Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

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

Vahhab Piranfar

vahab.p@gmail.com

Independent Researcher https://orcid.org/0000-0003-3653-5739

Research Article

Keywords: antimicrobial resistance, PlasmidCall, plasmid-chromosome classification, carbapenem-resistant Enterobacterales, antimicrobial resistance determinants, plasmid databases, blaKPC, short-read assembly

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11651433/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

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

Vahhab Piranfar

Independent Researcher, Jersey City, NJ, USA

Corresponding author: Vahhab Piranfar. E-mail: vahab.p@gmail.com. ORCID: 0000-0003-3653-5739.

Abstract

Antimicrobial resistance genes (ARGs) can lie on the bacterial chromosome, where they are passed on mainly with the host strain, or on plasmids, which can move them between bacteria. Most bacterial genomes are sequenced with short reads, which assemble into many fragments (contigs), so whether a contig, and any ARG on it, comes from a plasmid or from the chromosome usually has to be predicted. Plasmid classifiers are software tools that make this prediction for each contig from different kinds of evidence, and each tends to fail on different contigs. PlasmidCall v1.2-General is a logistic-regression model that combines the plasmid or chromosome calls of twelve such classifiers into one decision. Four gradient-boosting models, fitted on a larger training set, combine some or all of these calls with features of the contig, its neighbours in the assembly graph and its isolate. The five models were tested without refitting on 432 carbapenem-resistant Enterobacterales isolates from one hospital, none used in model development. For each isolate, a hybrid assembly built from long and short reads provided reference labels (plasmid or chromosome) for 30,986 of its 35,805 short-read contigs of at least 1 kb, whether or not they carried an ARG. PlasmidCall v1.2-General gave a decision on every contig: 94.9% of its plasmid calls were correct (precision 0.9491; 95% CI 0.8915-0.9642, resampling clusters of closely related isolates), and it found 91.8% of plasmid contigs (recall 0.9184). Its precision was 0.028 lower than a previously reported estimate; the two 95% intervals overlapped when clusters of related isolates were resampled, but not when single isolates were. The gradient-boosting models found 98.9-99.2% of plasmid contigs at precision 0.887-0.922. On contigs carrying ARGs, v1.2-General had precision 0.935 and recall 0.969, and on the 383 contigs carrying the carbapenemase gene *bla*KPC (post hoc), 0.994 and 0.970. PlasmidCall can therefore place resistance genes, including *bla*KPC, on a plasmid or the chromosome from short-read data alone, with a choice between higher precision and higher recall.

Keywords: antimicrobial resistance; PlasmidCall; plasmid-chromosome classification; carbapenem-resistant Enterobacterales; antimicrobial resistance determinants; plasmid databases

Introduction

Antimicrobial resistance genes (ARGs) in Enterobacterales can spread with their host clone or move between hosts on plasmids.[1] Both routes operate for carbapenemase genes,[2,3] and plasmid-borne carbapenemase genes can spread between bacterial species within a hospital in ways that

tracking clones alone does not detect,[2] yet genomic surveillance has tracked mainly clonal lineages and has often excluded plasmids.[3] Knowing whether an ARG lies on a plasmid or on the chromosome therefore matters for molecular epidemiology and risk assessment.[4]

Long-read sequencing can resolve complete plasmids,[5,6] but most public bacterial assemblies are built from short reads,[6] in which repetitive elements leave plasmids, particularly large ones, incompletely assembled,[7,8] so the replicon of origin of each contig, and of any ARG it carries, has to be inferred.[4]

Plasmid classifiers, tools that predict for each contig whether it comes from a plasmid or from the chromosome, draw on different evidence, from replicon markers and reference plasmids to sequence composition, protein distribution and the assembly graph (Table S1).[4,7,9-18] Alignment-based tools tend to miss diverged plasmids, whereas learning-based tools often have lower precision,[11] and a benchmark across Enterobacterales and enterococci found that performance depended mainly on plasmid representation in databases and on assembly contiguity,[6] so classifiers relying on different evidence can fail on different contigs.

Combining evidence is already common: PLASMe pairs alignment with a learned model,[11] and PlasmidEC takes a majority vote of three classifiers, its authors suggesting confidence-weighted votes as a refinement.[16] A fitted model can instead weight each classifier's call, abstention or failure, and add continuous scores, assembly-graph neighbours, contig features and isolate context.

This study describes PlasmidCall v1.2-General, the principal model, a logistic-regression meta-classifier whose only inputs are the calls of twelve existing classifiers (the component classifiers), and four gradient-boosting models (GB-Contig, GB-Panel, GB-Calls and GB-Full) that combine some or all of these calls with features of the contig, its neighbours in the assembly graph and its isolate. Each model learns, from genomes whose plasmids and chromosomes are known, how much weight to give each classifier's call, abstention or failure. The four gradient-boosting models share one estimator and one training set and differ only in which inputs they use, forming a prespecified ablation ladder rather than successive versions of one model. PlasmidCall's code and models are publicly available.[19]

The five models were evaluated without refitting in an independent cohort of 432 carbapenem-resistant Enterobacterales isolates of eight genera from one hospital,[2] none of them used in model development, against reference labels from hybrid assemblies of the same isolates. Because precision depends on plasmid prevalence,[20] results were stratified by representation in a PLSDB-derived plasmid database.[11,21]

Materials and methods

Study design and evaluation cohort

PlasmidCall v1.2-General was the principal model, and its precision the primary endpoint. The four gradient-boosting models were evaluated alongside it, with PlasmidCall v1.1 and a router as

secondary reference methods and the twelve component classifiers, whose outputs are the models' inputs, as comparators. Models, thresholds and the routing rule were fixed before the evaluation, which used reference labels from hybrid assemblies of the same isolates.

The evaluation cohort comprised every BioSample in NCBI BioProject PRJNA1133668[22] with both an Illumina and an Oxford Nanopore Technologies (ONT) run: 432 of 434 BioSamples of carbapenem-resistant Enterobacterales collected at a New York City hospital in 2009-2018.[2] By the depositors' species labels, which define the genus strata, the isolates span eight genera, led by *Klebsiella* (313) and *Enterobacter* (66) (Supplementary Methods S1). No BioSample or run accession is shared with any isolate used to train or tune an evaluated model.

Illumina reads were capped at an expected depth of 100× with SeqKit[23] and assembled with Unicycler[5] (SPAdes[24]), and every contig of at least 1 kb was eligible, whether or not it carried an ARG. The twelve component classifiers were run with fixed versions, databases and settings corresponding to the model-development inputs (Table S1), and their outputs were mapped to plasmid, chromosome, abstention, FAILED and MISSING states (Supplementary Methods S1). A contig was ARG-bearing if it carried at least one AMRFinderPlus[25] determinant of type and subtype AMR, in the core scope, with a drug class other than efflux.

Reference standard and database exposure

Unicycler built one hybrid assembly per isolate from its ONT reads and the same capped Illumina reads, and 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; and (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, before any prediction had been compared with a reference label, and no plasmid database contributed to any label. Each eligible contig was aligned with minimap2[26] to its own isolate's hybrid sequences, merging alignments of at least 95% identity per target. A contig took its best target's label if that target was labelled, covered at least 80% of the contig and every target within 0.10 of that coverage carried the same label; otherwise it was unresolved, excluded from every denominator and never counted as negative. All predictions were fixed before any long read was retrieved, and the reference standard was built without access to them.

Database exposure was computed without reference labels or predictions, by aligning each eligible contig with BLAST+ megablast[27] to the PLSDB-derived database distributed with PLASMe.[11,21,28] A contig was database-exposed if alignments of at least 99% identity covered at least 95% of its positions, and database-absent otherwise (Supplementary Methods S1). Exposure measures representation in one database, not novelty, and because that database holds no plasmid longer than 350 kb, it partly reflects plasmid size.

PlasmidCall v1.2-General and secondary reference methods

PlasmidCall v1.2-General is an L2-regularised logistic-regression meta-classifier whose only inputs are the component classifiers' calls, encoded as 102 features describing each classifier's state and abstention and the agreement of the panel (Table S2). It was fitted on 3,811 truth-labelled contigs (626 plasmid, 130 ARG-bearing) of 39 *E. coli* isolates, labelled from public complete genome assemblies of the same BioSamples, and calls plasmid at scores of at least 0.9285, the smallest score at which leave-one-isolate-out development precision reached 0.970 with at least ten plasmid calls (Supplementary Methods S2).

PlasmidCall v1.1, a histogram gradient-boosting classifier fitted on 1,460 resistance-gene-bearing contigs of a published benchmark,[6] has thresholds of 0.9524 (standard) and 0.9605 (high-confidence). The router, a predefined rule rather than a fitted model, sends ARG-bearing contigs to v1.1 and all others to v1.2-General, whose specified domain excludes ARG-bearing contigs. Both are distributed with v1.2-General in the PlasmidCall release,[19] so they were evaluated as secondary reference methods to show which of the released options suits ARG-bearing contigs.

Gradient-boosting models

The four gradient-boosting models share one estimator, histogram gradient boosting, and one training set of 13,195 truth-labelled contigs (2,889 plasmid) from 189 isolates, labelled as for v1.2-General. They differ only in their inputs, which their names describe (Table S2). GB-Contig combines features of the contig itself (the contig block: composition, read depth, assembly-graph features of the contig, taxon, ARG-bearing status and number of qualifying determinants) with the calls of two classifiers, MOB-recon and PlasmidFinder. GB-Calls adds the calls and abstentions of the other ten classifiers. GB-Full adds continuous classifier scores, the votes of neighbouring contigs in the assembly graph and isolate-level context. GB-Panel uses all twelve classifiers' calls, abstentions, scores and neighbour votes, without the contig block. Three prespecified contrasts form an ablation ladder: GB-Contig→GB-Calls and GB-Calls→GB-Full each add a block of inputs, and GB-Panel→GB-Calls exchanges classifier scores and neighbour votes for the contig block. In the deposited files, GB-Contig, GB-Panel, GB-Calls and GB-Full are named A, B_nb, C and C_sc_nb_iso.[29] Each model was applied at its balanced threshold, the score maximising the smaller of recall and precision over pooled out-of-fold development scores: 0.0951 (GB-Contig), 0.4004 (GB-Panel), 0.3116 (GB-Calls) and 0.3021 (GB-Full) (Supplementary Methods S2).

Performance endpoints and statistical analysis

Precision (the fraction of plasmid calls that were correct), recall (the fraction of plasmid contigs that were called plasmid), F1 and coverage (the fraction of contigs receiving a plasmid or chromosome decision) were computed per contig on truth-resolved eligible contigs, with plasmid as the positive class and every non-plasmid state counted as not plasmid. The primary endpoint was v1.2-General precision, with coverage co-primary, judged consistent with the earlier estimate, a previously reported precision of 0.9770 (isolate-clustered 95% CI 0.9642-0.9872)

from 150 genomes of six taxa,[30] if their 95% intervals overlapped.[31] The same criterion was prespecified for GB-Full against its development estimate and applied post hoc to GB-Contig, GB-Panel and GB-Calls, each against a development estimate fixed in the plan.

Secondary endpoints were v1.2-General recall and F1, v1.1 and router performance, and classifier coverage and failures; the gradient-boosting models had prespecified endpoints of precision and recall. Results were stratified by database exposure and, for v1.2-General, v1.1, the router and the classifiers, by genus and ARG-bearing status; because precision depends on plasmid prevalence, it was compared between methods only on the same contigs. PlasmidEC and PlaScope, run in *E. coli* configurations, and the high-confidence tiers of v1.1 and the router were descriptive only. Four sets of paired comparisons were made. The v1.2-General comparison set v1.2-General against each gradient-boosting model and each of the ten other classifiers, in database-absent precision and recall. The classifier comparison set each gradient-boosting model against those ten classifiers, in database-absent precision. The ablation ladder compared its three contrasts in all-contig precision and recall and in database-absent precision. The exposure-drop comparison set the same pairs as the classifier comparison against each other in the fall in precision from database-exposed to database-absent contigs, which also reflects the difference in plasmid prevalence. A time-stamped analysis plan specified the endpoints, strata and the classifier, ablation-ladder and exposure-drop comparisons before any prediction was compared with the reference standard; the v1.2-General comparison, the exposure stratification of v1.2-General, v1.1 and the router, exposure by reference label, and the sensitivity and additional analyses were post hoc (Supplementary Methods S3).

Uncertainty was estimated by resampling clusters of isolates formed by single linkage at an average nucleotide identity (ANI) of at least 99.5%, computed with skani.[32] In each of 4,000 bootstrap replicates, clusters were drawn with replacement and every quantity was recomputed, within strata where applicable, giving percentile 95% CIs.[33] ANI-cluster intervals are primary, including for the primary endpoint, and isolate-level intervals, resampling isolates instead, secondary. Paired differences were computed within each replicate on the same resampled contigs, and a difference was described as positive or negative only when its CI excluded zero; no adjustment was made for multiple comparisons. As prespecified, the primary endpoint was also estimated without isolates near-identical (ANI ≥99.5%, a threshold set post hoc) to any of 583 genomes used in earlier work, including the training genomes of v1.1 and v1.2-General. Analyses used Python, NumPy, pandas and scikit-learn[34] (Supplementary Methods S3).

Results

Hybrid assemblies labelled 30,986 of 35,805 contigs, and most plasmid contigs were database-exposed

Of 35,805 eligible contigs, 30,986 were resolved by the reference standard (8,743 plasmid, 22,243 chromosome) and 4,819 were unresolved (Figure 1A). The 18,516 database-absent contigs (59.8%) comprised 434 plasmid and 18,082 chromosomal contigs, so plasmid contigs

made up 2.3% of the stratum; the other 12,470 (40.2%) were database-exposed, and 8,309 of them (66.6%) were plasmid. Exposure, computed independently of the labels, was concentrated among plasmid-labelled contigs: 8,309 of 8,743 (95.0%) were database-exposed, against 4,161 of 22,243 chromosome-labelled contigs (18.7%). The 432 isolates formed 91 ANI clusters, the largest containing 285 isolates (66.0%; Figure 1B).

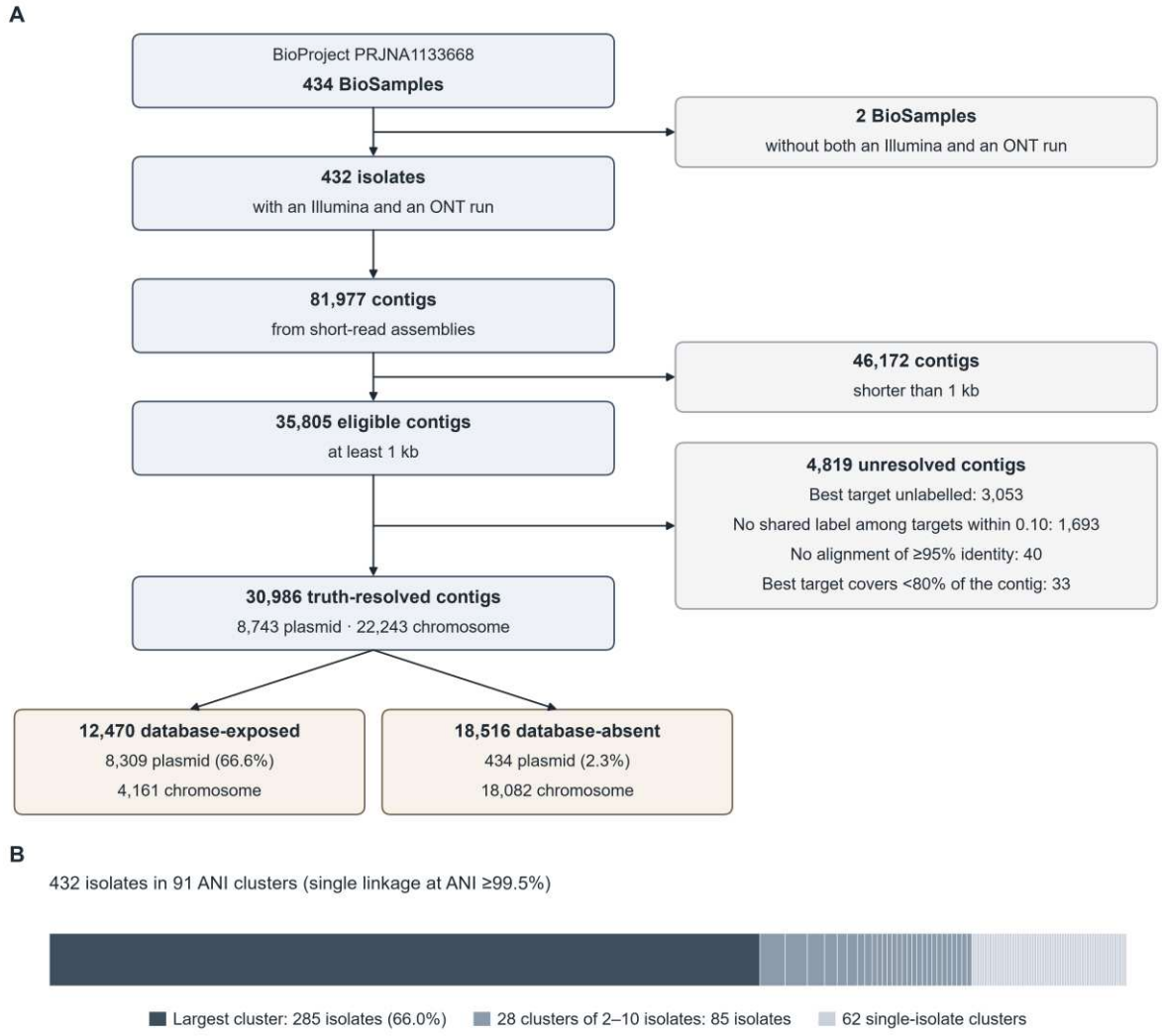


Figure 1. Cohort and contig disposition. (**A**) BioSamples, isolates and contigs from BioProject PRJNA1133668, from all contigs of the short-read assemblies to the truth-resolved contigs analysed and their database-exposure strata. Contigs shorter than 1 kb were not eligible. Unresolved contigs were excluded from every denominator and never counted as negatives. A contig was unresolved when its best target was an unlabelled hybrid sequence, when the targets within 0.10 of the best coverage did not all carry the same plasmid or chromosome label, when no alignment of at least 95% identity was found, or when the best target covered less than 80% of the contig (Methods, Reference standard and 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. Percentages give the plasmid share of each stratum. (**B**) The 432 isolates grouped into average nucleotide identity (ANI) clusters by single linkage at ANI of at least 99.5%. Each segment is one cluster, with width proportional to its number of isolates. These clusters are the resampling units of the ANI-cluster bootstrap (Methods, Performance endpoints and statistical analysis).

PlasmidCall v1.2-General favoured precision and decided every contig

PlasmidCall v1.2-General had a precision of 0.9491 (95% CI 0.8915-0.9642; ANI-cluster bootstrap) and coverage of 1.000, returning a decision for every truth-resolved contig (Table 1). Recall was 0.9184 (0.8681-0.9281), with 8,030 of 8,743 plasmid contigs recovered, and F1 was 0.9335. Of its 431 false plasmid calls, 425 were on database-exposed and 6 on database-absent contigs. On database-absent contigs (post hoc), its precision was 0.982 (0.961-0.993) and recall 0.740 (0.667-0.834), with 321 of 434 plasmid contigs recovered. In genera with at least 16 isolates, precision was at least 0.958 in *Klebsiella*, *Escherichia* and *Citrobacter*, 0.899 in *Enterobacter* and 0.688 in *Serratia*, whose interval was wide (0.628-1.000; Table 2).

Table 1. Performance of PlasmidCall v1.2-General, the four gradient-boosting models and the secondary reference methods 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

Denominators.

"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 count every non-plasmid state 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 (Methods, Performance endpoints and statistical analysis).

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. The high-confidence tiers of v1.1 and the router were not prespecified and are descriptive. 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.

Component classifiers. Results for the twelve component classifiers, on the same contigs and with the same columns, are in Table S3.

Table 2. Performance of PlasmidCall v1.2-General and of the secondary reference methods, PlasmidCall v1.1 and the router, by genus and ARG-bearing status

Stratum	Isolates	Truth-resolved contigs	Plasmid contigs	v1.2-General		v1.1		Router	
				Precision (95% CI)	Recall (95% CI)	Precision (95% CI)	Recall (95% CI)	Precision (95% CI)	Recall (95% CI)
Genus									
Klebsiella	313	23,287	7,022	0.958 (0.881-0.987)	0.929 (0.864-0.936)	0.953 (0.889-0.980)	0.637 (0.423-0.664)	0.960 (0.902-0.986)	0.862 (0.786-0.870)
Enterobacter	66	3,428	980	0.899 (0.829-0.971)	0.868 (0.834-0.915)	0.890 (0.722-0.970)	0.288 (0.181-0.501)	0.899 (0.820-0.966)	0.710 (0.678-0.772)
Escherichia	17	1,626	244	0.978 (0.941-1.000)	0.898 (0.807-0.964)	0.986 (0.966-1.000)	0.590 (0.323-0.809)	0.986 (0.972-1.000)	0.844 (0.754-0.913)
Citrobacter	16	991	320	0.986 (0.947-1.000)	0.891 (0.837-0.922)	0.978 (0.935-1.000)	0.409 (0.188-0.747)	0.984 (0.938-1.000)	0.747 (0.662-0.826)
Serratia	16	1,392	97	0.688 (0.628-1.000)	0.794 (0.500-1.000)	0.590 (0.533-1.000)	0.639 (0.000-0.867)	0.673 (0.626-1.000)	0.742 (0.286-0.800)
Proteus	2	109	35	0.914	0.914	1.000	0.057	0.944	0.486
Morganella	1	41	6	0.500	0.667	-	0.000	0.500	0.500
Raoultella	1	112	39	1.000	0.923	1.000	0.949	1.000	0.923
ARG-bearing status									
ARG-bearing	432	3,005	1,810	0.935 (0.834-0.960)	0.969 (0.953-0.979)	0.940 (0.830-0.960)	0.576 (0.399-0.610)	0.940 (0.830-0.960)	0.576 (0.399-0.610)
Not ARG-bearing	432	27,981	6,933	0.953 (0.907-0.967)	0.905 (0.844-0.917)	0.945 (0.868-0.961)	0.590 (0.420-0.625)	0.953 (0.907-0.967)	0.905 (0.844-0.917)

Strata. Genus is taken from the depositors' species labels (Methods, Study design and evaluation cohort), and ARG-bearing status is defined in Methods (Study design and evaluation cohort). Both sets of strata were prespecified. Isolates are those contributing at least one truth-resolved contig to the stratum.

Calls and thresholds. Precision and recall count every non-plasmid state as not plasmid. v1.2-General is shown at its fixed threshold of 0.9285, and v1.1 and the router at their standard thresholds.

95% CI. Percentile interval from 4,000 bootstrap replicates resampling average nucleotide identity clusters (single linkage at ≥99.5%; 91 clusters), with each quantity recomputed on the stratum's contigs.

Small genera. Intervals are not shown for *Proteus* (2 isolates), *Morganella* (1) and *Raoultella* (1). For the two single-isolate genera, every interval equals its point estimate; for *Proteus*, resampling can draw on only two isolates. Neither describes sampling uncertainty usefully.

Undefined value. v1.1 made no plasmid call among the *Morganella* contigs, so its precision there is undefined (-).

Router. The router sends ARG-bearing contigs to v1.1 and all other contigs to v1.2-General, so its ARG-bearing row repeats the values of v1.1 and its not-ARG-bearing row those of v1.2-General.

The gradient-boosting models recovered nearly all plasmid contigs at lower precision

At their balanced thresholds, the four gradient-boosting models had precision of 0.887-0.922 and recall of 0.989-0.992 on all truth-resolved contigs, and precision of 0.574-0.814 and recall of 0.846-0.894 on database-absent contigs (Table 1). Against the ten classifiers on database-absent contigs (classifier comparison; Figure 2B), 32 of the 40 precision differences excluded zero, 25 of them positive: all four models had higher precision than five of the ten classifiers and lower precision than PlasmidFinder, and for the other four, intervals included zero or the outcome depended on the model (Supplementary Results S1).

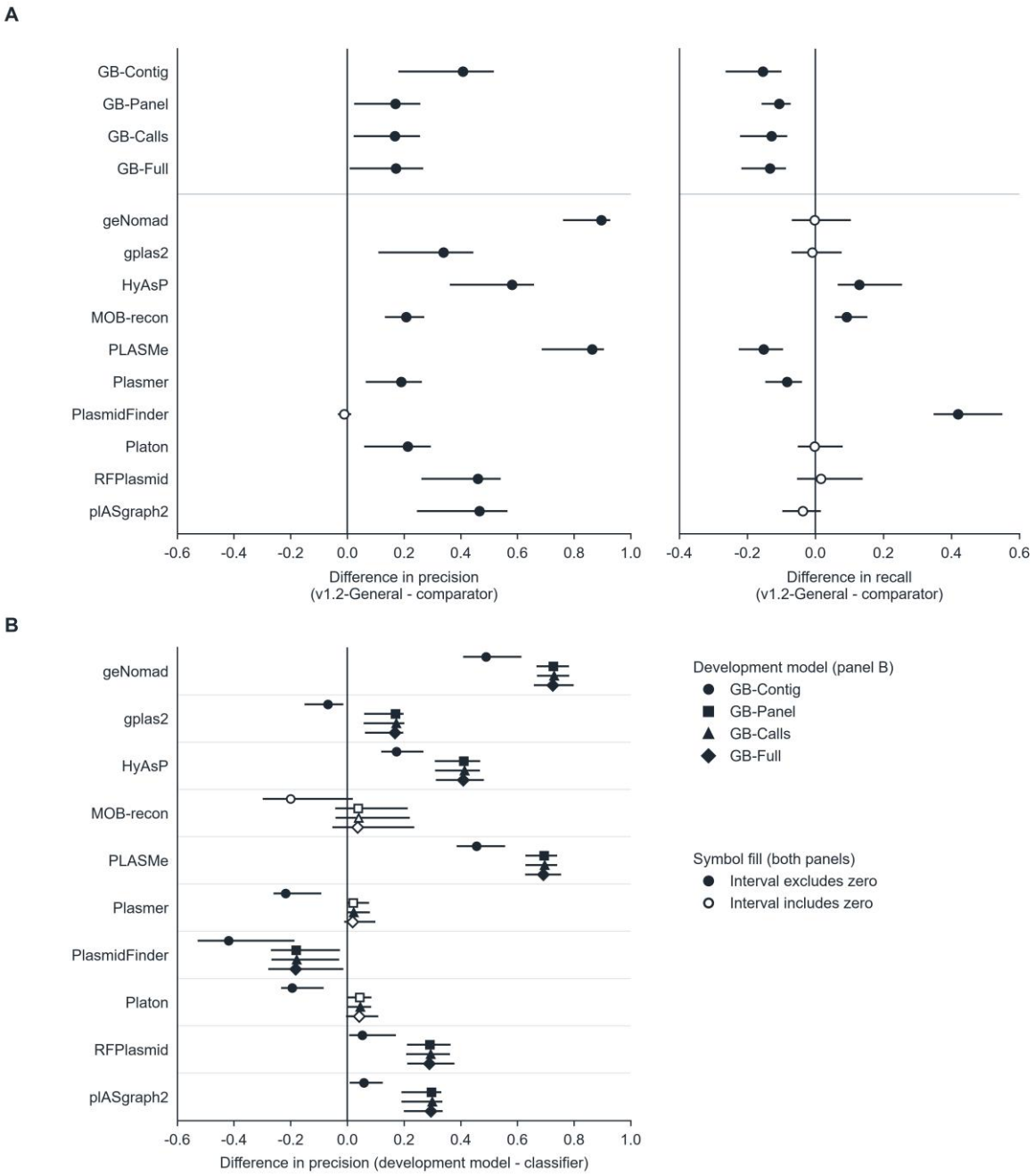


Figure 2. Paired differences on database-absent contigs. Each symbol is the difference between two methods scored on the same 18,516 database-absent truth-resolved contigs (434 plasmid). Each horizontal line is the 95% interval: the 2.5th to 97.5th percentiles of 4,000 bootstrap replicates resampling ANI clusters, with both methods recomputed on the same resampled contigs in each replicate. The vertical line marks zero difference. Filled symbols mark intervals that exclude zero, and open symbols intervals that include zero. Every bootstrap replicate gave a defined difference, and intervals are not adjusted for multiple comparisons. v1.2-General is shown at its fixed threshold (0.9285) and the four gradient-boosting models at their

balanced thresholds. Classifiers are listed in the order of Table S1. (**A**) PlasmidCall v1.2-General minus each of the four gradient-boosting models and each of the ten component classifiers other than PlasmidEC and PlaScope (14 comparisons), in precision (left) and recall (right). Positive values indicate higher values for v1.2-General. The v1.2-General comparison, post hoc. (**B**) Each gradient-boosting model minus each of the same ten classifiers, in precision. Positive values indicate higher precision for the gradient-boosting model, and symbol shape identifies the model (circles, GB-Contig; squares, GB-Panel; triangles, GB-Calls; diamonds, GB-Full). The classifier comparison, prespecified. gplas2 starts from PlasmidEC predictions made in the *E. coli* configuration. Database-absent contigs are not represented in the PLSDB-derived database.

In the ablation ladder (Figure 3), GB-Calls, which adds the ten further classifiers, had higher precision than GB-Contig, by 0.027 (95% CI 0.022 to 0.045) on all contigs and 0.240 (0.139 to 0.271) on database-absent contigs, and GB-Full had slightly higher all-contig precision than GB-Calls, by 0.008 (0.0003 to 0.009). Every other interval, including every recall interval, included zero (Supplementary Results S2).

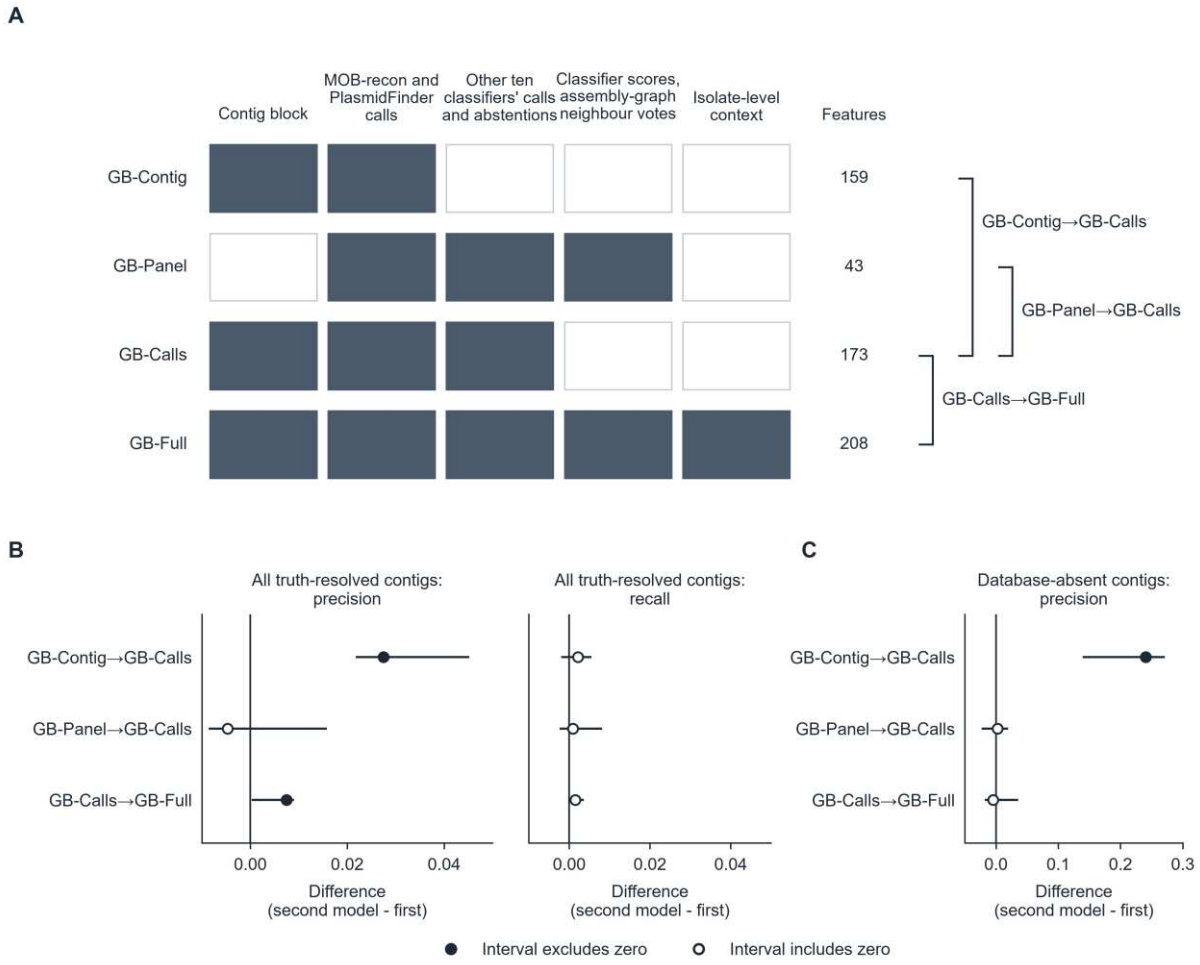


Figure 3. Prespecified ablation ladder of the four gradient-boosting models. (**A**) Input blocks of the four gradient-boosting models. Filled cells mark the blocks each model uses, and numbers

give total features. The models share one estimator and one training set and differ only in these inputs. The contig block comprises composition, read depth, single-contig assembly-graph features, taxon, ARG-bearing status and number of qualifying determinants. Brackets mark the three prespecified contrasts. GB-Contig→GB-Calls adds the other ten classifiers' calls and abstentions. GB-Panel→GB-Calls exchanges classifier scores and assembly-graph neighbour votes for the contig block. GB-Calls→GB-Full adds classifier scores, assembly-graph neighbour votes and isolate context. PlasmidCall v1.2-General, the principal model, is a separately fitted logistic-regression meta-classifier whose only inputs are the twelve component classifiers' calls; it is not part of the ladder. (**B**, **C**) Paired differences between the two models of each contrast in the independent evaluation cohort, computed as the second model minus the first, with every model at its balanced threshold. (**B**) Precision (left) and recall (right) on all 30,986 truth-resolved contigs (8,743 plasmid). (**C**) Precision on the 18,516 database-absent contigs (434 plasmid), on a wider scale than in **B**. Circles are full-cohort point estimates. Horizontal lines are 95% intervals: the 2.5th to 97.5th percentiles of 4,000 bootstrap replicates resampling ANI clusters, with both models recomputed on the same resampled contigs in each replicate. The vertical line marks zero difference. Filled circles mark intervals that exclude zero and open circles intervals that include zero; a bound of exactly zero counts as including zero. Every bootstrap replicate gave a defined difference, and intervals are not adjusted for multiple comparisons. Differences are computed from unrounded values, so they can differ by 0.001 from differences between the rounded values in Table 1. Database-absent contigs are not represented in the PLSDB-derived database.

No classifier had both higher precision and higher recall than v1.2-General

On all truth-resolved contigs, the ten classifiers other than PlasmidEC and PlaScope had precision of 0.553-0.981 and recall of 0.203-0.995; on database-absent contigs, their precision was 0.085-0.792 and recall 0.611-0.892, except PlasmidFinder (precision 0.993, recall 0.320; Table S3). Of the secondary reference methods, v1.1 had precision 0.944 and recall 0.587, and the router 0.951 and 0.837 (Table 1). All models and the router returned a decision for every truth-resolved contig, as did seven of the twelve classifiers.

On database-absent contigs (v1.2-General comparison, post hoc; Figure 2A), v1.2-General had higher precision and lower recall than each gradient-boosting model, at thresholds set by different rules, and higher precision than nine of the ten classifiers, the interval including zero for PlasmidFinder. Against a majority vote of the twelve classifiers (post hoc), v1.2-General had higher precision and lower recall on all truth-resolved contigs, by 0.041 and 0.046, and the F1 difference included zero; on database-absent contigs it made 6 false plasmid calls against 128 for the vote. On all truth-resolved contigs, no classifier had both higher precision and higher recall than v1.2-General. The fall in precision from database-exposed to database-absent contigs (exposure-drop comparison) was larger for the classifier in 24 of 40 differences and smaller in 5, all involving GB-Contig. Under isolate-level resampling, none of the 28 v1.2-General differences and 6 of the 40 gradient-boosting differences moved between excluding and including zero, and none changed direction (Supplementary Results S1).

On ARG-bearing contigs, v1.2-General recovered 711 more plasmid contigs than the router

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 not evaluated. On the 3,005 ARG-bearing contigs (1,810 plasmid), v1.2-General had precision 0.935 (95% CI 0.834-0.960) and recall 0.969 (0.953-0.979), and v1.1, to which the router sends these contigs, 0.940 (0.830-0.960) and 0.576 (0.399-0.610) (Table 2). The router therefore recovered 711 fewer plasmid contigs than v1.2-General, all of them ARG-bearing; post hoc, the recall of v1.2-General on these contigs exceeded that of the router by 0.393 (0.358 to 0.570), 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 on database-exposed chromosomal contigs carrying genes found on both replicon types (Supplementary Results S3).

blaKPC was plasmid-borne in 369 of 383 isolates, and v1.2-General identified 358 of them from short reads

Every isolate carried one KPC carbapenemase gene (*bla*KPC-3 in 250 isolates, *bla*KPC-2 in 171 and other or unassigned *bla*KPC alleles in 11), and no other carbapenemase gene was detected. Of the 432 KPC genes, 383 lay on truth-resolved contigs and 49 on unresolved contigs, and the reference standard labelled 369 of the 383 contigs plasmid and 14 chromosome. In a post hoc analysis, v1.2-General called 358 of the 369 plasmid contigs plasmid, and 2 of the 14 chromosomal contigs, giving precision 0.994 (0.965-1.000) and recall 0.970 (0.947-1.000). The router, which sends these contigs to v1.1, had precision 0.993 and recall 0.767.

Precision was consistent with the earlier estimates only when clusters of related isolates were resampled

The ANI-cluster precision interval of v1.2-General overlapped that of the earlier estimate, 0.9770 (0.9642-0.9872), by less than 0.0001, meeting the prespecified consistency criterion; the point estimate was 0.0279 lower. With isolate-level resampling, the interval type of the earlier estimate, the precision interval was 0.9346-0.9620 and lay wholly below the earlier one, by about 0.002. Plasmid contigs made up 28.2% of truth-resolved contigs here and 23.9% in the earlier evaluation; recall (0.9184 against 0.6631) and the false-positive rate (0.0194 against 0.0049) were both higher here, with non-overlapping intervals under either resampling unit. Consistency with the earlier estimate held without the 333 isolates near-identical (ANI ≥99.5%) to a genome used in earlier work, none of them from the same BioSample: precision was 0.9402 (0.8904-0.9782) and coverage 1.000.

The precision interval of GB-Full, 0.9221 (0.8745-0.9495), overlapped its development interval, 0.9619 (0.9495-0.9725), by approximately 0.000064, and those of GB-Contig, GB-Panel and GB-Calls also overlapped theirs, meeting the same criterion; each isolate-level interval lay wholly below its development interval (Supplementary Results S4).

Treating unresolved contigs as plasmid or as chromosome bounded precision between 0.800 and 0.957

The 4,819 unresolved contigs made up 13.5% of eligible contigs but 2.5% of eligible sequence. In post hoc bounding scenarios, counting every unresolved contig as plasmid gave v1.2-General a precision of 0.957 (0.910-0.970), and counting every one as chromosome 0.800 (0.727-0.829); the consistency criterion was met only in the former. In the truth-label sensitivity analysis, rule-4 chromosome labels underlay 115 of its 431 false plasmid calls; excluding contigs with rule-4 labels, or with chromosome labels on hybrid sequences shorter than 1 Mb, left 316 (precision 0.962, 0.906-0.974), with recall and all v1.2-General comparison results unchanged (Supplementary Results S4).

Discussion

Whether a resistance gene's contig is plasmid-borne or chromosomal informs genomic epidemiology and risk assessment of antimicrobial resistance.[4] PlasmidCall v1.2-General assigns contigs from the calls of existing classifiers, and its four gradient-boosting models add contig, assembly-graph and isolate information. Evaluated without refitting in an independent cohort, the five models provided complementary operating points: v1.2-General favoured precision with complete coverage, whereas the gradient-boosting models recovered nearly all plasmid contigs.

Reference labels came from hybrid assemblies of the same isolates, independently of any classifier or plasmid database. Against these labels, the precision of v1.2-General was consistent with the earlier estimate when clusters of related isolates were resampled, also after removing isolates near-identical to genomes used in earlier work, but not when single isolates were resampled. Plasmid contigs were more frequent here than in the earlier collection, so prevalence would, if anything, have raised precision; the lower precision accompanied a higher false-positive rate, with higher recall. The result also depended on the resampling unit: one ANI cluster held two-thirds of the isolates, consistent with clonal transmission in this collection,[2] and resampling whole clusters accounts for this dependence.[33]

At a fixed recall and false-positive rate, precision falls as plasmid contigs become rarer,[20] as among database-absent contigs, where v1.2-General made few false plasmid calls and recovered about three-quarters of the plasmid contigs, consistent with a threshold chosen for precision. Nearly all its false plasmid calls fell on database-exposed chromosomal contigs, which may consist of mobile genetic elements that move between chromosomes and plasmids[35] or carry labels assigning plasmid-derived sequence to the chromosome. Larger falls in precision for the classifiers in 24 of 40 exposure-drop differences are consistent with the benchmark finding that database representation is a major determinant of plasmid detection,[6] although falls also depend on prevalence and false-positive rates.[20]

Because the v1.2-General comparison set v1.2-General against the gradient-boosting models at thresholds chosen by different rules, it compares operating points rather than ranking the models. The gradient-boosting models met the consistency criterion against their development estimates

and, at their balanced thresholds, had higher database-absent precision than five of the ten classifiers (classifier comparison). In the ablation ladder, GB-Calls had higher precision than GB-Contig, and GB-Full slightly higher all-contig precision than GB-Calls; with model-specific thresholds and one training set, these contrasts describe configurations rather than causal effects of each block. Against a majority vote of the same calls, v1.2-General shifted the operating point towards precision, with no difference in F1.

On ARG-bearing contigs, including those carrying KPC genes, v1.2-General recovered nearly all plasmid contigs with a precision point estimate close to that of v1.1, so routing these contigs to v1.1 cost 711 recovered plasmid contigs in this cohort. v1.2-General was specified for contigs without ARGs, although its training data included 130 ARG-bearing contigs, so this result is post hoc. In this hospital collection, *bla*KPC lay on a plasmid in nearly every isolate whose KPC contig could be labelled, consistent with the frequent horizontal transfer of *bla*KPC reported for these isolates,[2] and v1.2-General recovered nearly all of these plasmid locations from short reads alone. Genus differences partly reflect differences in plasmid prevalence, and pooled estimates largely reflect *Klebsiella*.

This study has limitations. Comparisons with existing classifiers are asymmetric: the models use classifier outputs as inputs and were fitted on labelled genomes, whereas each classifier ran as specified in Table S1; PlasmidEC and PlaScope ran in *E. coli* configurations, and gplas2 starts from PlasmidEC's *E. coli*-preset predictions. v1.2-General was fitted on *E. coli* alone at native depth, whereas cohort reads were capped at 100×, and labels for training and for the earlier and development estimates came from public complete genome assemblies, whereas cohort labels came from the hybrid assemblies built for this study. The cohort comes from one hospital and contains only carbapenem-resistant isolates, several genera have few isolates, exposure was measured against one database and is related to plasmid status, and the estimates describe the contigs that the reference standard could label.

In conclusion, PlasmidCall v1.2-General is a precision-oriented classifier with complete coverage, and its four gradient-boosting models offer higher-recall operating points; together they assign short-read contigs, and the resistance determinants they carry, to plasmid or chromosome. Because it placed nearly all *bla*KPC-bearing contigs correctly from short reads (post hoc), short-read data already in public archives may be used to ask whether carbapenemase genes spread with clones or on plasmids. For genomic surveillance of antimicrobial resistance, precision should be reported with recall, by database exposure and with cluster-level uncertainty, because it depends on plasmid prevalence and its uncertainty, in clonal collections, on the resampling unit.

Acknowledgements

The author acknowledges Google for providing access to computational resources, including a server with 256 CPU cores and 1 TB of RAM, which supported the computational analyses performed in this study.

Funding

No external funding was received for this study. The study was conducted as independent research.

Declaration of competing interest

The author declares no competing interests.

Data availability

Illumina and Oxford Nanopore reads are in NCBI BioProject PRJNA1133668.[22] Short-read and hybrid assemblies, the accession and run manifest, reference labels, classifier calls, predictions, scores, full result tables (including those for the classifier and exposure-drop comparisons and by genus) and the analysis plan are in figshare (CC BY 4.0).[36] Native outputs of the twelve classifiers and of AMRFinderPlus, with execution records, are in Harvard Dataverse (CC BY 4.0).[37] PlasmidCall v1.2-General and v1.1 models and code are in PlasmidCall release v1.1.0 (MIT License),[19] archived from https://github.com/piranfar/PlasmidCall (tag v1.1.0). The four gradient-boosting models and the code that applied them,[29] and the analysis code of this study with the software versions used,[38] are in figshare (MIT License). The source of the earlier estimate is in Zenodo.[30] Classifier versions, settings and databases are given in Table S1.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author used Hermes Agent (Nous Research), running Anthropic's Claude Opus 5.5 large language model, in order to edit the English-language text of the manuscript, including its adaptation to the journal's format, and to organise the Python code. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

1. Rozwandowicz M, Brouwer MSM, Fischer J et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J Antimicrob Chemother 2018; 73: 1121-37. https://doi.org/10.1093/jac/dkx488.
2. 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.
3. David S, Cohen V, Reuter S et al. Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among *Klebsiella pneumoniae*. Proc Natl Acad Sci U S A 2020; 117: 25043-54. https://doi.org/10.1073/pnas.2003407117.

4. 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.

5. 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.

6. 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.

7. 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.

8. Arredondo-Alonso S, Willems RJ, van Schaik W et al. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb Genom 2017; 3: e000128. https://doi.org/10.1099/mgen.0.000128.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

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

18. 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.

19. Piranfar V. PlasmidCall: code and frozen models for plasmid-origin classification of short-read contigs [software]. Version 1.1.0. figshare, 2026. https://doi.org/10.6084/m9.figshare.34018380.v1.

20. Brabec J, Komárek T, Franc V et al. On model evaluation under non-constant class imbalance. In: Computational Science - ICCS 2020. Lecture Notes in Computer Science, vol. 12140. Cham: Springer International Publishing, 2020; 74-87. https://doi.org/10.1007/978-3-030-50423-6_6.

21. 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.

22. [dataset] Columbia University Medical Center, 2024, Carbapenem-resistant Enterobacterales from a New York City Hospital sequencing data, NCBI BioProject, PRJNA1133668.

23. 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.

24. 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.

25. 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.

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

27. 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.

28. [dataset] Tang X, 2023, PLASMe database, Zenodo, https://doi.org/10.5281/zenodo.8046934.

29. Piranfar V. PlasmidCall: code and frozen models for plasmid-origin classification of short-read contigs [software]. Version 2. figshare, 2026. https://doi.org/10.6084/m9.figshare.34018380.v2.

30. [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.

31. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. Am Stat 2001; 55: 182-6. https://doi.org/10.1198/000313001317097960.

32. 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.

33. Field CA, Welsh AH. Bootstrapping clustered data. J R Stat Soc Series B Stat Methodol 2007; 69: 369-90. https://doi.org/10.1111/j.1467-9868.2007.00593.x.

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

35. Partridge SR, Kwong SM, Firth N et al. Mobile genetic elements associated with antimicrobial resistance. Clin Microbiol Rev 2018; 31: e00088-17. https://doi.org/10.1128/CMR.00088-17.

36. [dataset] Piranfar V, 2026, Independent evaluation of PlasmidCall in 432 carbapenem-resistant Enterobacterales isolates: assemblies, reference labels, fixed predictions and evaluation results, figshare, https://doi.org/10.6084/m9.figshare.34021401.

37. [dataset] Piranfar V, 2026, Independent evaluation of PlasmidCall in 432 carbapenem-resistant Enterobacterales isolates: native classifier outputs and execution records, Version 1.0, Harvard Dataverse, https://doi.org/10.7910/DVN/Q5JNX9.

38. Piranfar V. Independent evaluation of PlasmidCall in 432 carbapenem-resistant Enterobacterales isolates: analysis code [software]. figshare, 2026. https://doi.org/10.6084/m9.figshare.34063578.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- P114PreprintSupplementaryMaterial.pdf