower score. Those scores came from five-fold cross-validation grouped by genomic kinship. The thresholds are 0.0951 (GB-Contig), 0.4004 (GB-Panel), 0.3116 (GB-Calls) and 0.3021 (GB-Full).

The balanced rule was chosen on development data, and thresholds from four-fifths-data models were applied to full-data refits. No development model was refitted to the cohort or had its threshold changed. In models with a taxon feature, isolates of genera absent from training (*Proteus*, *Morganella* and *Raoultella*) were scored with taxon treated as missing.

S3. Endpoints, comparisons and statistical analysis

Endpoints and comparisons

Metrics were computed per contig on truth-resolved eligible contigs, with plasmid as the positive class:

- precision (TP/[TP+FP]);
- recall (TP/[TP+FN]);
- F1;
- coverage (the fraction of contigs receiving a decision).

The primary all-contig reading counts every non-plasmid state as not plasmid. The decided-only reading removes abstentions and failures. Precision is identical under both readings.

The planned primary endpoint was v1.2-General precision, with coverage co-primary. It was compared with the earlier estimate, a previously reported performance estimate of 0.9770 (isolate-clustered 95% CI 0.9642-0.9872), computed on 9,371 truth-resolved contigs from 150 genomes, one per isolate, 25 each of *Citrobacter* spp., *Enterobacter* spp., *Klebsiella pneumoniae*, *Serratia* spp., *Enterococcus faecalis* and *Enterococcus faecium*[11]. The earlier interval resampled isolates, whereas the primary interval here resamples clusters of isolates defined by average nucleotide identity (ANI; Statistical analysis). The two were declared consistent if their 95% intervals overlapped.

The plan also specified that GB-Full precision be compared in the same way with its development estimate, 0.9619 (95% CI 0.9495-0.9725), obtained at its balanced threshold from pooled out-of-fold development scores with a cluster bootstrap.

The plan stated that its outcome descriptions apply to each development model. For GB-Contig, GB-Panel and GB-Calls, precision was compared with the development estimate listed for each model at its own balanced threshold, using the corresponding cluster bootstrap interval: GB-Contig, 0.925 (95% CI, 0.909-0.940); GB-Panel, 0.951 (0.936-0.965); and GB-Calls, 0.957 (0.944-0.969). This model-specific application was post hoc; the development estimates themselves were fixed in the plan before the cohort results were available.

Secondary endpoints were:

- recall and F1 of v1.2-General;
- precision, recall and F1 of v1.1 and the router;
- coverage and failures of each classifier.

All results were stratified by database exposure. The classifiers and PlasmidCall v1.2-General, v1.1 and the router were further stratified by genus and by ARG-bearing status. Because precision, unlike recall, depends on the prevalence of plasmid contigs, precision was compared between methods only on the same contigs: all truth-resolved contigs, or the contigs of one stratum.

The methods evaluated alongside v1.2-General were:

- the four development models, at their balanced thresholds;
- the twelve component classifiers;
- the secondary reference methods, v1.1 and the router, including their high-confidence tiers.

PlasmidEC and PlaScope, run in *E. coli* configurations, were descriptive only, as were the high-confidence tiers of v1.1 and the router. The paired comparisons were:

1. v1.2-General comparison: v1.2-General against each of the four development models and each of the ten other classifiers (14 comparisons), in database-absent precision and recall;

2. Classifier comparison: each development model against each of those ten classifiers, in database-absent precision;

3. Ablation ladder: the contrasts GB-Contig→GB-Calls, GB-Panel→GB-Calls and GB-Calls→GB-Full, in all-contig precision and recall and in database-absent precision;

4. Exposure-drop comparison: the fall in precision from database-exposed to database-absent contigs (precision on database-exposed minus precision on database-absent contigs), compared between each development model and each of those ten classifiers.

The exposure-drop comparison alone compares precision between strata, and no fall was adjusted for prevalence. A method's fall therefore reflects the difference in plasmid prevalence between the strata as well as any change in its recall or in the fraction of chromosomal contigs it calls plasmid, and it is not a measure of the method alone. The two falls in each exposure-drop difference were computed on the same contigs, so both methods face the same pair of prevalences, but the size of the difference still depends on those prevalences.

A time-stamped analysis plan specified the following before any prediction was compared with the reference standard:

- the primary and secondary endpoints;
- the genus and ARG strata;
- the development-model endpoints (precision and recall of each development model at its balanced threshold, on truth-resolved contigs overall and by database exposure);
- the classifier, ablation-ladder and exposure-drop comparisons.

The following were added afterwards and are reported as post hoc:

- the exposure stratification of PlasmidCall v1.2-General, v1.1 and the router;
- the v1.2-General comparison;
- the label check;
- the analyses in the final two subsections (Sensitivity analysis and Additional post hoc analyses).

Analyses specified in the plan but not completed were not included in the reported results.

Statistical analysis

Uncertainty was estimated by resampling genomic clusters formed by single linkage at an ANI of at least 99.5%. ANI was computed all-against-all between the short-read assemblies with skani v0.2.2[12] (triangle mode, minimum aligned fraction 15%). The plan also required the primary endpoint to be reported with and without cohort isolates near-identical to any of 583 genomes used in earlier work, including the training genomes of v1.1 and v1.2-General; near-identity, which the plan did not define, was taken as an ANI of at least 99.5% (skani, minimum aligned fraction 15%).

In each of 4,000 replicates, clusters were drawn with replacement from the whole cohort, and every quantity was recomputed on the drawn clusters' contigs, within strata where applicable. The 95% intervals are the 2.5th and 97.5th replicate percentiles. Point estimates use the full cohort. The ANI-cluster interval is primary, including for the primary endpoint. Isolate-level intervals, from the same procedure with isolates resampled in place of clusters, are secondary.

Paired differences were computed within each replicate on the same resampled contigs, using unrounded values rather than the rounded values reported. They are signed as follows:

- v1.2-General minus the other method;
- development model minus classifier;
- for each contrast X→Y, Y minus X;
- for the exposure-drop comparison, the classifier's fall minus the development model's.

A difference was described as positive or negative only when its confidence interval excluded zero; a bound exactly equal to zero was counted as including zero. Replicates in which a quantity was undefined

were excluded from its interval; their number is reported for paired differences. No adjustment was made for multiple comparisons.

Analyses used Python 3.14.6, NumPy 2.5.0, pandas 3.0.3 and scikit-learn 1.9.0[13]. Bootstrap resampling used a fixed random seed.

Sensitivity analysis

Chromosome labels on hybrid sequences shorter than 1 Mb are the least supported. They arise from rule 1 in assemblies lacking a sequence of at least 1 Mb, or from rule 4. Database-absent contigs with such labels were excluded, and precision and the v1.2-General comparison were recomputed. A narrower variant excluded only rule-4 labels. Both exclusions were also applied to v1.2-General on database-exposed contigs and on all truth-resolved contigs. Because they remove only chromosome labels, these exclusions can raise precision but cannot lower it, and they leave recall unchanged.

In two bounding scenarios, every unresolved eligible contig was counted as plasmid, and then every one as chromosome, and v1.2-General's precision and recall were recomputed on all 35,805 eligible contigs.

Additional post hoc analyses

The recall and false-positive rate (FP/[FP+TN]) of v1.2-General, which unlike precision do not depend on plasmid prevalence, were compared with those of the earlier evaluation, whose intervals were computed from its deposited data[11] with the same bootstrap.

A majority-vote baseline, the rule used in the earlier evaluation, called a contig plasmid when more of the twelve classifiers called it plasmid than chromosome; abstentions and FAILED and MISSING states were not counted, and ties counted as chromosome. v1.2-General was compared, as paired differences (Statistical analysis):

- with this baseline, in precision, recall and F1, on all truth-resolved contigs and on database-absent contigs;
- with each of the ten classifiers other than PlasmidEC and PlaScope, in the same metrics on all truth-resolved contigs;
- with the router, in precision and recall on ARG-bearing contigs.

Performance on contigs carrying a carbapenemase gene was also summarised.

Supplementary Results

S1. Component classifiers, secondary reference methods and paired comparisons

Component classifiers and secondary reference methods

On all truth-resolved contigs, the ten classifiers other than PlasmidEC and PlaScope had precision of 0.553 to 0.981 and recall of 0.203 to 0.995 (Table S3). On database-absent contigs, they had precision of 0.085 to 0.792 and recall of 0.611 to 0.892, except for PlasmidFinder, which had precision 0.993 (95% CI, 0.968-1.000; ANI-cluster bootstrap) and recall 0.320 (0.233-0.372).

Of the secondary reference methods, PlasmidCall v1.1 had precision 0.944 (95% CI, 0.864-0.959; ANI-cluster bootstrap) and recall 0.587 (0.419-0.622), and the router had precision 0.951 (0.900-0.964) and recall 0.837 (0.766-0.852); both are secondary endpoints.

The four gradient-boosting models, PlasmidCall v1.1 and the router also returned a decision for every truth-resolved contig, as did seven of the twelve classifiers. Coverage was 0.887 for gplas2, 0.991 for MOB-recon, 0.970 for PlaScope, 0.996 for PlasmidFinder and 0.960 for plASgraph2. On the 30,986 truth-resolved contigs, failed runs affected only gplas2 (437 contigs), MOB-recon (281) and PlasmidFinder (130), no expected output was missing, and AMRFinderPlus annotation succeeded on all contigs. Run on each isolate one after another, each in a container limited to 8 CPUs and 24 GB of memory, the twelve classifiers took a median of 27 min of wall-clock time per isolate (interquartile range, 22-31 min).

In post hoc paired comparisons on all truth-resolved contigs, v1.2-General had higher precision and lower recall than the majority vote of the twelve classifiers, by +0.041 (95% CI, 0.030 to 0.056; ANI-cluster bootstrap) and -0.046 (-0.055 to -0.033), and the F1 difference, -0.002 (-0.011 to 0.009), included zero. v1.2-General made 431 false plasmid calls and missed 713 plasmid contigs, against 852 and 307 for the vote. On database-absent contigs, its precision exceeded that of the vote by 0.257 (0.032 to 0.377), with 6 false plasmid calls against 128, and the recall difference, -0.037 (-0.084 to 0.013), included zero. Against the ten classifiers other than PlasmidEC and PlaScope on all truth-resolved contigs, v1.2-General had higher precision than nine and lower precision than PlasmidFinder, higher recall than seven, lower recall than PLASMe and Plasmer and a recall interval including zero against MOB-recon, and higher F1 than nine. It had both higher precision and higher recall than geNomad, gplas2, HyAsP, Platon, RFPlasmid and plASgraph2, and no classifier had both higher precision and higher recall than v1.2-General.

Paired comparisons on database-absent contigs

In the post hoc v1.2-General comparison, PlasmidCall v1.2-General was compared with the four gradient-boosting models and the ten classifiers other than PlasmidEC and PlaScope on the same database-absent contigs (Figure 2A). Differences are v1.2-General minus the other method. PlasmidCall v1.2-General and the development models were compared at thresholds chosen by different rules: v1.2-General at its fixed threshold, set to reach a precision of 0.970 in its own development data, and each development model at its balanced threshold (Supplementary Methods S2). Against each development model, v1.2-General had higher precision and lower recall. The precision differences were +0.408 for GB-Contig (95% CI, 0.179 to 0.518; ANI-cluster bootstrap), +0.170 for GB-Panel (0.024 to 0.257), +0.167 for GB-Calls (0.022 to 0.257) and +0.172 for GB-Full (0.008 to 0.268), and the recall differences were -0.154 (-0.264 to -0.099), -0.106 (-0.158 to -0.073), -0.129 (-0.222 to -0.082) and -0.134 (-0.218 to -0.086), respectively. On these contigs, v1.2-General made 6 false plasmid calls and recovered 321 plasmid contigs, against 85 to 288 false plasmid calls and 367 to 388 recovered plasmid contigs for the development models. Against the ten classifiers, precision was higher for v1.2-General than for nine (differences +0.190 to +0.897); for PlasmidFinder, the difference, -0.011 (-0.033 to 0.014), included zero. Recall was lower for v1.2-General than for PLASMe and Plasmer and higher than for HyAsP, MOB-recon and PlasmidFinder, and the interval included zero for geNomad, gplas2, Platon, RFPlasmid and plASgraph2. Isolate-level resampling changed none of the 28 verdicts, and no bootstrap replicate was undefined.

In the prespecified classifier comparison, the database-absent precision of each development model was compared with that of each of the ten classifiers other than PlasmidEC and PlaScope, giving 40 paired differences (Figure 2B). Differences are development model minus classifier, so positive values indicate

higher precision for the development model. The 95% interval excluded zero for 32 differences: 25 positive and 7 negative. All four gradient-boosting models had higher precision than geNomad, HyAsP, PLASMe, RFPlasmid and plASgraph2 (differences +0.053 to +0.729) and lower precision than PlasmidFinder (differences -0.419 to -0.179). No interval excluded zero for MOB-recon. For gplas2, Plasmer and Platon, the verdict differed between models. GB-Contig had lower precision than all three, and GB-Calls had higher precision than all three. GB-Panel and GB-Full had higher precision than gplas2, with intervals including zero for Plasmer and Platon. Two intervals had a bound within 0.001 of zero: GB-Calls minus Plasmer, +0.023 (0.0003 to 0.080), and GB-Panel minus Platon, +0.043 (-0.0004 to 0.086). Under isolate-level resampling, 6 of the 40 verdicts changed between excluding and including zero, and none changed direction.

Fall in precision from database-exposed to database-absent contigs

In the prespecified exposure-drop comparison, the fall in precision from database-exposed to database-absent contigs was compared between each development model and each of the ten classifiers other than PlasmidEC and PlaScope (40 paired differences, the classifier's fall minus the development model's): the 95% interval (ANI-cluster bootstrap) excluded zero with a larger fall for the classifier in 24 differences (seven classifiers each for GB-Panel and GB-Calls, six for GB-Full and four for GB-Contig), with a smaller fall in 5, all involving GB-Contig, and included zero in 11.

S2. Ablation ladder of the four gradient-boosting models

In the prespecified ablation ladder, the three ladder contrasts compared development models that differ only in their input blocks (Supplementary Methods S2; Figure 3A); each difference is the second model minus the first. On all truth-resolved contigs (Figure 3B), precision was higher for GB-Calls than for GB-Contig, by 0.027 (95% CI, 0.022 to 0.045; ANI-cluster bootstrap), and higher for GB-Full than for GB-Calls, by 0.008 (0.0003 to 0.009). The GB-Panel→GB-Calls difference in precision, -0.005 (-0.009 to 0.016), included zero. No recall difference excluded zero: +0.002 (-0.002 to 0.006) for GB-Contig→GB-Calls, +0.001 (-0.002 to 0.008) for GB-Panel→GB-Calls, and +0.001 (0.0000 to 0.004) for GB-Calls→GB-Full, whose lower bound was exactly zero. On database-absent contigs (Figure 3C), precision was higher for GB-Calls than for GB-Contig, by 0.240 (0.139 to 0.271). The GB-Panel→GB-Calls difference, +0.002 (-0.023 to 0.020), and the GB-Calls→GB-Full difference, -0.004 (-0.018 to 0.035), included zero. Under isolate-level resampling, two of the nine intervals excluded zero where the ANI-cluster interval had not: all-contig precision for GB-Panel→GB-Calls (-0.008 to -0.001) and recall for GB-Calls→GB-Full (0.0005 to 0.003).

S3. Genus strata, ARG-bearing contigs and KPC genes

In the prespecified genus strata (95% CIs, ANI-cluster bootstrap), PlasmidCall v1.2-General had precision and recall of 0.958 (0.881-0.987) and 0.929 (0.864-0.936) for *Klebsiella* (313 isolates), 0.899 (0.829-0.971) and 0.868 (0.834-0.915) for *Enterobacter* (66), 0.978 (0.941-1.000) and 0.898 (0.807-0.964) for *Escherichia* (17), 0.986 (0.947-1.000) and 0.891 (0.837-0.922) for *Citrobacter* (16) and 0.688 (0.628-1.000) and 0.794 (0.500-1.000) for *Serratia* (16). In the other three genera, which together comprised four isolates, v1.2-General recovered 32 of 35 plasmid contigs, with 3 false plasmid calls, in the two *Proteus* isolates; 4 of 6, with 4 false plasmid calls, in the *Morganella* isolate; and 36 of 39, with none, in the *Raoultella* isolate (Table 2).

Of the 6,428 determinants meeting the ARG-bearing definition, 765 (11.9%), none of them a carbapenemase gene, lay on contigs shorter than 1 kb and were therefore not evaluated. In the prespecified stratum of 3,005 ARG-bearing contigs (1,810 plasmid), PlasmidCall v1.2-General had precision 0.935 (95% CI, 0.834-0.960; ANI-cluster bootstrap) and recall 0.969 (0.953-0.979), and PlasmidCall v1.1 had precision 0.940 (0.830-0.960) and recall 0.576 (0.399-0.610). The router sends ARG-bearing contigs, which lie outside v1.2-General's specified domain (Supplementary Methods S2), to v1.1, so its calls on these contigs were those of v1.1 (Table 2). On all truth-resolved contigs, the router recovered 7,319 plasmid contigs and v1.2-General 8,030; the whole difference of 711 plasmid contigs lay on ARG-bearing contigs. Among the ten classifiers other than PlasmidEC and PlaScope, precision on ARG-bearing contigs ranged from 0.820 for geNomad (recall 0.614) to 1.000 for PlasmidFinder (recall 0.081). PlasmidEC and PlaScope, run in *E. coli* configurations (Table S1) and reported descriptively, had precision 0.925 and 0.645 and recall 0.999 and 0.998, respectively.

Routing ARG-bearing contigs to v1.1 lost 730 plasmid contigs that v1.2-General recovered and gained 19, and it avoided 59 of v1.2-General's false plasmid calls while adding 5. In a post hoc paired comparison on these contigs, the recall of v1.2-General exceeded that of the router by 0.393 (95% CI, 0.358 to 0.570; ANI-cluster bootstrap), and the precision difference, -0.004 (-0.042 to 0.053), included zero.

Of the 121 false plasmid calls of v1.2-General on ARG-bearing contigs, 119 fell among the 143 database-exposed chromosomal ARG-bearing contigs, on contigs carrying genes found on both replicon types, such as *sul1*, *aadA2* and *bla*$_{\text{OXA-1}}$. Among the 1,052 database-absent chromosomal ARG-bearing contigs, it made 2 false plasmid calls, none on contigs carrying intrinsic chromosomal genes such as *fosA*, *oqxAB* or *bla*$_{\text{SHV-11}}$. Every isolate carried one KPC carbapenemase gene (*bla*$_{\text{KPC-3}}$ in 250 isolates, *bla*$_{\text{KPC-2}}$ in 171 and other *bla*$_{\text{KPC}}$ alleles in 11), and no other carbapenemase gene was detected. Of the 432 KPC genes, 383 lay on truth-resolved contigs, 369 of them plasmid, and 49 on unresolved contigs. In a post hoc analysis of the 383 KPC-bearing contigs, v1.2-General had precision 0.994 (95% CI, 0.965-1.000) and recall 0.970 (0.947-1.000), and the router had precision 0.993 and recall 0.767.

S4. Consistency with the earlier and development estimates, and reference-standard sensitivity

Consistency with the earlier and development estimates

The ANI-cluster precision interval of PlasmidCall v1.2-General overlapped the isolate-clustered interval of the earlier estimate, 0.9770 (95% CI, 0.9642-0.9872), so precision was consistent with the earlier estimate under the prespecified criterion. The overlap was less than 0.0001, below the four-decimal display precision, and the point estimate was 0.0279 lower. With isolate-level resampling, the interval type used for the earlier estimate, the precision interval was 0.9346-0.9620 and lay wholly below the earlier interval, by about 0.002.

Plasmid contigs made up 28.2% of the truth-resolved contigs here and 23.9% (2,241 of 9,371) in the earlier evaluation. In a post hoc comparison of quantities that do not depend on that prevalence, the recall of v1.2-General was higher than in the earlier evaluation, 0.9184 against 0.6631 (95% CI, 0.6205-0.7052), and so was its false-positive rate, 0.0194 against 0.0049; for both quantities, the two intervals did not overlap under either resampling unit.

At an ANI of at least 99.5%, 333 of the 432 isolates, contributing 24,143 of the 30,986 truth-resolved contigs, were near-identical to at least one genome used in earlier work, none of them from the same BioSample. Without these isolates (99 isolates), precision was 0.9402 (95% CI, 0.8904-0.9782; ANI-

cluster bootstrap), against 0.9491 with them, coverage remained 1.000 and the consistency verdict was unchanged.

GB-Full precision was 0.9221 (95% CI, 0.8745-0.9495; ANI-cluster bootstrap). The interval overlapped that of its development estimate, 0.9619 (95% CI, 0.9495-0.9725; cluster bootstrap on development data), so precision was consistent with development under the prespecified criterion. The overlap was approximately 0.000064. The overlap was calculated from the unrounded interval bounds. With isolate-level resampling, the precision interval was 0.9062-0.9369 and lay wholly below the development interval.

For GB-Contig, GB-Panel and GB-Calls, precision was 0.887 (95% CI, 0.836-0.915; ANI-cluster bootstrap), 0.919 (0.863-0.941) and 0.915 (0.872-0.945), respectively, against development estimates of 0.925 (95% CI, 0.909-0.940; cluster bootstrap on development data), 0.951 (0.936-0.965) and 0.957 (0.944-0.969). Each interval overlapped its corresponding development interval, so precision was consistent with development for all three models under the same criterion. With isolate-level resampling, a secondary analysis, each of the three precision intervals lay wholly below its corresponding development interval; this did not change the primary ANI-cluster verdicts.

Unresolved contigs and truth-label sensitivity

The 4,819 unresolved contigs made up 13.5% of eligible contigs but 2.5% of eligible sequence; their median length was 1,979 bp, against 21,424 bp for resolved contigs, and half were shorter than 2 kb. v1.2-General called 1,574 of them (32.7%) plasmid. In post hoc bounding scenarios on all 35,805 eligible contigs, counting every unresolved contig as plasmid gave precision 0.957 (95% CI, 0.910-0.970) and recall 0.708 (0.424-0.809), and counting every one as chromosome gave precision 0.800 (0.727-0.829) and recall 0.918. The consistency criterion was met in the first scenario and not in the second; the primary estimate and its verdict describe truth-resolved contigs.

In the post hoc truth-label sensitivity analysis, excluding the 602 database-absent contigs whose chromosome label rested on a hybrid sequence shorter than 1 Mb reduced the false plasmid calls of PlasmidCall v1.2-General on database-absent contigs from 6 to 4; its precision changed from 0.982 to 0.988 (95% CI, 0.967-0.998; ANI-cluster bootstrap), and its recall was unchanged at 0.740. Excluding only the 346 contigs whose chromosome label came from rule 4 gave the same values. Under each exclusion, the 14 paired precision differences of the v1.2-General comparison were recomputed, and none of the 28 recomputed differences changed its verdict under either resampling scheme. No bootstrap replicate was undefined for any recomputed difference.

Across all truth-resolved contigs, 115 of the 431 false plasmid calls of v1.2-General (26.7%) rested on rule-4 chromosome labels, 113 of them among the 425 on database-exposed contigs. Excluding the 540 contigs with rule-4 labels left 316 false plasmid calls and gave a precision of 0.962 (95% CI, 0.906-0.974; ANI-cluster bootstrap), with recall unchanged; excluding the 808 contigs whose chromosome label rested on a hybrid sequence shorter than 1 Mb gave the same values. On database-exposed contigs, both exclusions removed the same 113 false plasmid calls, and precision changed from 0.948 to 0.961 (0.899-0.973). Even if every rule-4 label were wrong, at least 316 of the 431 false plasmid calls would remain; the data do not establish that these labels are incorrect.

Table S1. The twelve component plasmid classifiers. Every isolate was run with the same version, input, settings and database.

Classifier	Version	Input	Settings	Database or model
geNomad[14]	1.9.8	Contigs	End-to-end workflow; default settings	geNomad database v1.5 (doi:10.5281/zenodo.8339387)
gplas2[15]	1.1.0	Assembly graph	PlasmidEC predictions (*E. coli* preset) as the initial classification; default settings	None
HyAsP[16]	1.0.0	Assembly graph	Maximum plasmid length 750 kb; otherwise default settings	HyAsP NCBI RefSeq database (doi:10.6084/m9.figshare.8001827.v1)
MOB-recon[17]	MOB-suite 3.1.7	Contigs	Default settings	Database bundled in the MOB-suite 3.1.7 container
PLASMe[7]	1.1 (commit ef0409b)	Contigs	Coverage 0.9, identity 0.9, probability 0.5	PLASMe database (doi:10.5281/zenodo.8046934)
PlaScope[18,19]	Centrifuge 1.0.4	Contigs	Up to 1,000 assignments per contig	PlaScope *E. coli* database v1.1 (doi:10.5281/zenodo.1311641)
Plasmer[20]	0.1	Contigs	Minimum length 0 bp; chromosome length threshold 100 Mb	Plasmer database v1.0 (doi:10.5281/zenodo.7030675)
PlasmidEC[15]	1.3	Assembly graph	*E. coli* preset: Centrifuge, Platon and RFPlasmid (Enterobacteriaceae model), majority vote	PlaScope *E. coli* database v1.1; Platon database v1.5.0
PlasmidFinder[21]	2.0.1	Contigs	BLAST method; default thresholds	PlasmidFinder database 2.2.0 (commit 49c24c1)
Platon[22]	1.6	Contigs	Default settings	Platon database v1.5.0 (doi:10.5281/zenodo.4066768)
RFPlasmid[23]	0.0.18	Contigs	Enterobacteriaceae model; Jellyfish k-mer counting	Models distributed with RFPlasmid 0.0.18
plASgraph2[24]	1.0.0 (commit 7bc4e92)	Assembly graph	Default settings	ESKAPEE model distributed with plASgraph2 1.0.0

Table S2. Estimators, training data, input features and thresholds of PlasmidCall v1.2-General and the four gradient-boosting models

Model	Estimator	Training data	Features	Inputs	Threshold
PlasmidCall v1.2-General	Logistic-regression meta-classifier; L2-regularised fitting (C = 1)	3,811 truth-labelled contigs (626 plasmid, 130 ARG-bearing) of 39 *E. coli* isolates	102	One-hot indicators of each classifier's seven states (84); one abstention indicator per classifier (12); six panel summaries: the numbers of plasmid, chromosome and valid calls; the plasmid and chromosome fractions of valid calls; and the larger of those fractions (agreement)	0.9285
GB-Contig	Histogram gradient boosting with 1,000 iterations, no early stopping, otherwise default settings and a fixed seed	13,195 truth-labelled contigs (2,889 plasmid) from 189 isolates	159	A contig block (composition, read depth, single-contig assembly-graph features, taxon, ARG-bearing status and number of qualifying determinants), plus MOB-recon and PlasmidFinder calls	0.0951
GB-Panel			43	All twelve classifiers' calls, abstentions, scores and neighbour votes, without the contig block	0.4004
GB-Calls			173	GB-Contig plus the other ten classifiers' calls and abstentions	0.3116
GB-Full			208	GB-Calls plus continuous classifier scores, assembly-graph neighbour votes and isolate-level context	0.3021

Valid calls. Valid calls exclude abstentions and FAILED and MISSING states.

Thresholds. Contigs scoring at least 0.9285 are called plasmid by v1.2-General. Each balanced threshold is the score that maximises the smaller of recall and precision over pooled out-of-fold development scores, with ties resolved to the lower score.

Table S3. Performance of PlasmidCall v1.2-General, the four gradient-boosting models, the secondary reference methods and the component classifiers on truth-resolved contigs, overall and by database exposure

Method	Status	All					Exposed		Absent		
		Precision (95% CI)	Recall (95% CI)	F1	Coverage	False plasmid calls	Precision (95% CI)	Recall (95% CI)	Precision (95% CI)	Recall (95% CI)	False plasmid calls
Principal model											
PlasmidCall v1.2-General	Primary; strata post hoc	0.9491 (0.8915-0.9642)	0.918 (0.868-0.928)	0.934	1.000	431	0.948 (0.884-0.963)	0.928 (0.887-0.935)	0.982 (0.961-0.993)	0.740 (0.667-0.834)	6
Four gradient-boosting models											
GB-Contig	Prespecified endpoints	0.887 (0.836-0.915)	0.989 (0.978-0.991)	0.935	1.000	1,100	0.910 (0.850-0.937)	0.994 (0.982-0.996)	0.574 (0.467-0.799)	0.894 (0.842-0.973)	288
GB-Panel	Prespecified endpoints	0.919 (0.863-0.941)	0.990 (0.974-0.993)	0.953	1.000	760	0.925 (0.859-0.943)	0.998 (0.992-0.999)	0.812 (0.727-0.957)	0.846 (0.803-0.923)	85
GB-Calls	Prespecified endpoints	0.915 (0.872-0.945)	0.991 (0.980-0.993)	0.951	1.000	809	0.920 (0.867-0.945)	0.997 (0.991-0.998)	0.814 (0.728-0.956)	0.869 (0.828-0.940)	86
GB-Full	Prespecified endpoints	0.922 (0.875-0.950)	0.992 (0.982-0.994)	0.956	1.000	733	0.928 (0.870-0.948)	0.999 (0.993-1.000)	0.810 (0.718-0.968)	0.873 (0.837-0.940)	89
Secondary reference methods											
PlasmidCall v1.1	Secondary; strata post hoc	0.944 (0.864-0.959)	0.587 (0.419-0.622)	0.724	1.000	304	0.943 (0.851-0.958)	0.591 (0.416-0.625)	0.969 (0.953-1.000)	0.512 (0.423-0.564)	7
PlasmidCall v1.1 high-confidence tier	Descriptive	0.949 (0.873-0.963)	0.561 (0.401-0.594)	0.705	1.000	264	0.947 (0.861-0.962)	0.565 (0.398-0.597)	0.982 (0.967-1.000)	0.495 (0.406-0.553)	4
Router	Secondary; strata post hoc	0.951 (0.900-0.964)	0.837 (0.766-0.852)	0.890	1.000	377	0.950 (0.892-0.963)	0.844 (0.774-0.857)	0.981 (0.959-0.993)	0.710 (0.636-0.798)	6
Router high-confidence tier	Descriptive	0.944 (0.843-0.964)	0.116 (0.076-0.125)	0.207	1.000	60	0.944 (0.847-0.963)	0.118 (0.079-0.126)	0.946 (0.760-1.000)	0.081 (0.035-0.110)	2
Component classifiers											
geNomad	Comparator	0.586 (0.564-0.669)	0.657 (0.647-0.702)	0.619	1.000	4,061	0.901 (0.850-0.928)	0.652 (0.644-0.697)	0.085 (0.058-0.216)	0.742 (0.690-0.777)	3,466
gplas2	Comparator	0.919 (0.851-0.937)	0.859 (0.790-0.873)	0.888	0.887	659	0.938 (0.858-0.953)	0.865 (0.798-0.878)	0.642 (0.539-0.876)	0.749 (0.700-0.797)	181
HyAsP	Comparator	0.872 (0.793-0.891)	0.893 (0.841-0.902)	0.882	1.000	1,142	0.910 (0.826-0.929)	0.908 (0.879-0.914)	0.401 (0.326-0.618)	0.611 (0.508-0.676)	396
MOB-recon	Comparator	0.871 (0.816-0.903)	0.937 (0.865-0.951)	0.903	0.991	1,217	0.875 (0.819-0.915)	0.952 (0.903-0.962)	0.774 (0.705-0.851)	0.647 (0.549-0.712)	82
PLASMe	Comparator	0.553 (0.530-0.690)	0.995 (0.986-0.996)	0.711	1.000	7,025	0.668 (0.642-0.831)	1.000 (0.999-1.000)	0.118 (0.080-0.293)	0.892 (0.867-0.940)	2,894
Plasmer	Comparator	0.824 (0.806-0.899)	0.987 (0.966-0.991)	0.898	1.000	1,845	0.825 (0.805-0.901)	0.996 (0.984-0.998)	0.792 (0.721-0.916)	0.823 (0.787-0.892)	94

Method	Status	All					Exposed		Absent		
		Precision (95% CI)	Recall (95% CI)	F1	Coverage	False plasmid calls	Precision (95% CI)	Recall (95% CI)	Precision (95% CI)	Recall (95% CI)	False plasmid calls
PlasmidFinder	Comparator	0.981 (0.954-0.989)	0.203 (0.166-0.211)	0.337	0.996	35	0.980 (0.947-0.988)	0.197 (0.155-0.205)	0.993 (0.968-1.000)	0.320 (0.233-0.372)	1
Platon	Comparator	0.932 (0.869-0.946)	0.870 (0.788-0.887)	0.900	1.000	554	0.941 (0.868-0.955)	0.876 (0.799-0.891)	0.768 (0.690-0.920)	0.742 (0.670-0.806)	97
RFPlasmid	Comparator	0.874 (0.797-0.892)	0.880 (0.845-0.887)	0.877	1.000	1,111	0.900 (0.819-0.918)	0.888 (0.866-0.894)	0.521 (0.441-0.721)	0.724 (0.642-0.770)	289
plASgraph2	Comparator	0.919 (0.862-0.939)	0.784 (0.772-0.839)	0.846	0.960	607	0.957 (0.896-0.979)	0.784 (0.774-0.840)	0.515 (0.420-0.734)	0.776 (0.689-0.882)	317
Component classifiers, descriptive only (*E. coli *configurations)											
PlaScope	Descriptive	0.595 (0.559-0.666)	0.985 (0.956-0.990)	0.742	0.970	5,864	0.812 (0.791-0.898)	0.990 (0.964-0.995)	0.088 (0.061-0.203)	0.882 (0.854-0.933)	3,963
PlasmidEC	Descriptive	0.897 (0.828-0.918)	0.958 (0.908-0.968)	0.926	1.000	962	0.914 (0.834-0.931)	0.967 (0.925-0.975)	0.625 (0.521-0.860)	0.793 (0.753-0.851)	206

Denominators and reading.

"All" means all truth-resolved contigs: 30,986, of which 8,743 are plasmid. These are eligible contigs (at least 1 kb) that the reference standard labelled plasmid or chromosome. Unresolved contigs are excluded from every denominator.

"Exposed" means database-exposed: 12,470 contigs (8,309 plasmid). "Absent" means database-absent: 18,516 contigs (434 plasmid).

Precision, recall, F1 and false plasmid calls use the reading in which every non-plasmid state counts as not plasmid.

Database exposure. A contig is database-exposed if at least 95% of its length is covered by alignments of at least 99% identity to the PLSDB-derived database distributed with PLASMe. All other contigs are database-absent: not represented in that database, which does not mean novel. Plasmid prevalence differs between strata and precision depends on prevalence, so precision is compared between methods within a column. A method's difference between database-exposed and database-absent precision therefore partly reflects prevalence and is not a measure of the method alone (Supplementary Methods S3, Endpoints and comparisons).

Coverage. The number of truth-resolved contigs receiving a plasmid or chromosome decision, divided by 30,986. Abstentions and failed runs count as undecided.

95% CI. Percentile interval from 4,000 bootstrap replicates resampling average nucleotide identity clusters (single linkage at ≥99.5%; 91 clusters).

Thresholds. v1.2-General is shown at its fixed threshold of 0.9285, and v1.1 and the router at their standard thresholds. The four gradient-boosting models are shown at their balanced thresholds, chosen on development data.

Descriptive rows. Four rows are reported descriptively and were not included in the prespecified paired comparisons:

PlasmidEC and PlaScope, run in *E. coli* configurations on all isolates.

The high-confidence tiers of v1.1 and the router, which were not prespecified. These tiers count as plasmid only v1.1's high-confidence calls (threshold 0.9605). The router's tier therefore makes plasmid calls only on ARG-bearing contigs, which the router sends to v1.1.

gplas2 uses PlasmidEC's predictions as its initial classification. In a benchmark of 12 detection tools across Enterobacterales and enterococci, PlaScope, PlasmidEC, Plasmer and gplas2 had the highest F1 scores; that benchmark ran PlasmidEC, whose predictions gplas2 used, with species-specific settings where available[10].

References

1. Gomez-Simmonds A, Annavajhala MK, Seeram D et al. Genomic epidemiology of carbapenem-resistant Enterobacterales at a New York City hospital over a 10-year period reveals complex plasmid-clone dynamics and evidence for frequent horizontal transfer of *bla*$_{KPC}$. Genome Res 2024; 34: 1895-907. https://doi.org/10.1101/gr.279355.124.

2. Shen W, Le S, Li Y et al. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 2016; 11: e0163962. https://doi.org/10.1371/journal.pone.0163962.

3. Wick RR, Judd LM, Gorrie CL et al. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 2017; 13: e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

4. 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: 455-77. https://doi.org/10.1089/cmb.2012.0021.

5. Feldgarden M, Brover V, Gonzalez-Escalona N et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep 2021; 11: 12728. https://doi.org/10.1038/s41598-021-91456-0.

6. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018; 34: 3094-100. https://doi.org/10.1093/bioinformatics/bty191.

7. Tang X, Shang J, Ji Y et al. PLASMe: a tool to identify PLASMid contigs from short-read assemblies using transformer. Nucleic Acids Res 2023; 51: e83. https://doi.org/10.1093/nar/gkad578.

8. Schmartz GP, Hartung A, Hirsch P et al. PLSDB: advancing a comprehensive database of bacterial plasmids. Nucleic Acids Res 2022; 50: D273-8. https://doi.org/10.1093/nar/gkab1111.

9. Camacho C, Coulouris G, Avagyan V et al. BLAST+: architecture and applications. BMC Bioinformatics 2009; 10: 421. https://doi.org/10.1186/1471-2105-10-421.

10. Teixeira M, Souque C, Worby CJ et al. Circling in on plasmids: benchmarking plasmid detection and reconstruction tools for short-read data from diverse species. Brief Bioinform 2025; 26: bbaf589. https://doi.org/10.1093/bib/bbaf589.

11. [dataset] Piranfar V, 2026, Prospective evaluation of PlasmidCall on 150 bacterial genomes across six taxa: assemblies, reference labels, classifier calls and evaluation evidence, Version 1.0.0, Zenodo, https://doi.org/10.5281/zenodo.22086357.

12. Shaw J, Yu YW. Fast and robust metagenomic sequence comparison through sparse chaining with skani. Nat Methods 2023; 20: 1661-5. https://doi.org/10.1038/s41592-023-02018-3.

13. Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011; 12: 2825-30.

14. Camargo AP, Roux S, Schulz F et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol 2024; 42: 1303-12. https://doi.org/10.1038/s41587-023-01953-y.

15. Paganini JA, Kerkvliet JJ, Vader L et al. PlasmidEC and gplas2: an optimized short-read approach to predict and reconstruct antibiotic resistance plasmids in *Escherichia coli*. Microb Genom 2024; 10: 001193. https://doi.org/10.1099/mgen.0.001193.

16. Müller R, Chauve C. HyAsP, a greedy tool for plasmids identification. Bioinformatics 2019; 35: 4436-9. https://doi.org/10.1093/bioinformatics/btz413.

17. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom 2018; 4: e000206. https://doi.org/10.1099/mgen.0.000206.

18. Royer G, Decousser JW, Branger C et al. PlaScope: a targeted approach to assess the plasmidome from genome assemblies at the species level. Microb Genom 2018; 4: e000211. https://doi.org/10.1099/mgen.0.000211.

19. Kim D, Song L, Breitwieser FP et al. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 2016; 26: 1721-9. https://doi.org/10.1101/gr.210641.116.

20. Zhu Q, Gao S, Xiao B et al. Plasmer: an accurate and sensitive bacterial plasmid prediction tool based on machine learning of shared k-mers and genomic features. Microbiol Spectr 2023; 11: e04645-22. https://doi.org/10.1128/spectrum.04645-22.

21. Carattoli A, Zankari E, García-Fernández A et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 2014; 58: 3895-903. https://doi.org/10.1128/AAC.02412-14.

22. Schwengers O, Barth P, Falgenhauer L et al. Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. Microb Genom 2020; 6: mgen000398. https://doi.org/10.1099/mgen.0.000398.

23. van der Graaf-van Bloois L, Wagenaar JA, Zomer AL. RFPlasmid: predicting plasmid sequences from short-read assembly data using machine learning. Microb Genom 2021; 7: 000683. https://doi.org/10.1099/mgen.0.000683.

24. Sielemann J, Sielemann K, Brejová B et al. plASgraph2: using graph neural networks to detect plasmid contigs from an assembly graph. Front Microbiol 2023; 14: 1267695. https://doi.org/10.3389/fmicb.2023.1267695.