A systematic comparison of tools for predicting antimicrobial resistance from nanopore sequence data

1.1 Author names

Natalie Ring[1,2,†], Alison Low[2], Rhodri Evans[2], Marianne Keith[2], Gavin K. Paterson[1,2], David Gally[1], Tim Nuttall[2], Dylan Neil Clements[2], J. Ross Fitzgerald[1]

1.2 Affiliation(s)

[1]The Roslin Institute, University of Edinburgh, UK
[2] The Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, UK

1.3 Corresponding author and email address

Natalie Ring, nring@ed.ac.uk

1.4 Keywords

AMR, AST, nanopore sequencing, metagenomics, *in silico* prediction

1.5 Repositories:

Sequencing data for all samples sequenced for this study are available at the NCBI under BioProject ID PRJNA1292816 (SRA accessions for all datasets used here are available in **Supplementary Table S1**).

2. Abstract

Antimicrobial resistance (AMR) presents a pressing need to ensure that the right antimicrobials are used to target the right microbes at the right time. Ideally, the appropriate antimicrobial is selected after patient samples have been cultured and assessed with antimicrobial sensitivity testing (AST). However, the time needed for culture-based diagnosis leads to immediate empirical treatment, often with broad-spectrum and/or high-tier antimicrobials. Direct nanopore metagenomic whole genome sequencing to identify pathogens and predict their antimicrobial resistance is a rapid and patient-side alternative. A limitation of this approach is potential inconsistencies in *in silico* predicted AMR phenotypes. Here, we benchmarked the current performance of *in silico* AMR prediction strategies for nanopore-generated long read data. Using nanopore data paired with AST phenotyping for 201 samples, we assessed the impact of basecalling mode, data volume, and assembly strategy, and compared the performance of eight *in silico* AMR prediction tools with seven AMR databases. We found that basecalling accuracy mode does not affect the overall accuracy of *in silico* AMR predictions, but assembly strategy and data volume both do. Prediction tools using the ResFinder database scored best for balanced accuracy (0.80 ± 0.02 for both ResFinder and ABRicate), whilst DeepARG scored best for sensitivity (0.65 ± 0.03). However, even the best performing *in silico* AMR prediction strategy missed some resistance identified by lab-based AST. *In silico* AMR prediction can therefore supplement lab-based AST, but cannot yet replace it.

3. Impact statement

Antimicrobial resistance (AMR) is threatening modern standards of human and veterinary healthcare. Rapid and patient-side diagnostic tests are needed to diagnose bacterial infections and allow clinicians

to select effective antibiotics. Current tests based on bacterial cultures take several days, which may delay diagnosis and treatment, or lead to inappropriate "just in case" treatment while waiting for the results. In contrast, nanopore metagenomic whole genome sequencing can identify bacterial infections and predict which antibiotics will be effective in minutes to hours. However, the accuracy of these tests is uncertain. We therefore compared the performance of eight AMR prediction tools and seven databases of AMR determinants, using 201 bacterial samples with known antibiotic susceptibility and resistance. We found that the sensitivity (i.e. false negative rate), specificity (i.e. false positive rare) and overall accuracy of the tools and databases varied. In particular, even the best performing AMR prediction methods missed some AMR. Therefore, while these tools are useful for rapid and patient-side diagnosis and treatment decisions, they still have limitations and should be used alongside bacterial cultures and antibiotic sensitivity testing.

4. Data summary

Sequencing data for the samples sequenced for this study are available at the NCBI under BioProject ID PRJNA1292816 (SRA accessions for all datasets used here are available in **Supplementary Table S1**).

All commands and code used can be found at:
https://github.com/nataliering/nanopore_AMR_tools/

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

5. Introduction

Antimicrobial resistance (AMR) is an immediate and ongoing global threat that will impact health, food security and economic development (1, 2). Over eight million annual AMR-associated deaths are forecasted by the year 2050 (3). The selection pressure from inappropriate and indiscriminate antimicrobial use exerts a selection pressure favouring the development and dissemination of bacterial AMR (4). To avoid this, we need to use the right antimicrobial, for the right infection, at the right time. This is a one health challenge that encompasses human, production animal and companion animal healthcare.

Faster diagnosis to confirm a bacterial infection and determine its antimicrobial susceptibility will improve antimicrobial stewardship. The current gold-standard methods involve culturing samples to identify the bacteria, followed by antibiotic sensitivity tests (ASTs). Commonly used ASTs include disk diffusion, minimum inhibitory concentration (MIC) test strips and broth microdilution assays (5, 6). These tests require a minimum of 16 hours, but it often takes significantly longer to obtain the results. Immediate empirical antimicrobial treatment may be unnecessary (where there is no bacterial infection) and/or may employ unnecessarily broad-spectrum or high-tier antimicrobials through fear of compromising patient outcomes. Excessive and inappropriate antimicrobial use is a major driver of AMR (4). In addition, culture-based ASTs are effectively "a sample of a sample" (i.e. patient sample – culture – culture sample – AST) and favour less fastidious bacteria. AST results may not therefore reflect the whole bacterial population and antimicrobial susceptibility patterns in the patient's infection.

Metagenomic whole genome sequencing (mWGS) using Oxford Nanopore Technologies' (ONT) long-read sequencers is a rapid culture-free technique that involves extraction and sequencing of all genomic DNA (gDNA) in a sample. The unbiased DNA sequencing data can be used to identify pathogens and to detect genes and mutations associated with AMR in a little as 10 minutes (7, 8) . Progress on developing mWGS-based rapid diagnostics pipelines has accelerated in recent years, and

ONT's platforms have been used to pilot rapid diagnostic methods for mWGS from clinical samples such as sputum, respiratory fluid, blood, urine, and skin swabs (9-24).

Robust, reliable and reproducible bioinformatic tools and pipelines will be vital to implement rapid mWGS-based diagnostics in clinical settings (25). A multi-laboratory study found wide discordance between the AMR phenotypes predicted by different laboratories using the same genomic datasets (26). This discordance occurred as a direct result of differences in how each laboratory chose to process the data, and the bioinformatic tools and databases used for the AMR predictions. Similarly, a comparison of tools to identify AMR from Illumina short-read datasets showed wide variability in their sensitivity, specificity and balanced accuracy (27). No single tool performed equally well across every antibiotic class; some performed very well in terms of sensitivity but relatively poorly in terms of specificity.

Here, we analysed the performance of eight different AMR detection tools: ABRicate (28), StarAMR (29), ResFinder (30), AMRFinderPlus (31), c-SSTAR (32), RGI (33), DeepARG (34) and AMR++ (35) on datasets produced using the ONT sequencers that would be used for mWGS-based rapid diagnostics. We sourced a collection of 201 ONT datasets with matched lab-based AST phenotypic AMR profiles. We included tools which require the assembly and annotation of sequencing reads into consensus contigs as well as read-based tools. The tools use different methods to detect AMR determinants, including *k*-mer-based alignment, Hidden Markov Models (HMM), and deep machine learning. Some tools offer a choice of AMR databases (e.g. ABRicate) whilst others are based around their own database (e.g. ResFinder, DeepARG). We tested all possible combinations and options for the tools as well as addressing issues such as volume of data required, the downstream effects of different basecalling accuracy options, and the optimal assembly strategy. We also compared the performance of tools at the level of antibiotic class and individual antibiotics.

6. Methods

Data collection and curation

An NCBI PubMed database search was conducted in May 2025, using various combinations of the terms "nanopore", "AMR", "MinION" and "AST". The resulting literature hits were manually reviewed to determine if 1) their raw sequencing data was publicly available and 2) their datasets included *vitro* antimicrobial sensitivity testing (AST) results. If a paper fulfilled both of the above criteria, the relevant datasets were downloaded, and the following metadata was recorded for each sample (full data in **Supplementary Table S1**): sample accession, sample type, MinION flow cell version, library preparation kit, basecaller used (and, where available, version and accuracy mode), species identified in original publication, AST results, and original publication reference.

Read preparation

Data were downloaded in fastq format from the NCBI sequence read archive (SRA) using fasterq-dump v2.11.3 (36). NanoStat v1.6.0 (37) was used to determine mean read length, number of reads, and overall number of bases. Datasets were excluded if their mean read length was less than the standard nanopore quality cut-off of 1,000 bp (38-41) or their overall number of bases was less than 25 Mb (around 5X coverage of the "average" bacterial genome (42-45)). A set of 64 additional isolates (**Supplementary Table S7**) were sequenced specifically for this study, representing species not otherwise found in the above literature search. See **Supplementary Materials SM1** for full DNA extraction, library preparation and sequencing methods, and **Supplementary Table S1** for sample accessions.

Porechop v0.2.4 (46) was used to trim adapters from the fastq read sets which passed quality control (QC), and Filtlong v0.2.0 (38) was used to filter longer read sets down to 0.5 Gb of the longest and highest quality fastq reads. Where an AMR tool based its predictions on raw fastq reads alone, the output Filtlong reads were used.

Genome assembly and draft annotation

The trimmed reads were assembled with two different long-read assembly tools, Minimap2 v2.24-r1122/Miniasm v0.3-r179 (47, 48) and Flye v2.9 (49). The Flye assembly was polished using Medaka v2.0.1 (50), and the Flye, Miniasm and Medaka assemblies were all annotated with Prokka v1.14.6 (51). All three assemblies (Miniasm, Flye and Medaka), along with the raw reads and annotations where applicable, were used to assess the performance of the tested AMR tools to determine the optimal assembly strategy. More details about each assembly tool can be found in the **Supplementary Materials** (**SM2**).

AMR tools and databases tested

We tested a variety of assembly- and read-based tools, with every relevant database available to each tool. The tools and databases are summarised in **Table 1** and more details are in the **Supplementary Materials** (**SM3**).

For tools which used assembled contigs for AMR prediction, the Miniasm, Flye and Medaka assemblies were all tested with each tool/database combination. See **Supplementary Table S2** for a full list of the 42 tool/database combinations.

Some of the tools tested here, such as ResFinder, offer species-specific AMR prediction for a limited number of species, facilitating the detection of AMR caused by species-specific point mutations. Where a sample was known to contain one of these species, all possible combinations were also trialled with this mode. **Supplementary Table S3** shows which tools were used with which species. Samples containing multiple species were not included in the species-specific tests.

HAMRonization v1.1.9 (54) was used to convert the results of all tools into a consistent tabular form**.** See **Supplementary Methods SM4** for details of the additional steps which were taken to ensure standardisation of the way results were reported across different tools.

Table 1. AMR detection tools and databases tested here

Tool	Version	AMR database(s)	AMR database version	Input file(s)
ABRicate (28)	1.0.1	NCBI(31), CARD(33), ResFinder(30), ARG-ANNOT(52), MEGARes 2.0(53)	14 Jan 2025	Assembly
AMRFinderPlus (31)	4.0.23	NCBI	2025-03-25.1	Assembly and Annotations
ResFinder (30)	4.7.2	ResFinder & PointFinder	2.6.0 & 4.1.1	Reads or Assembly
c-SSTAR (32)	2.1.0	ResGANNOT	NA	Assembly
RGI (33)	6.0.4	CARD	4.0.1	Assembly, or Assembly and Annotations
StarAMR (29)	0.11.0	ResFinder	2024-08-06	Assembly
DeepARG (34)	1.0.2	DeepARG DB	NA	Reads, or Assembly and Annotations
AMR++ (35)	3.0.6	MEGARes	NA	Reads

Calculating performance metrics

For each sample, the original AST results were collated to create a per-antibiotic "phenotypic AMR" table. In addition, the Comprehensive Antibiotic Resistance Database (CARD) ontology (33) was used to assign antibiotics in our phenotypic AMR table to their drug class: aminoglycosides, beta-lactams, fluoroquinolones, glycopeptides, macrolides-lincosamides-streptogramins (MLS), phenicols, sulfonamides, tetracyclines, trimethoprim (diaminopyrimidines), and polypeptides. Rarer antibiotic classes were excluded from analysis, due to insufficient numbers of resistant samples in our dataset. **Supplementary Table S5** shows which antibiotics were included in each class and which classes/antibiotics were excluded. A second table was then produced, which listed each sample's phenotypic AST results in terms of drug class (**Supplementary Table S6**). For each sample in our phenotypic AMR tables, each antibiotic or class was assigned "R" (resistant), "S" (susceptible) or "NA" (not tested) based on the AST results from the original publications. Where a result in the phenotypic AMR table was "I" (Intermediate), it was considered to be resistant.

Each set of results was then analysed at the level of specific antibiotic and antibiotic class. The tool results were then compared with our phenotypic AMR table, and each antibiotic/class was scored as True Positive (TP), False Positive (FP), True Negative (TN) or False Negative (FN) for each sample. A TP occurred when both the phenotypic AMR result and the *in silico* tool prediction were "R" for an antibiotic/class for that sample. A FP occurred when the phenotypic result was "S" but the tool predicted "R". A TN occurred when both the ground truth result and the tool prediction were "S". A FN occurred when the phenotypic result was "R" but the tool predicted "S". Where an antibiotic or class was not AST tested for a sample ("NA"), that antibiotic or class was excluded from the results comparison.

Performance metrics were then calculated for each tool according to the equations shown in **Table 2**.

Calculating statistical differences

Each set of results were tested for normality and variance using Shapiro-Wilk (55) and Levene's Tests (56). Non-parametric Kruskal-Wallis tests (57) were used to test for significance across all variables with Dunn's tests (58) used to determine pairwise differences between the variables.

Comparing super accuracy, high accuracy and fast basecalling mode

A subset of 64 monoculture samples (all those for which raw data was available, **Supplementary Table S7**) were used to conduct a comparison of the accuracy of AMR prediction when using different basecalling accuracy modes. Where necessary, the unbasecalled fast5 files were converted into pod5 format using pod5 v0.2.15 (59). The Dorado standalone basecaller (versions 0.9.6 and 1.0.2 for data from SQK-RBK004 and SQK-RBK114.24 respectively) was then used to basecall the reads with each of the three accuracy modes: fast, high accuracy (HAC) and super accuracy (SUP), creating three basecalled read sets per sample. Each read set was processed through the data preparation, AMR

Table 2. Equations used to calculate our performance metrics

Metric	Equation
Sensitivity	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Balanced Accuracy	$\frac{Sensitivity + Specificity}{2}$

prediction, and analysis pipeline as described above (although performance metrics were calculated per sample instead of per tool). The metrics were statically analysed between the three levels of accuracy.

Assessing the effect of data volume on AMR prediction accuracy

A subset of 61 samples (all those with at least 500 Mb of starting data, **Supplementary Table S8**) were used to assess the effect of initial data volume on the eventual accuracy of AMR prediction. Each read set was processed through the data preparation, AMR prediction, and statistical analysis pipeline as above, except that instead of filtering to 500 Mb with Filtlong the read sets were filtered to 250 Mb, 100 Mb, 50 Mb and 25 Mb. Including the read set originally filtered to 500 Mb, this produced five different sets of AMR predictions, representing 100X, 50X, 20X, 10X and 5X coverage of the "average" length bacterial genome (42). The performance metrics for each level of filtering were compared in the same way as the basecalling modes, as described above.

Comparing monocultured isolates with metagenomic samples

Of the 201 total datasets gathered here, 151 were from monocultured isolates and were used in the main AMR tool comparison described above. The remaining 50 datasets were from samples with potential host contamination or containing multiple species; these datasets were analysed separately to assess AMR tool performance for "difficult" sample types.

Tool resource usage

To compare resource usage by each tool tested here, each tool was run on one sample (NR149) basecalled with super accuracy Dorado v0.9.6, assembled with Flye and polished with Medaka, originally 155 Mb of reads). As the AMR prediction tools were being run, the system was limited to 2 CPUs and a total of 8 GB RAM. Wall clock time and maximum RAM usage were recorded using `/usr/bin/time -v`.

7. Results

A curated dataset of 201 samples to evaluate AMR prediction tools

A total of 201 samples met our selection criteria of having both ONT reads and AST phenotypes available, as well as passing our QC process. **Supplementary Table S1** gives the full metadata of all analysed samples and **Table 3** gives the references of the original papers. The samples were sequenced on a variety of different flow cell types (R9.4.1, R9.5, R10.4.1), sequencing libraries were prepared using a variety of kits (for example, SQK-RPB004), and basecalling was performed with a variety of different basecallers (Albacore, Guppy, Dorado) and basecaller software versions. Starting volumes of data ranged from 28 Mb to 16.5 Gb, with a mean of 662.51 Mb, median of 175.90 Mb and standard deviation of 2031.42 Mb. Mean read length per dataset ranged from 1.01 to 11.12 kbp, with a mean of 4.43 kbp, median of 4.00 kbp and standard deviation of 1.87 kbp. 151 samples were monocultured isolates, while 50 contained multiple species or potential host contamination. The monocultured isolates were used for the majority of the tests here; the remaining 50 were used to test how the tools performed with more "difficult" samples.

27 species were included, with *Escherichia coli* the most common species (40 samples) and *Staphylococcus* the most widely represented genus (40 samples across 7 species). On average, each species was represented by 6.3 samples (range=1-40 samples). For the purposes of statistical power, only antibiotic classes recorded as resistant in ASTs from at least 5 (~2%) of our samples were

considered (**Supplementary Table S5**). Of these, beta-lactam resistance was most common (52.24% of samples). The mean number of samples in our dataset resistant to each antibiotic class was 35.7 (SD=30.11). See **Supplementary Material SF1** for full species and AST details.

Basecalling accuracy mode does not significantly affect overall AMR prediction accuracy

A subset of 64 monocultured samples were basecalled with Dorado in three different accuracy modes: fast, high accuracy (HAC) and super accuracy (SUP). Each set of reads was then processed through the rest of the assembly, analysis and statistics pipeline with the results stratified by basecalling accuracy. **Figure 1** shows that no significant differences were found between the accuracy modes across any of the performance metrics. All samples in our dataset were therefore carried forward for further analysis, regardless of original basecalling accuracy

Flye+Medaka produces the best balanced accuracy, whilst reads-only produce the best sensitivity across all tools

Performance metrics were calculated for all 151 monocultured samples across all tool/database combinations, then stratified by assembly strategy ("flye", "medaka", "miniasm" and "reads"). As shown in **Figure 2**, although no tool performed significantly better across all metrics, the reads-only strategy produced the best sensitivity, whilst the Flye assembly polished with Medaka produced marginally better results than Miniasm and reads for balanced accuracy. The Medaka and reads-only strategies were therefore taken forwards for more detailed analysis across all tools.

ResFinder-based tools produce the best performance metrics across all samples

The mean score for each performance metric across all monocultured samples was calculated and stratified according to prediction strategy (tool, database, Medaka assembly vs reads) to determine the best performing tool overall. **Figure 3** shows the results for sensitivity, specificity and balanced accuracy (the results for precision and F1-Score can be found in **Supplementary Figures SF1 and SF2**). For balanced accuracy, which considers both sensitivity and specificity, the top five strategies are all based on the ResFinder database, with the ResFinder tool itself and ABRicate with its own version of the ResFinder database performing better (0.80 ± 0.02 for ResFinder in read-based mode and both tools in assembly mode) than StarAMR (0.75 ± 0.02) with its own version of the ResFinder database.

Table 3. Original sources for our QC passed dataset samples

Sample reference	Number of QC passed samples
Bialasiewicz et al. 2019 **(60)**	1
Charalampous et al. 2019 **(12)**	27
Cherdtrakulkiat et al. 2020 **(61)**	3
Chopjitt et al. 2020 **(62)**	4
Conzemius et al. 2022 **(63)**	1
Harris Patrick et al. 2024 **(64)**	39
Jahnen et al. 2025 **(65)**	8
Kostova et al. 2024 **(66)**	13
Leggett et al. 2020 **(67)**	4
Lim et al. 2019 **(68)**	2
Ring et al. 2023 **(9)**	17
Smith et al. 2020 **(69)**	10
Ring et al. 2026 (this study)	63
Udaondo et al. 2020 **(70)**	4
Unemo et al. 2016 **(71)** & Sanderson et al. 2020 **(14)**	5

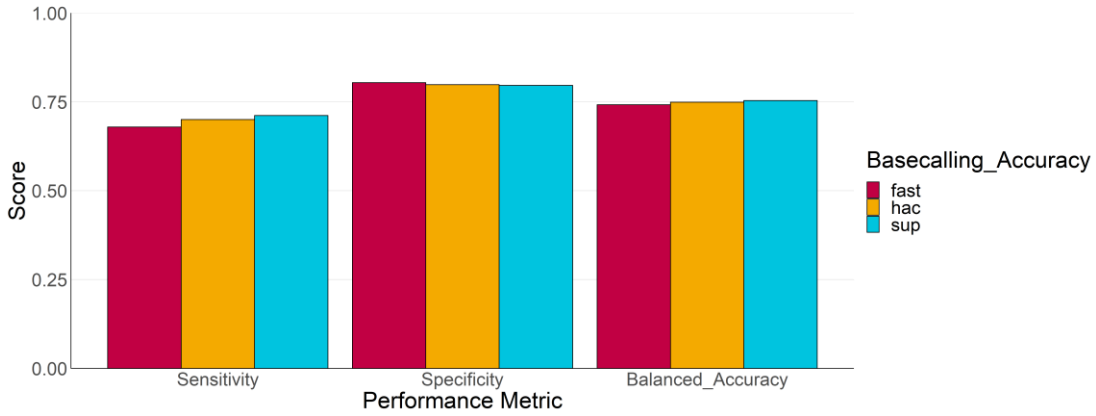

Figure 1. Performance metrics of our 64-sample subset which were basecalled with three different basecalling accuracy modes in Dorado. Metrics were calculated across all tools and all samples, then stratified by basecalling accuracy; fast, high accuracy ("hac") and super accuracy ("sup"). No significant differences were observed, according to the Kruskal-Wallis test.

Across the whole 151-sample monoculture dataset, sensitivity scores were generally low. The read-based DeepARG mode scored the highest at 0.65 ± 0.03. The lowest sensitivity scores were for both RGI strategies (assembly and annotations) at 0.09 ± 0.01, meaning that most AMR genes were not called.

Conversely, both RGI modes ranked the highest in specificity (0.93 ± 0.1 for assembly and annotation-based). However, as specificity relates to the absence of false positive calls, RGI scoring highly does not mean that it performed particularly well; rather, RGI made the fewest calls simply because it made so few positive calls (false or true) in general. Read-based DeepARG performed best in sensitivity, but worst in specificity (0.52 ± 0.02), indicating a high number of apparent false positives alongside the total positives. In contrast, the two modes of the ResFinder tool score relatively well for specificity: 0.88 ± 0.02 for both read-based and assembly.

Overall, although read-based DeepARG scored best for sensitivity, the ResFinder tool with ResFinder database in either assembly or read mode performed best across all metrics when taking the means of the 151 monocultured samples analysed here.

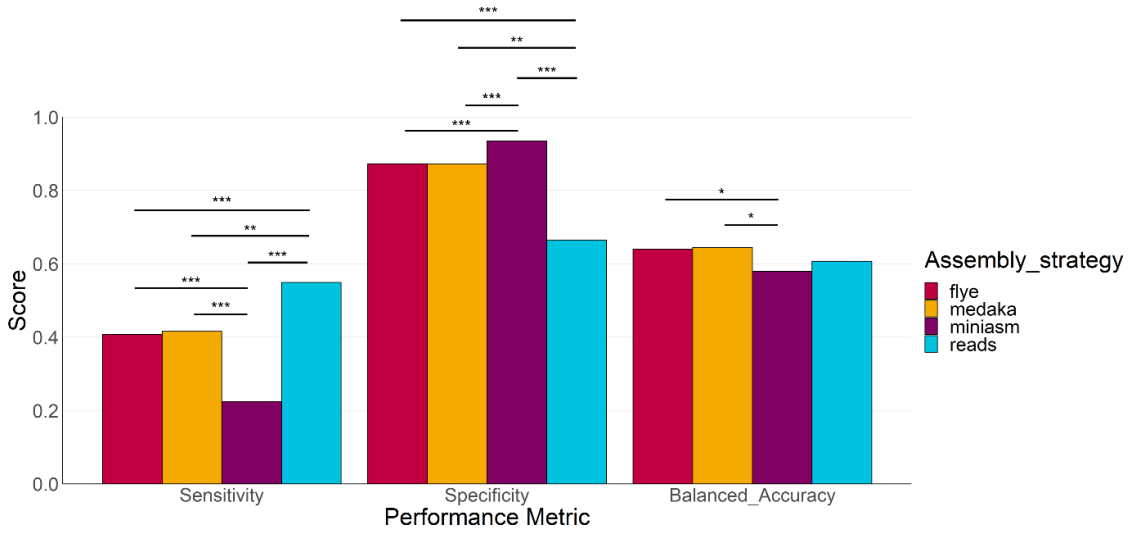

Figure 2. Overall performance metrics across 151 monocultured samples and all tools, stratified by assembly strategy. "flye" is assembly with Flye in nano-hq mode, "medaka" is assembly with Flye followed by polishing with Medaka, "miniasm" is rapid assembly with Minimap2 and Miniasm, and "reads" is prediction from raw reads, with no assembly step (of the tools test here, only AMR++ and ResFinder include a "read-only" mode). Statistical significance was calculated using the Kruskal-Wallis Test and Dunn's Test.

*** indicates p<0.0005, ** indicates p<0.005, * indicates p<0.05

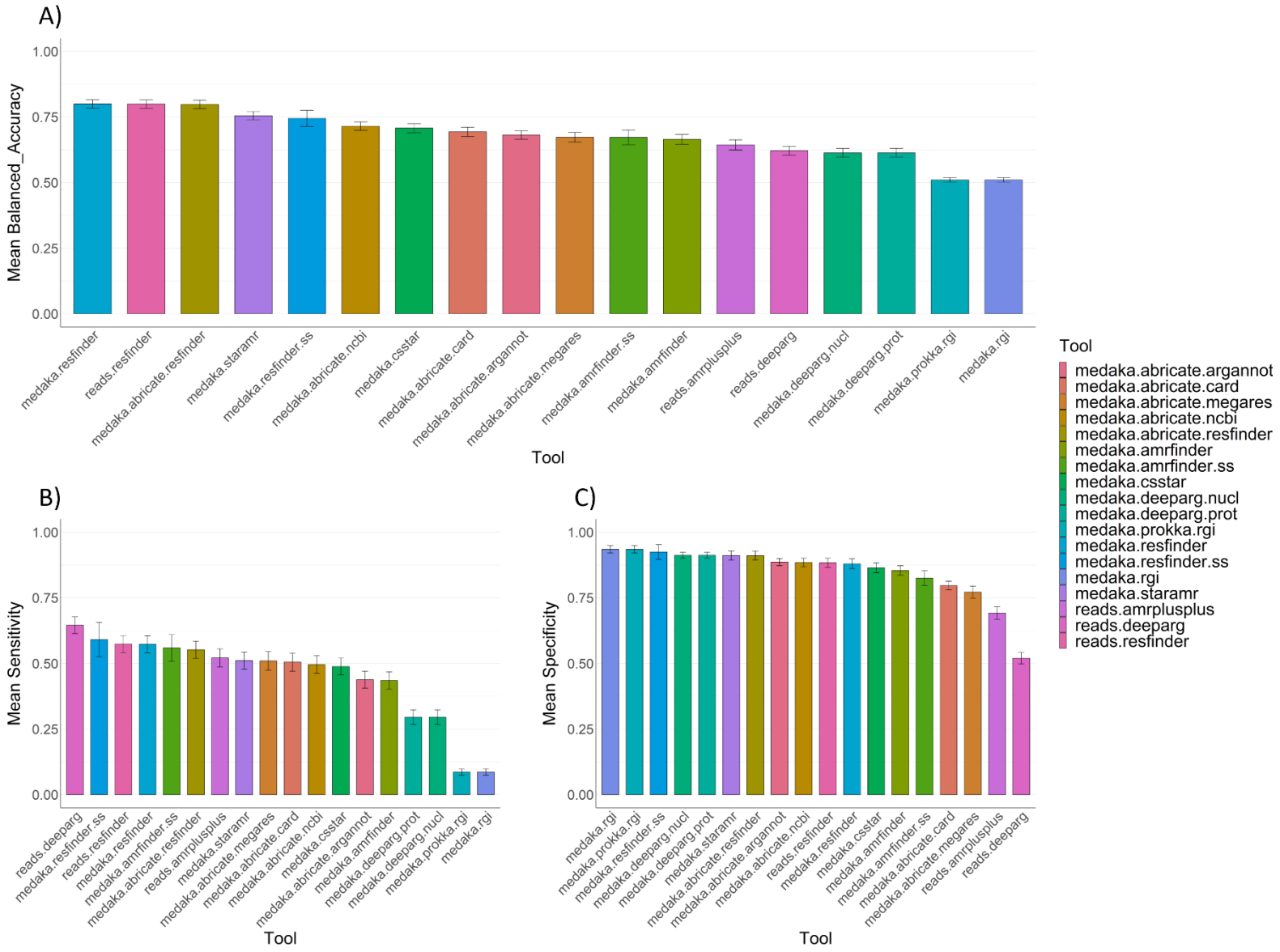

Figure 3. Performance metrics for each AMR prediction strategy, calculated at the level of specific antibiotics per sample then stratified by prediction strategy. Prediction strategy names follow the general structure assembly/reads.tool.(database).(ss), where (database) is only relevant for those tools with multiple database options and (SS) is only relevant for those tools with a species-specific mode. A) shows the mean balanced accuracy per strategy across all 151 monocultured samples in our dataset, whilst B) shows the mean sensitivity and C) shows the mean specificity. Error bars show Standard Error.

Making predictions based on antibiotic classes improves sensitivity, but not specificity

We tested whether grouping antibiotics into their classes (**Supplementary Table S4**) produced better predictions, as some of the tools tested reported AMR on the level of class only. **Figure 4** shows the per-class results for balanced accuracy, sensitivity and specificity, stratified by tool/database/reads vs assembly. The strategies which performed best on the level of specific antibiotic also performed best here: ResFinder-based tools were the top five scorers in balanced accuracy, whilst read-based DeepARG performed best in sensitivity but worst in specificity. Compared to predictions for specific antibiotics, read-based and assembly-based ResFinder scored worse for specificity: 0.88 per specific antibiotic and 0.81 per antibiotic class. In contrast, sensitivity was significantly improved: 0.57 per specific antibiotic and 0.73 per antibiotic class for ResFinder, and 0.65 per antibiotic and 0.80 per class for DeepARG.

Some antibiotic classes are predicted more accurately than others

We also investigated whether AMR predictions for some classes were more robust than for others. **Figure 5** shows the mean balanced accuracy score across our 151 monocultured samples for all prediction strategies, stratified by antibiotic class. This varied significantly according to antibiotic class, ranging from 0.47 for fluoroquinolones to 0.80 for glycopeptides.

Figure 6 visualises the balanced accuracy of AMR predictions for each antibiotic class further stratified by prediction strategy, to identify potential strengths and weaknesses. All tools have relatively strong results for both glycopeptides and trimethoprim, but weaker results for fluoroquinolones. The ResFinder-based tools scored highly across most classes. Tetracyclines showed the highest within-class variability according to prediction strategy, with some (ResFinder-based) scoring well and others (ABRicate with the MegaRES or CARD databases) scoring poorly.

AMR prediction tools for samples with mixed populations or host contamination perform similarly to more simple samples

To test how confounding factors such as host DNA contamination affected the prediction accuracy of our AMR tools, we processed 50 metagenomic/mixed species samples through the same analysis pipeline. This showed the same pattern as the monoculture results: ResFinder-based tools performed best for overall balanced accuracy whilst DeepARG performed best for sensitivity but worst for specificity. However, although the performance pattern remained the same the actual scores observed for the “difficult” samples were lower than those of the monocultured samples. Read-based ResFinder scored 0.75 ± 0.02 for balanced accuracy compared to 0.80 ± 0.02 for the monocultures, whilst DeepARG scored 0.58 ± 0.03 for sensitivity compared to 0.65 ± 0.03 for monocultures.

Increasing starting data volume raises the overall accuracy of AMR predictions

The effect of data volume on the AMR prediction performance was assessed by analysing a subset of our samples filtered to 500 Mb, 250 Mb, 100 Mb, 50 Mb and 25 Mb. **Figure 7** shows that for each sequential increase in starting data volume, a significant improvement is seen in the balanced accuracy of the AMR predictions. However, this decreased with higher data volumes: the incremental improvement in sensitivity between 25 Mb and 50 Mb is larger than that between 250 Mb and 500 Mb. Although not tested here, the trends in **Figure 7** suggest that 1000 Mb of data would produce better predictions still.

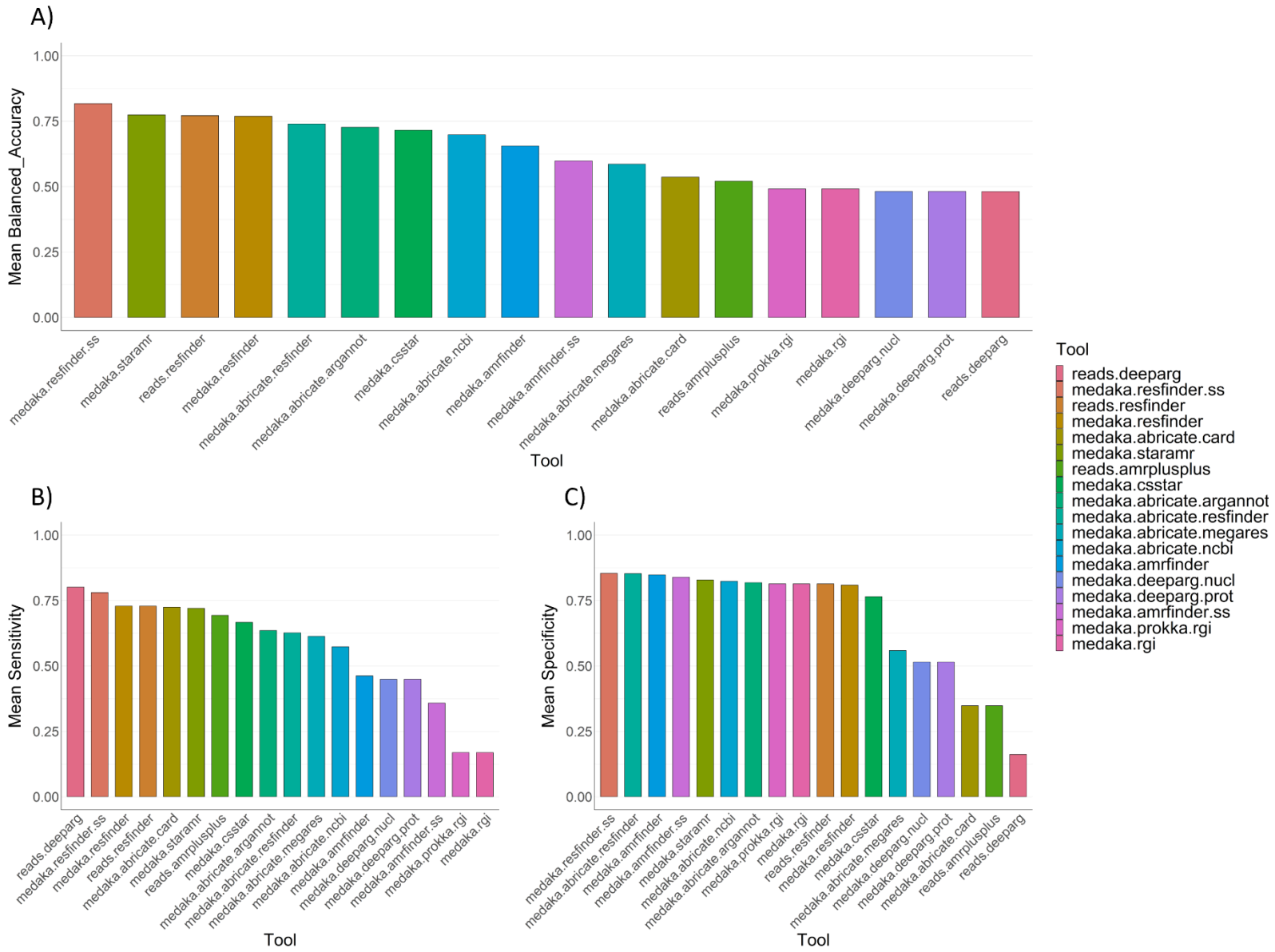

Figure 4. Performance metrics for each AMR prediction strategy, calculated at the level of antibiotic class and stratified by prediction strategy. Prediction strategy names follow the general structure assembly/reads.tool.(database).(ss), where (database) is only relevant for those tools with multiple database options and (SS) is only relevant for those tools with a species-specific mode. A) shows the mean balanced accuracy per strategy across all 151 monocultured samples in our dataset, whilst B) shows the mean sensitivity and C) shows the mean specificity.

318

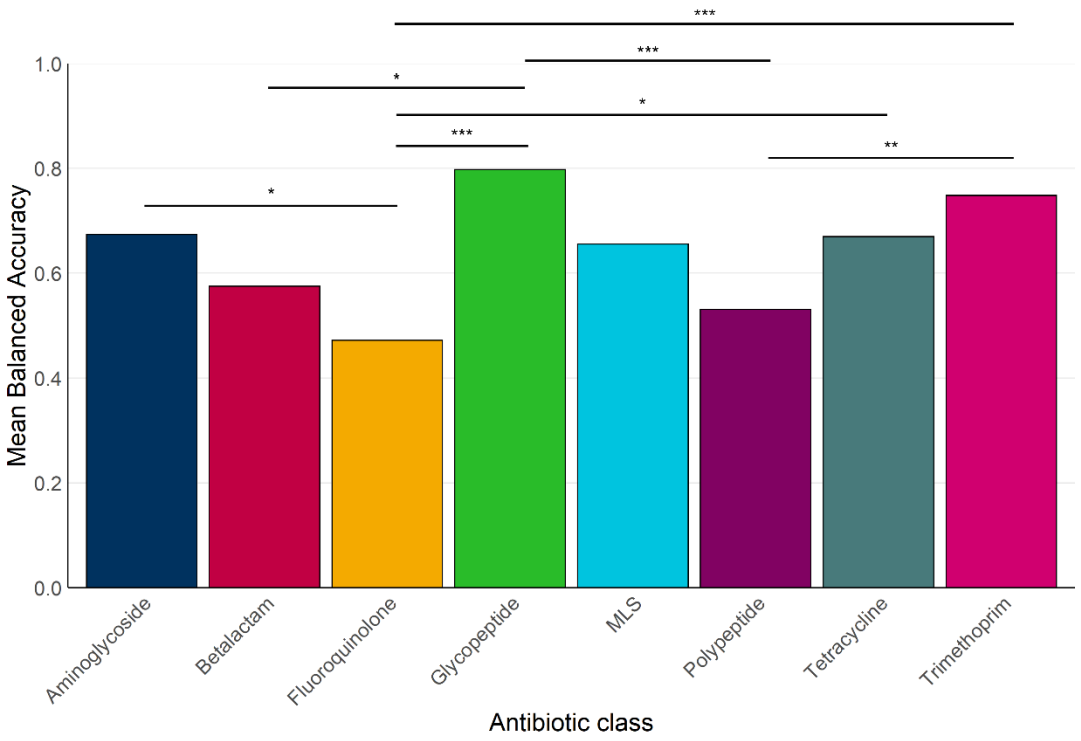

Figure 5. Mean balanced accuracy across all monocultured samples for all prediction strategies, stratified by antibiotic class. Statistical significance was calculated using the Kruskal-Wallis Test and Dunn's Test. P-values show pairwise differences.

*** indicates p<0.0005, ** indicates p<0.005, * indicates p<0.05

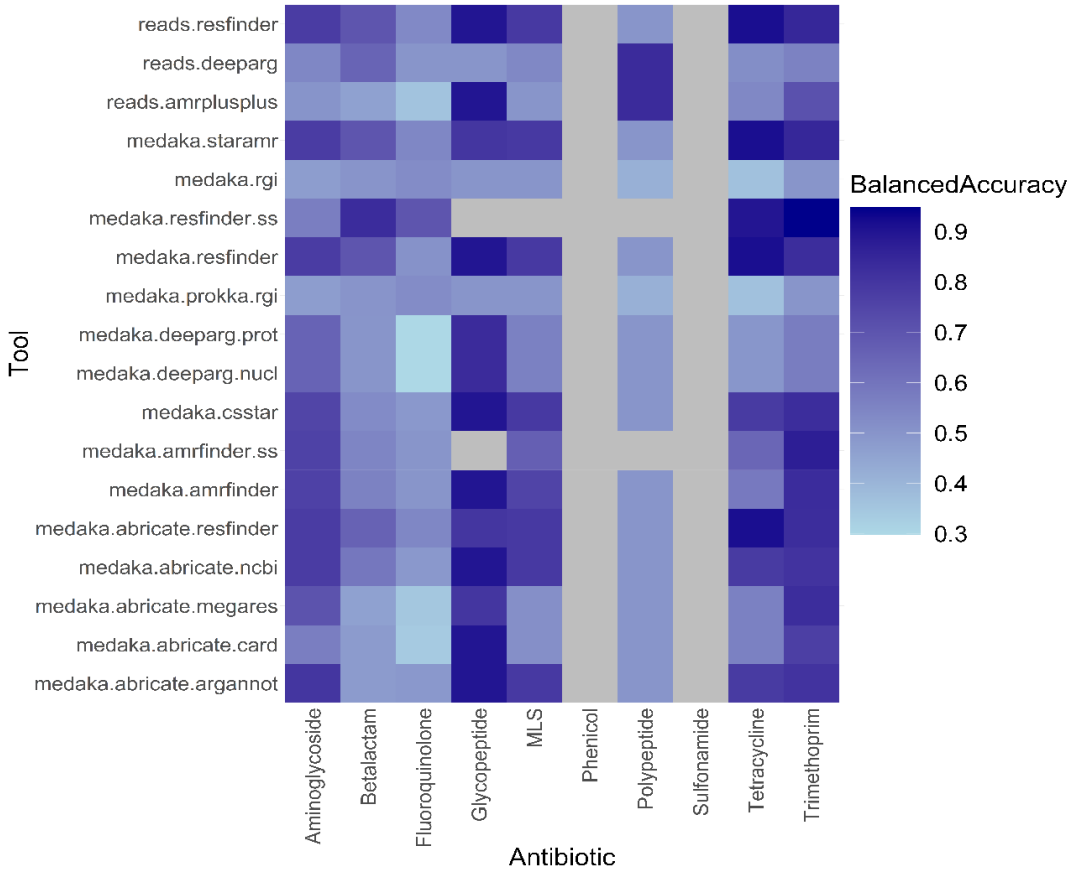

Figure 6. Heatmap showing the performance of each prediction strategy for each antibiotic class, in terms of balanced accuracy. The results for Phenicol and Sulfonamide are grey because too few samples had positive AST results to calculate meaningful statistics. Likewise, the two species-specific strategies have grey cells for several classes for which the samples for those species had limited AST-proven resistance.

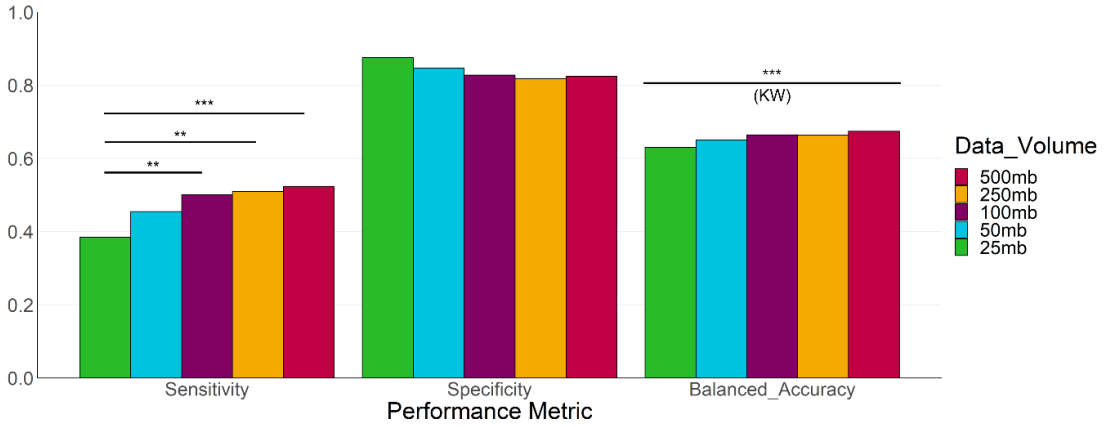

Figure 7. Performance metrics of our 61-sample subset which were filtered to different volumes of raw data by Filtlong, from 500 mb down to 25 mb. Metrics were calculated across all tools and all samples, then stratified by starting volume of data. Statistical significance was calculated using the Kruskal-Wallis Test and Dunn's Test.

*** indicates p<0.0005, ** indicates p<0.005, * indicates p<0.05.
*** (KW) indicates p<0.0005 in the Kruskal-Wallis test, but no significant pairwise differences were observed in the Dunn's Test.

Tool resource usage

The wall clock time and maximum RAM required were recorded for each tool/database combination run for a single sample (NR149) with an initial volume of data of 155 Mb (**Figure 8** and **Supplementary Table S10**). All tool/database combinations ran quickly; run times ranged from 1 second (ABRicate with any database) to 66 seconds (AMRFinderPlus) with a mean of 16.6 seconds (SD=15.12). Maximum RAM usage was relatively low and easily within the resources of most machines (range 77 MB to 1.15 GB, mean=375 MB, SD=364 MB). Maximum RAM usage did not necessarily correlate with time; although AMRFinderPlus took the most time of the eight tools, it required only the fourth highest maximum RAM. Overall, c-SSTAR and ABRicate required the least resources in terms of both time and maximum RAM usage, whilst AMR++ and DeepARG required the most.

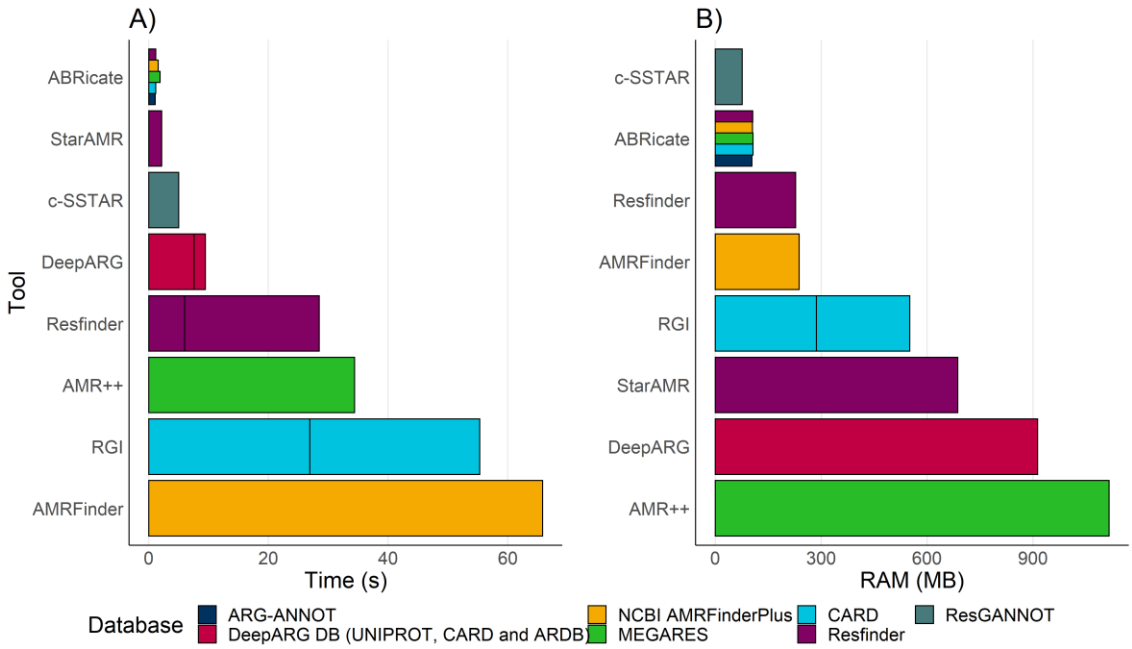

Figure 8. Resource usage per tool and database. A) shows the wall clock time taken to run each tool on a single sample (NR149, starting data volume 155 Mb) in seconds. B) shows the maximum RAM required by each tool whilst processing the same sample.

8. Discussion

Benchmarking AMR prediction tool performance with ONT data could facilitate future clinical implementation

In this study, we set out to determine how well existing computational tools can detect AMR in genomic DNA datasets derived from ONT sequencers, by comparing *in silico* AMR predictions with paired *in vitro* AST results. Although there are many *in silico* AMR detection tools available, the majority were originally designed with Illumina short reads in mind. Indeed, some commonly-used tools (e.g. ARIBA (72), SRST2 (73), Groot (74)) could not be tested here, as they are for short read data only. The new Amira tool is designed specifically for long reads; however, at the time of testing it was only available for a limited number of species and thus we did not include it (75).

Even those tools that can be used with ONT data were still originally optimised to detect AMR genes and mutations based on the characteristics of Illumina reads. These are short (<500 bp) reads with high (often >99.5%) raw accuracy, and any errors tending to occur as substitutions at the end of reads (76). These tools therefore may not perform as well with reads from ONT sequencers, which are long (typically >1,000 bp) and historically had a higher raw error (1-5%), with errors tending to be insertions or deletions (indels) (77). Although raw accuracy is much higher using the latest ONT sequencing chemistry and basecaller (R10.4.1 flow cells and Dorado commonly produce raw error <1% (78)), many of the datasets analysed here (147/201) were generated using older ONT chemistry or basecallers (R9.4.1 flow cells and Albacore or Guppy). Tools designed with an Illumina error profile in mind may therefore have been too strict, resulting in missed AMR calls. For example, the default mode of RGI only calls exact matches as hits, thus explaining its surprisingly poor performance here; even minimal indels could cause frameshifts in the gene sequences, which would prevent their accurate translation into protein sequences leading to false negative results (33). Applying RGI's "--include_loose" flag could allow for higher sensitivity with ONT data.

With metagenomic ONT sequencing likely to used diagnostically in clinical settings, it will be vital that robust computational methods are used to identify species and detect AMR profiles. Accurately classifying AMR in genomic DNA datasets is also useful in other contexts, such as monitoring the environmental load of AMR-carrying bacteria within healthcare or One Health settings (79-82). Whilst there are well-established and maintained guidelines for the interpretation of *in vitro* lab-based ASTs (83), no such guidelines currently exist for interpreting *in silico* DNA-based tests. Understanding the strengths and weaknesses of the possible strategies to detect AMR from genomic data will therefore be important when determining which methods to implement in clinical settings. This has been done for short reads (27) but, until now, not for ONT long read sequencing data.

Choice of basecalling mode does not significantly affect AMR prediction accuracy, but choice of assembly strategy does

Before comparing the performance of the AMR prediction tools and databases, we optimised the strategy for preparing sample data. The first step in analysing ONT data is basecalling, i.e translating the raw electrical signal (the "squiggle") into a DNA sequence. Early basecallers such as Albacore, which was used to basecall 34 of the samples in our dataset, did not have different basecalling modes. However, Guppy and Dorado have three accuracy modes: fast, high accuracy (HAC) and super accuracy (SUP). SUP basecalling requires much higher computing resources and takes an order of magnitude longer than fast basecalling, albeit with the benefit of significantly increased per-base accuracy in the resulting fastq reads. HAC basecalling is middle-ground option between fast and SUP. ONT do not

currently publish an expected read accuracy for fast basecalling, but quote a median raw read accuracy of 99.4% and 99.7% respectively for HAC and SUP (84).

SUP basecalling within a rapid diagnostic timeframe would require a GPU-powered computer or GridION, which may not be economically viable in all clinical settings. However, despite a slight trend in our results towards more accurate AMR predictions from fast to HAC to SUP the differences were not significant. Our results therefore indicate that even fast basecalling, which is feasible using a standard laptop with 2-4 CPUs and 8-16 Gb RAM, may produce AMR predictions with comparable accuracy to HAC or SUP basecalling. Further, our comparison of the computational resources required for each AMR prediction tool showed that the RAM required is well within the realms of a standard laptop (we tested each tool with 2 CPUs and 8 Gb RAM); using fast basecalling could thereby enable all analysis to be carried out locally, on the same machine.

The next step in data preparation after basecalling is assembly into contigs. A previous comparison of AMR prediction tools focussing on short read data included tools that analysed reads but not assembled contigs (27). Although assembling reads into contigs requires additional processing steps and therefore additional time, the processing of individual reads into a consensus sequence greatly improves the per-base accuracy (85). For ONT data, which may be less accurate than Illumina data at the read level, computing a consensus may therefore result in more accurate AMR predictions. Our results show that Miniasm, which exchanges accuracy for speed, performed the worst overall, whilst Flye and Flye+Medaka produced results that were not significantly different. This finding is unsurprising, since newer sequencing chemistries and basecallers may not need post-assembly polishing (86). However, since a relatively large number of our samples were not produced using the latest chemistry and basecaller, we chose to use the Medaka assemblies instead of Flye when analysing the performance of each AMR prediction tool separately.

ResFinder DB outperforms the other databases tested here in both per-antibiotic and per-class strategies

We tested a variety of tools with a variety of databases, in both assembly and read-based modes. Where a tool allowed either assembly or read mode to be used, we trialled both. Likewise, where a tool allowed the selection of AMR databases (e.g. ABRicate), we trialled all available. ResFinder, CARD, MEGARes and NCBI's AMRFinderPlus are among the most commonly used AMR gene databases (87), and all were trialled with more than one tool. Our results show that, in terms of balanced accuracy, the ResFinder database outperforms the others for AMR prediction from ONT data, independent of the tool used to screen the database. Of the ResFinder-based tools, the standalone ResFinder tool performed best.

ResFinder performed almost exactly the same when used with assembled contigs or with raw reads, suggesting that important AMR information from the reads is not lost during consensus building. ResFinder performed best in species-specific mode, although the ResFinder command line tool only includes species-specific screening for a very limited number of species. Of note, the ResFinder webserver (88) offers species-specific screening with PointFinder for far more species than the command line tool. In addition, the webserver is user-friendly and outputs an easily interpreted results report. One accurate AMR prediction strategy for rapid diagnostics would therefore be to simply use the ResFinder webserver with either assembled fasta or raw fastq reads, selecting any species known to be in the sample. However, this relies on a shared public web service; analysis turnaround time may therefore vary with server load and concurrent usage. In addition, the need to upload genetic data to an online tool could raise ethical concerns, particularly when working with metagenomic samples originally from humans, which may still contain human DNA sequences.

Are all false positives really false positives?

Although DeepARG outperformed other strategies in sensitivity, its poor score for specificity suggests a high likelihood of false positives compared to AST. If used clinically to support a choice of antibiotic, DeepARG could therefore lead to more antibiotics being excluded than by AST. However, the term "false positive" may be misleading. There are instances in which isolates with known resistance determinants are deemed susceptible by ASTs. This suggests that the AMR gene is present but not phenotypically expressed at levels detectable under standard laboratory conditions (89, 90). Such "silent" or poorly expressed resistance genes still reflect a latent AMR potential that could be clinically relevant, particularly if expression is induced *in vivo* or under antibiotic selection pressure. In addition, ASTs are not 100% accurate, with reproducibility and other outcomes ranging from 90-99% (91). Despite its seemingly worse performance than Resfinder, DeepARG may therefore represent a more conservative option for complementing clinical decisions, and is a strong choice for surveillance of the full resistance gene load in metagenomic samples.

Is *in silico* AMR prediction ready for use in a clinical setting with ONT data?

In this study, we identified an optimal strategy for preparing ONT data for AMR prediction, and compared the performance of AMR prediction tools and databases. Tools using the ResFinder database strike the best balance between sensitivity and specificity, while DeepARG performs the best in terms of just sensitivity.

However, even the best performing tool missed some AMR reported by phenotypic AST. Although no diagnostic test is perfect, in order for a test to be trusted for making clinical decisions, its accuracy would ideally approach 100% (92). DeepARG, the best tool for sensitivity, detected only 65% of the AMR in our AST results. A number of factors may have led to this. Our dataset represented a wide variety of sequencing qualities, with some samples dating back to 2019 or earlier, sequenced on less accurate flow cells and basecalled with less accurate algorithms. Some of our samples had minimal data, and our data volume test shows that even 500 Mb might not capture all AMR present in a sample. In addition, some samples contained species less well-represented in the AMR databases and/or more difficult to detect (e.g. *Mannheimia haemolytica*, *Capnocytophaga canismorsus* and *Mycobacterium tuberculosis*). Nonetheless, we included these to ensure our testing was representative of the spectrum of pathogens that could be present in clinical mWGS samples.

Predicting AMR based on antibiotic class instead of specific antibiotic may lead to better sensitivity where there is predictable susceptibility or resistance to all drugs with a class; our results showed an improvement from 65% to 80% for DeepARG. Using this strategy, if a gene was detected which was related to AMR in a specific antibiotic, all other antibiotics from that same class would automatically be excluded. For example, if a gene related to gentamicin resistance was detected, all macrolides would be excluded and an entirely different antibiotic class would need to be used. However, predictive antibiotic susceptibility is variable between classes, bacterial species and AMR genes. In our study, performance for aminoglycosides, glycopeptides, tetracyclines and trimethoprim were reasonably good, whilst that for beta-lactams was less impressive, and performance for fluoroquinolones was very poor. In genomic terms, this may also reflect gaps in AMR databases as well as point mutations rather than specific AMR genes (e.g. in fluoroquinolone resistance) (93). As most tools do not offer a species-specific mode or the option to screen for point mutations, sensitivity to AMR caused by point mutations will be low. Care should therefore be taken with any class-based strategy to avoid excluding antibiotics that may be effective, or including those where resistance is likely. This is particularly important in serious infections and/or multi-drug-resistant scenarios where

clinical outcomes depend on a rapid and accurate choice of antibiotic, and fewer antibiotics are available.

Conclusions

In conclusion, limitations remain for users considering the implementation of AMR prediction tools and databases with nanopore long reads in clinical settings. This mirrors an earlier study with Illumina short reads (27). Universal guidelines for how *in silico* predictions should be interpreted for each antibiotic and antibiotic class are needed. In the meantime, the results of *in silico* predictions should be used with caution, with clinical decisions being made on a case-by-case basis, taking into account the case history, pathogen species, antibiotic class and available drugs. Based on our findings, *in silico* AMR prediction using mWGS is ready to support and augment the results of *in vitro* AST, particularly in rapid patient-side diagnosis and treatment decisions, but it is not yet ready to replace it.

9. Author statements

9.1 Author contributions

Conceptualization: N.R.; Data curation: N.R., A.L., G.K.P.; Formal analysis: N.R.; Funding acquisition: J.R.F.; Investigation: N.R., A.L., R.E., M.K.; Methodology: N.R.; Project administration: D.N.C, J.R.F.; Resources: N.R., A.L., G.K.P., T.N.; Supervision: D.G., D.N.C, J.R.F.; Validation: N.R.; Visualization: N.R.; Writing - original draft: N.R., D.N.C., J.R.F.; Writing - review & editing: N.R., A.L., R.E., M.K., G.K.P., T.N., D.G., D.N.C, J.R.F.

9.2 Conflicts of interest

The author(s) declare that there are no conflicts of interest affecting this work.

9.3 Funding information

This work, and the samples sequenced herein, contribute to a project originally funded by a Dogs Trust Canine Welfare Grant, Direct Diagnostic Genomics Towards Effective Antimicrobial Use ("Dogstails"). The study was further supported by institute strategic grant funding ISP2: BBS/E/D/20002173 and BBS/E/RL/230002C from the Biotechnology and Biological Sciences Research Council (United Kingdom).

9.4 Ethical approval

No ethical approval was required for this study.

9.5 Acknowledgements

The authors are grateful to the Easter Bush Pathology microbiology laboratory for the storage and provision of the samples sequenced specifically for this study. In addition, we thank the authors of the original studies for making their data available for the remainder of the samples analysed here.

10. References

1. Antimicrobial Resistance C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet (London, England). 2022;399(10325):629-55.
2. World Health Organization. Antibiotic Resistance 2020 [Available from: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance.
3. Naghavi M, Vollset SE, Ikuta KS, Swetschinski LR, Gray AP, Wool EE, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. The Lancet. 2024;404(10459):1199-226.

4. Nuttall T. Choosing the best antimicrobial for the job. Vet Rec. 2013;172(1):12-3.
5. Kirby WM, Yoshihara GM, Sundsted KS, Warren JH. Clinical usefulness of a single disc method for antibiotic sensitivity testing. Antibiotics annual. 1956:892-7.
6. Bolmstrom A, Arvidson S, M E, A K. A novel technique for direct quantification of antimicrobial susceptibility of microorganisms. Twenty-Eighth Interscience Conference on Antimicrobial Agents and Chemotherapy; Los Angeles, CA: American Society for Microbiology, Washington, DC.; 1988. p. 325.
7. Břinda K, Callendrello A, Ma KC, MacFadden DR, Charalampous T, Lee RS, et al. Rapid inference of antibiotic resistance and susceptibility by genomic neighbour typing. Nature Microbiology. 2020;5(3):455-64.
8. Steinig E, Pitt M, Aglua I, Suttie A, Greenhill A, Heather C, et al. Genomic neighbor typing for bacterial outbreak surveillance. bioRxiv. 2022:2022.02.05.479210.
9. Ring N, Low AS, Wee B, Paterson GK, Nuttall T, Gally D, et al. Rapid metagenomic sequencing for diagnosis and antimicrobial sensitivity prediction of canine bacterial infections. Microbial Genomics. 2023;9(7):001066.
10. Charalampous T, Alcolea-Medina A, Snell LB, Alder C, Tan M, Williams TGS, et al. Routine respiratory metagenomics service for intensive care unit patients. medRxiv. 2023:2023.05.15.23289731.
11. Charalampous T, Alcolea-Medina A, Snell LB, Williams TGS, Batra R, Alder C, et al. Evaluating the potential for respiratory metagenomics to improve treatment of secondary infection and detection of nosocomial transmission on expanded COVID-19 intensive care units. Genome Medicine. 2021;13(1):182.
12. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783-92.
13. Yang L, Haidar G, Zia H, Nettles R, Qin S, Wang X, et al. Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: a feasibility and clinical validity study. Respiratory Research. 2019;20(1):265.
14. Sanderson ND, Swann J, Barker L, Kavanagh J, Hoosdally S, Crook D, et al. High precision Neisseria gonorrhoeae variant and antimicrobial resistance calling from metagenomic Nanopore sequencing. Genome research. 2020;30(9):1354-63.
15. Street TL, Barker L, Sanderson ND, Kavanagh J, Hoosdally S, Cole K, et al. Optimizing DNA Extraction Methods for Nanopore Sequencing of Neisseria gonorrhoeae Directly from Urine Samples. Journal of Clinical Microbiology. 2020;58(3):e01822-19.
16. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, et al. Same-day diagnostic and surveillance data for tuberculosis via whole genome sequencing of direct respiratory samples. J Clin Microbiol. 2017.
17. Avershina E, Frye SA, Ali J, Taxt AM, Ahmad R. Ultrafast and Cost-Effective Pathogen Identification and Resistance Gene Detection in a Clinical Setting Using Nanopore Flongle Sequencing. Frontiers in Microbiology. 2022;13.
18. Taxt AM, Avershina E, Frye SA, Naseer U, Ahmad R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. Scientific reports. 2020;10(1):7622.
19. Yang L, Haidar G, Zia H, Nettles R, Qin S, Wang X, et al. Metagenomic identification of severe pneumonia pathogens with rapid Nanopore sequencing in mechanically-ventilated patients. medRxiv. 2019:19002774.
20. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J Antimicrob Chemother. 2017;72(1):104-14.
21. Whittle E, Yonkus JA, Jeraldo P, Alva-Ruiz R, Nelson H, Kendrick ML, et al. Optimizing Nanopore Sequencing for Rapid Detection of Microbial Species and Antimicrobial Resistance in Patients at Risk of Surgical Site Infections. mSphere. 2022;7(1):e0096421.

22. Janes VA, Matamoros S, Munk P, Clausen P, Koekkoek SM, Koster LAM, et al. Metagenomic DNA sequencing for semi-quantitative pathogen detection from urine: a prospective, laboratory-based, proof-of-concept study. The Lancet Microbe. 2022;3(8):e588-e97.
23. Irwin AD, Coin LJM, Harris PNA, Cotta MO, Bauer MJ, Buckley C, et al. Optimising Treatment Outcomes for Children and Adults Through Rapid Genome Sequencing of Sepsis Pathogens. A Study Protocol for a Prospective, Multi-Centre Trial (DIRECT). Frontiers in Cellular and Infection Microbiology. 2021;11.
24. Azami ME, Lanet V, Beaulieu C, Griffon A, Schicklin S, Mahé P, et al. BacT-Seq, a Nanopore-Based Whole-Genome Sequencing Workflow Prototype for Rapid and Accurate Pathogen Identification and Resistance Prediction from Positive Blood Cultures: A Feasibility Study. Diagnostics (Basel, Switzerland). 2026;16(1).
25. Libuit KG, Doughty EL, Otieno JR, Ambrosio F, Kapsak CJ, Smith EA, et al. Accelerating bioinformatics implementation in public health. Microbial Genomics. 2023;9(7).
26. Doyle RM, O'Sullivan DM, Aller SD, Bruchmann S, Clark T, Coello Pelegrin A, et al. Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. Microb Genom. 2020;6(2).
27. Marini S, Mora RA, Boucher C, Robertson Noyes N, Prosperi M. Towards routine employment of computational tools for antimicrobial resistance determination via high-throughput sequencing. Briefings in bioinformatics. 2022;23(2):bbac020.
28. Seemann T. ABRicate 2022 [Available from: https://github.com/tseemann/abricate.
29. Bharat A, Petkau A, Avery BP, Chen JC, Folster JP, Carson CA, et al. Correlation between Phenotypic and In Silico Detection of Antimicrobial Resistance in Salmonella enterica in Canada Using Staramr. Microorganisms. 2022;10(2).
30. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial Chemotherapy. 2020;75(12):3491-500.
31. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Scientific reports. 2021;11(1):12728.
32. de Man TJ, Limbago BM. SSTAR, a Stand-Alone Easy-To-Use Antimicrobial Resistance Gene Predictor. mSphere. 2016;1(1).
33. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Research. 2020;48(D1):D517-D25.
34. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 2018;6(1):23.
35. Bonin N, Doster E, Worley H, Pinnell LJ, Bravo JE, Ferm P, et al. MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing. Nucleic Acids Res. 2023;51(D1):D744-d52.
36. NCBI. SRA Tools 2014 [Available from: https://github.com/ncbi/sra-tools.
37. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics (Oxford, England). 2018;34(15):2666-9.
38. Wick RR. Filtlong: quality filtering tool for long reads 2017 [Available from: https://github.com/rrwick/Filtlong.
39. Sun J, Li R, Chen C, Sigwart JD, Kocot KM. Benchmarking Oxford Nanopore read assemblers for high-quality molluscan genomes. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2021;376(1825):20200160.

40. Brejová B, Boršová K, Hodorová V, Čabanová V, Gafurov A, Fričová D, et al. Nanopore sequencing of SARS-CoV-2: Comparison of short and long PCR-tiling amplicon protocols. PloS one. 2021;16(10):e0259277.
41. Lee S, Nguyen LT, Hayes BJ, Ross EM. Prowler: a novel trimming algorithm for Oxford Nanopore sequence data. Bioinformatics (Oxford, England). 2021;37(21):3936-7.
42. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. Functional & integrative genomics. 2015;15(2):141-61.
43. Jiang Y, Jiang Y, Wang S, Zhang Q, Ding X. Optimal sequencing depth design for whole genome re-sequencing in pigs. BMC Bioinformatics. 2019;20(1):556.
44. Bagger FO, Borgwardt L, Jespersen AS, Hansen AR, Bertelsen B, Kodama M, et al. Whole genome sequencing in clinical practice. BMC Med Genomics. 2024;17(1):39.
45. Process V, Ambavaram MMR, Vasantgadkar S, Khanal S, Werner M, Berkeley MA, et al. Optimization of DNA Fragmentation Techniques to Maximize Coverage Uniformity of Clinically Relevant Genes Using Whole Genome Sequencing. Diagnostics [Internet]. 2025; 15(18):[2294 p.].
46. Wick RR. Porechop: Adapter trimmer for Oxford Nanopore reads 2017 [Available from: https://github.com/rrwick/Porechop.
47. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics (Oxford, England). 2016;32(14):2103-10.
48. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics (Oxford, England). 2018;34(18):3094-100.
49. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540-6.
50. Oxford Nanopore Technologies. Medaka 2017 [Available from: https://github.com/nanoporetech/medaka.
51. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England). 2014;30(14):2068-9.
52. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014;58(1):212-20.
53. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res. 2019.
54. Mendes I, Griffiths E, Manuele A, Fornika D, Tausch SH, Le-Viet T, et al. hAMRonization: Enhancing antimicrobial resistance prediction using the PHA4GE AMR detection specification and tooling. bioRxiv. 2024:2024.03.07.583950.
55. Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples). Biometrika. 1965;52:591-611.
56. Levene H. Robust Tests for Equality of Variances. In: Olkin I, editor. Contributions to Probability and Statistics. Palo Alto: Stanford University Press; 1960. p. 278-92.
57. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association. 1952;47(260):583-621.
58. Dunn OJ. Multiple comparisons using rank sums. Technometrics. 1964;6(3):241-52.
59. Oxford Nanopore Technologies. POD5 File Format 2022 [Available from: https://pod5-file-format.readthedocs.io/en/latest/.
60. Bialasiewicz S, Duarte TPS, Nguyen SH, Sukumaran V, Stewart A, Appleton S, et al. Rapid diagnosis of Capnocytophaga canimorsus septic shock in an immunocompetent individual using real-time Nanopore sequencing: a case report. BMC Infectious Diseases. 2019;19(1):660.
61. Cherdtrakulkiat T, Wongsurawat T, Jenjaroenpun P, Sutheeworapong S, Leelawiwat W, Hickey Andrew C, et al. Complete Genome Sequences of Three Neisseria gonorrhoeae Isolates from Thailand with Multidrug Resistance and Multilocus Sequence Type 1903. Microbiology resource announcements. 2020;9(19):10.1128/mra.00198-20.

62. Chopjitt P, Wongsurawat T, Jenjaroenpun P, Boueroy P, Hatrongjit R, Kerdsin A. Complete Genome Sequences of Four Extensively Drug-Resistant Acinetobacter baumannii Isolates from Thailand. Microbiology resource announcements. 2020;9(40):10.1128/mra.00949-20.
63. Conzemius R, Bergman Y, Májek P, Beisken S, Lewis S, Jacobs EB, et al. Automated antimicrobial susceptibility testing and antimicrobial resistance genotyping using Illumina and Oxford Nanopore Technologies sequencing data among Enterobacteriaceae. Front Microbiol. 2022;13:973605.
64. Harris Patrick NA, Bauer Michelle J, Lüftinger L, Beisken S, Forde Brian M, Balch R, et al. Rapid nanopore sequencing and predictive susceptibility testing of positive blood cultures from intensive care patients with sepsis. Microbiology Spectrum. 2024;12(2):e03065-23.
65. Jahnen J, Hanke D, Kadlec K, Schwarz S, Krüger-Haker H. Antimicrobial Resistance in Pasteurella multocida Isolates from Bovine Mastitis Can Be Associated with Multidrug-Resistance-Mediating Integrative and Conjugative Elements (ICEs). Antibiotics (Basel, Switzerland). 2025;14(2).
66. Kostova V, Hanke D, Kaspar H, Fiedler S, Schwarz S, Krüger-Haker H. Macrolide resistance in Mannheimia haemolytica isolates associated with bovine respiratory disease from the German national resistance monitoring program GERM-Vet 2009 to 2020. Front Microbiol. 2024;15:1356208.
67. Leggett RM, Alcon-Giner C, Heavens D, Caim S, Brook TC, Kujawska M, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nature microbiology. 2020;5(3):430-42.
68. Lim A, Naidenov B, Bates H, Willyerd K, Snider T, Couger MB, et al. Nanopore ultra-long read sequencing technology for antimicrobial resistance detection in Mannheimia haemolytica. Journal of microbiological methods. 2019;159:138-47.
69. Smith C, Halse TA, Shea J, Modestil H, Fowler RC, Musser KA, et al. Assessing Nanopore Sequencing for Clinical Diagnostics: a Comparison of Next-Generation Sequencing (NGS) Methods for Mycobacterium tuberculosis. Journal of clinical microbiology. 2020;59(1):e00583-20.
70. Udaondo Z, Jenjaroenpun P, Wongsurawat T, Meyers E, Anderson C, Lopez J, et al. Two Cases of Vancomycin-Resistant Enterococcus faecium Bacteremia With Development of Daptomycin-Resistant Phenotype and its Detection Using Oxford Nanopore Sequencing. Open Forum Infectious Diseases. 2020;7(6):ofaa180.
71. Unemo M, Golparian D, Sánchez-Busó L, Grad Y, Jacobsson S, Ohnishi M, et al. The novel 2016 WHO Neisseria gonorrhoeae reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. J Antimicrob Chemother. 2016;71(11):3096-108.
72. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genom. 2017;3(10):e000131.
73. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Medicine. 2014;6(11):90.
74. Rowe WPM, Winn MD. Indexed variation graphs for efficient and accurate resistome profiling. Bioinformatics (Oxford, England). 2018;34(21):3601-8.
75. Anderson D, Lima L, Le T, Judd L, Wick R, Iqbal Z. Amira: detection of AMR genes directly from long reads using gene-space de Bruijn graphs. bioRxiv. 2025:2025.05.16.654303.
76. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. NAR genomics and bioinformatics. 2021;3(1):lqab019.
77. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. PloS one. 2021;16(10):e0257521.
78. Zidane N, Rodrigues C, Bouchez V, Rethoret-Pasty M, Passet V, Brisse S, et al. Accurate genotyping of three major respiratory bacterial pathogens with ONT R10.4.1 long-read sequencing. Genome Res. 2025;35(8):1758-66.

79. Jauneikaite E, Baker KS, Nunn JG, Midega JT, Hsu LY, Singh SR, et al. Genomics for antimicrobial resistance surveillance to support infection prevention and control in health-care facilities. The Lancet Microbe. 2023;4(12):e1040-e6.
80. Baker KS, Jauneikaite E, Hopkins KL, Lo SW, Sánchez-Busó L, Getino M, et al. Genomics for public health and international surveillance of antimicrobial resistance. The Lancet Microbe. 2023;4(12):e1047-e55.
81. Baker KS, Jauneikaite E, Nunn JG, Midega JT, Atun R, Holt KE, et al. Evidence review and recommendations for the implementation of genomics for antimicrobial resistance surveillance: reports from an international expert group. The Lancet Microbe. 2023;4(12):e1035-e9.
82. Muloi DM, Jauneikaite E, Anjum MF, Essack SY, Singleton DA, Kasudi MR, et al. Exploiting genomics for antimicrobial resistance surveillance at One Health interfaces. The Lancet Microbe. 2023;4(12):e1056-e62.
83. European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters 2025.
84. Oxford Nanopore Technologies. Nanopore sequencing accuracy 2025 [Available from: https://nanoporetech.com/platform/accuracy.
85. Kim BY, Gellert HR, Church SH, Suvorov A, Anderson SS, Barmina O, et al. Single-fly genome assemblies fill major phylogenomic gaps across the Drosophilidae Tree of Life. PLoS Biol. 2024;22(7):e3002697.
86. Wick R. Ryan Wick's bioinformatics blog [Internet]2023. Available from: https://rrwick.github.io/2023/05/05/ont-only-accuracy-with-r10.4.1.html.
87. Backs-Hüsemann D, Reinert J. Embryobildung durch isolierte Einzelzellen aus Gewebekulturen vonDaucus carota. Protoplasma. 1970;70(1):49-60.
88. DTU Center for Genomic Epidemiology. ResFinder Web Server 2025 [Available from: https://genepi.dk/resfinder.
89. Bolkenov B, Lee KY, Atwill ER, Pitesky M, Rickard M, Hung-Fan M, et al. Phenotypic and genotypic characterization of antimicrobial resistance of non-typhoidal Salmonella from retail meat in California. International Journal of Food Microbiology. 2024;421:110785.
90. Do Nascimento V, Day MR, Doumith M, Hopkins KL, Woodford N, Godbole G, et al. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of enteroaggregative Escherichia coli isolated from cases of diarrhoeal disease in England, 2015–16. Journal of Antimicrobial Chemotherapy. 2017;72(12):3288-97.
91. Jones RN. Method preferences and test accuracy of antimicrobial susceptibility testing: updates from the College of Amercian Pathologists Microbiology Surveys Program. Archives of pathology & laboratory medicine. 2001;125(10):1285-9.
92. Santini A, Man A, Voidăzan S. Accuracy of Diagnostic Tests. Journal of critical care medicine (Universitatea de Medicina si Farmacie din Targu-Mures). 2021;7(3):241-8.
93. Jacoby GA. Mechanisms of Resistance to Quinolones. Clinical Infectious Diseases. 2005;41(Supplement_2):S120-S6.