Benchmarking software to predict antibiotic resistance phenotypes in shotgun metagenomes using simulated data

Emily F. Wissel[A], Brooke M. Talbot[B], Bjorn A. Johnson[C], Robert A Petit III[D] , Vicki Hertzberg[A], Anne Dunlop[E], Timothy D. Read[D,F]

Author affiliations and ORCID

A: Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, US
B: Population Biology, Ecology, and Evolution Program, Graduate Division of Biological and Biomedical Science, Emory University, Atlanta, GA, US
C: Cockrell School of Engineering, The University of Texas at Austin, Austin, TX
D: Division of Infectious Diseases, Department of Medicine, School of Medicine, Emory University, Atlanta, GA, US
E: Department of Gynecology & Obstetrics, Emory University School of Medicine
F: Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, US

Author Info
EFW: ewissel@emory.edu https://orcid.org/0000-0003-2275-8456
BMT: https://orcid.org/**0000-0001-5246-7209**
BAJ: https://orcid.org/0000-0002-6460-2444
RAP: https://orcid.org/**0000-0002-1350-9426**
VH: https://orcid.org/**0000-0002-8834-4363**
ALD**: https://orcid.org/0000-0002-5092-8136**
TDR: tread@emory.edu ORCID:0000-0001-8966-9680

Abstract

The use of shotgun metagenomics for AMR detection is appealing because data can be generated from clinical samples with minimal processing. Detecting antimicrobial resistance (AMR) in clinical genomic data is an important epidemiological task, yet a complex bioinformatic process. Many software tools exist to detect AMR genes, but they have mostly been tested in their detection of genotypic resistance in individual bacterial strains. It is important to understand how well these bioinformatic tools detect AMR genes in shotgun metagenomic data.

We developed a software pipeline, hAMRoaster (https://github.com/ewissel/hAMRoaster), for assessing accuracy of prediction of antibiotic resistance phenotypes. For evaluation purposes, we simulated a short read (Illumina) shotgun metagenomics community of eight bacterial pathogens with extensive antibiotic susceptibility testing profiles. We benchmarked nine open source bioinformatics tools for detecting AMR genes that 1) were conda or Docker installable, 2) had been actively maintained, 3) had an open source license, and 4) took FASTA or FASTQ files as input. Several metrics were calculated for each tool including sensitivity, specificity, and F1 at three coverage levels.

This study revealed that tools were highly variable in sensitivity (0.25 - 0.99) and specificity (0.2 - 1) in detection of resistance in our synthetic FASTQ files despite similar databases and methods implemented. Tools performed similarly at all coverage levels (5x, 50x, 100x). Cohen's kappa revealed low agreement across tools.

Importance

Software selection for metagenomic AMR prediction should be driven by the context of the clinical/research questions and tolerance for true and false negative results. As the prediction software and databases are in a state of constant refinement, the approach used here—creating synthetic communities containing taxa and phenotypes of interest along with using hAMRoaster to assess performance of candidate software—offers a template to aid researchers in selecting the most appropriate strategy.

Keywords: antimicrobial resistance, bioinformatics, metagenomics

Tweet: Introducing a new pipeline for comparing results from #AMR tools from @emily_wissel @tdread_emory and others!

hAMRoaster compares detected AMR genes to known resistance, and returns a table with metrics for comparing results across tools.

Introduction

Antibiotic resistant bacterial infections pose a serious threat to public health. Particularly concerning is that the burden of multi-drug resistant pathogens is increasing globally, creating complex clinical scenarios in which there are limited (if any) therapeutic options. In the United States alone, multi-drug resistant infections cost over $4.5 billion annually and kill over 35,000 people each year.[1] Genes that confer antimicrobial resistance (AMR) are increasingly present in commensal members of the human microbiome and are recognized as an important reservoir for conferring pathogen resistance through horizontal gene transfer.[2,3] Detecting AMR potential through non-culture based, high throughput DNA sequencing and bioinformatic approaches is of growing relevance and importance. Two key approaches to mitigating AMR infections are antibiotic stewardship and AMR surveillance. While antibiotic stewardship focuses on prescribing antibiotics appropriately, AMR surveillance focuses on describing AMR genes already present in a community.

AMR surveillance is a key strategy in understanding the threat of AMR. Currently, AMR surveillance typically relies on phenotypic characterization through culture or genotypic characterization through molecular diagnostics based on PCR and hybridization techniques.[4] However, there is a move toward genome-based methods [5] with the Illumina short-read platform being the dominant platform for data generation at the present time.[6] Direct sequencing of clinical samples using shotgun metagenomic approaches is of growing interest for minimizing sample processing and for fully characterizing the commensal members of the microbiome. However, the bioinformatic tools that currently exist to detect AMR have typically not been assessed for their performance on shotgun metagenomic sequence data. Further, as is common with software developed in academic settings, tools are not always maintained or easy to install.

Software managers like conda and docker help to alleviate this problem, however, it can still be difficult for those without a bioinformatics background to understand the state of the tools and select the best one for their needs.

As shotgun metagenomic sequencing is emerging as a powerful tool for detecting and controlling AMR,[7] it is essential to understand how well these tools perform with these data. In addition to testing these tools against a widely available data type, they should be compared against samples with extensive phenotypic resistance (acquired and mutational AMR genes).

This analysis aims to compare a set of existing bioinformatic tools in their ability to accurately identify AMR genes in a community. We describe a software pipeline, hAMRoaster, that provides statistics on accuracy of software when the presence of phenotypes is known. As shotgun metagenomic data is more often used in research and surveillance, and likely soon in clinical diagnostics,[8] we believe this approach of validating tools using synthetic data will be important in selecting the most appropriate software.

Methods

For a schematic overview of the methods, see **Supplementary Figure One.**

Development of a software pipeline, hAMRoaster, to assess results of antibiotic resistance prediction

hAMRoaster was written as a Python script to take three inputs: a) the text output of AMR prediction run tool on a FASTQ or FASTA test file, such as a text file processed through hAMRonization,[9] b) a list of known phenotypes associated with the test file and c) (optional) a tab formatted table which matches antibiotic drugs with their drug class. If option c) is not specified a default table is used. The outputs of the program are a set of performance metrics that include sensitivity and specificity. A conda installable version of the software was deposited in the Bioconda[10] database. The Github site for the software is https://github.com/ewissel/hAMRoaster.

hAMRoaster requires, as input, a formatted results table of runs by AMR detection tools. This table is identical to that produced by the hAMRonization[9] software**.** hAMRonization is conda installable and can compile the outputs of many AMR tools into a unified format. shortBRED[11] and fARGene[12] are not included in hAMRonization at the time of analysis, so hAMRoaster can take the path to the raw output for these tools and partially match it to the hAMRonization output.

hAMRoaster requires an input to the "known" phenotypic resistance in the mock community (--AMR_key flag of hAMRoaster), such as a result of susceptibility testing tables that are available from NCBI Biosamples. Antibiotics in the table of known resistances are matched to their respective drug classes. Results classified as "susceptible" or "intermediate" in susceptibility testing are filtered out so only resistant instances are considered. In cases where susceptibility testing occurred with two or more agents, each agent is considered independently (e.g. resistance to "amoxicillin-tetracycline" was treated as resistance to "amoxicillin" and "tetracycline" independently). Each identified AMR gene is labeled with its corresponding drug class for comparison. In instances where a gene confers resistance to multiple drug classes, the detected gene is split into multiple rows so that each conferred resistance can be independently compared to what is known from the susceptibility testing. Gene to drug class linkage is verified using the CARD database[13] when applicable. Any genes corresponding to 'unknown' or 'other' drug classes (including hypothetical resistance genes) are excluded from further analysis. Genes that confer resistance to an antibiotic that was only effective in combination with another drug (e.g. clavulanic acid in amoxicillin-clavulanic acid) are classified as 'Other' and excluded from analysis.

A detected AMR gene is labeled as a true positive by hAMRoaster if the drug class matched to an AMR gene corresponds to a drug class represented in the mock community. Similarly, a false positive is coded as a drug class that is called by the software, but tested as susceptible in the mock community (--AMR key parameter). Observed AMR genes are labeled "Unknown" if the corresponding drug class is not tested in the mock community and not included in the AMR key file. Once true/false positives and true/false negatives are determined per tool, hAMRoaster calculates sensitivity, specificity, precision, accuracy, recall, and percent unknown.

Creation of a synthetic mock communities of antibiotic resistance bacteria

Bacterial members of the base mock community were chosen from NCBI's BioSample Database[14] and met the following criteria: (1) the strain had extensive antibiotic susceptibility testing data using CLSI or EUCAST testing standards as part of the public NCBI BioSample record; (2) the strain was isolated from human tissue; (3) the strain was the cause of a clinical infection; (4) the FASTA was available to download from NCBI BioSample Database.[14] Eight bacteria, each representing a different species, with overlapping resistance to 43 antibiotics across 18 drug classes, were selected for the mock community (**Table 1**). The included taxa were *Acinetobacter baumannii* MRSN489669, *Citrobacter freundii* MRSN12115, *Enterobacter cloacae* 174, *Escherichia coli* 222, *Klebsiella pneumoniae* CCUG 70742, *Pseudomonas aeruginosa* CCUG 70744, *Neisseria gonorrhoeae* SW0011, and *Staphylococcus aureus* LAC (Table 1).

Paired-end FASTQs were simulated by ART [15] using default parameters for HiSeq 2500 at three levels of average sequence coverage (5x, 50x, and 100x) and are available on FigShare (https://figshare.com/account/home#/projects/125974). Simulated FASTQs were subsequently concatenated to resemble shotgun metagenomics reads, and metaSPAdes[16] was used to create assembled contigs. The FASTQs were simulated with approximately equal numbers of reads of each genome.

Running antibiotic prediction software on mock communities

All tools for AMR prediction were run on the mock community at all coverage levels using default settings for either simulated FASTQ or assembled contigs. When both options were available, assembled contigs were run.

Statistical Analysis

Data were analyzed in Python v3.7.7 and plotted in R v4.0.4. In initial runs we found that some tools provided results with a very high number of observed AMR genes because of multiple overlapping matches on the same gene. Because of this, we condensed the results so that the first observed gene is included in the dataset and subsequent genes that start before the observed end of that gene were not included. Unweighted Cohen's kappa was calculated for each pairwise combination of tools to test agreement between tools.

Data Availability

All data and code is available on the hAMRoaster GitHub repository (https://github.com/ewissel/hAMRoaster) and figshare (for large FASTQ files; https://figshare.com/account/home#/projects/125974)

Results

Selection of nine open source, conda-installable tools for detection of antibiotic resistance phenotypes

To identify tools for antibiotic resistance prediction, we used a multi-headed search strategy. We searched PubMed using terms “AMR”, “antibiotic resistance genes”, “bioinformatics”, and “antimicrobial resistance”. We also searched GitHub using the same set of terms. Once an initial list of tools was compiled, we performed a second PubMed literature review including the search terms from above plus the names of the tools (“tool 1” OR “tool 2”). We also used Twitter to ask the research community what bioinformatic tools they use to identify AMR (link available in supplementary materials). These searches identified 16 potential tools to identify AMR genes (**Table 2**). The search for tools concluded on March 1, 2021.

In order for an identified tool to be considered eligible for comparison, it had to meet the following criteria: (1) be conda or Docker installable; (2) have source code publicly available in a data repository and be actively maintained (defined as tool updates or GitHub responses within the last year); (3) have an open source license; and (4) take FASTQs or FASTAs as input files. Nine tools met the criteria to be included in this analysis: ABRIcate [17], fARGene [18] ResFinder [19], shortBRED[11], RGI [20], AMRFinderPlus [21], starAMR [22] , sraX [23], and deepARG [24]. PointFinder also qualified[25], but was a subtool of ResFinder and only identified mutational resistance for some organisms, so it was excluded from analysis. The code used to install and run all tools is available on the hAMRoaster GitHub.

Identified tools fell into two groups - those that aligned reads to a database, and those that compared reads against some model of AMR (Table 2).

ABRIcate

ABRIcate v.1.0.1 took contig FASTA files as inputs and compared reads against a large database created by compiling several existing database, including NCBI AMRFinder Plus,[21] CARD,[20] ResFinder,[19] ARG-ANNOT,[26] MEGARES,[27] EcOH,[28] PlasmidFinder,[29] VFDB,[30] and Ecoli_VF.[31] ABRIcate reported on acquired AMR genes and not mutational resistance.

shortBRED

shortBRED[11] v0.9.3 used a set of marker genes to search metagenomic data for protein families of interest. The bioBakery[32] team published an AMR gene marker database built from 849 AR protein families derived from the ARDB[33] v1.1 and independent curation alongside shortBRED, which is used in this study.

fARGene

fARGene[12,18] v.0.1 uses Hidden Markov Models to detect AMR genes from short metagenomic data or long read data. This was different from most other tools which compare the reads directly. fARGene has three pre-built models for detecting resistance to quinolone, tetracycline, and beta lactamases, which were tested in this study.

RGI

RGI[20] v5.1.1 used protein homology and SNP models to predict ‘resistomes’. It used CARD’s protein homolog models as a database. RGI predicts open reading frames (ORFs) using Prodigal,[34] detects homologs with DIAMOND,[35] and matches to CARD’s database and model cut off values.

ResFinder

ResFinder[19] v4.0 was available both as a web-based application or the command line. We used ResFinder 4 in this study, which was specifically designed for detecting genotypic resistance in phenotypically resistant samples. ResFinder aligned reads directly to its own curated database without need for assembly.

deepARG

deepARG[24] v.2.0 used a supervised deep learning based approach for antibiotic resistance gene annotation of metagenomic sequences. It combines three databases—CARD, ARDB, and UNIPROT—and categorizes them into resistance categories.

sraX

sraX[23] v.1.5 was built as a one step tool; in a single command, sraX downloads a database and aligns contigs to this database with DIAMOND[35]. By default, sraX uses CARD, though other options can be specified. As we use default settings for all tools, only CARD was used in this study for sraX.

starAMR

starAMR[22,36] v.0.7.2 uses BLAST+[37] to compare contigs against a combined database with data from ResFinder, PointFinder, and PlasmidFinder.

AMR Finder Plus

AMR Finder Plus[21] v.3.9.3 uses BLASTX[38] translated searches and hierarchical tree of gene families to detect AMR genes. The database was derived from the Pathogen Detection Reference Gene Catalog[39] and was compiled as part of the National Database of Antibiotic Resistant Organisms (NDARO).

Performance of selected tools on a mock bacterial community containing 43 laboratory confirmed AMR phenotypes

Each software tool was run against a synthetic mock community of 8 bacteria at three coverage levels that expressed 43 antibiotic resistance phenotypes. To assess sensitivity and specificity, we developed a new software pipeline called hAMRoaster (Harmonized AMR Output compAriSon Tool 'ER').

Range of phenotypes detected

Overall, the number of AMR genes detected across all tools ranged from 13 to over 700 at 100x coverage (**Table 3).** For some tools, genes detected did not match to a tested phenotype in the mock community, so the prediction fell into the "unknown" category. Among the tools tested, AMR Finder Plus had the highest degree of unclassifiable/unknown results (observed AMR gene not testing in the mock community)**.** An overview of these results are available in **Figure One**.

Sensitivity and Specificity

The highest sensitivity for phenotype detection ranged from >0.99 (RGI) to 0.23 (sraX) at the lowest coverage levels (**Fig. 2**). In general, coverage did not greatly affect sensitivity, with the exception of sraX, which increased to 0.53 at the highest level. fARGene and deepARG had a high sensitivity value (>0.90) at all coverage levels. RGI, deepARG, and fARGene are all tools that compare reads to a model of AMR instead of aligning reads directly to a database, indicating that this method may be appropriate when high sensitivity values are preferred. As a note, in this dataset, there were only 2 possible true negatives because only two drug classes were always susceptible to antibiotics in those two drug classes when tested (nitrofuran and polypeptide).

When all software predictions were combined we achieved the 0.99 sensitivity across the coverage (**Supplementary Table 1**). However this came at the cost of low specificity 0.11 .

Specificity in this study is artificially low for most tools because the number of possible true negatives is low (only two). Therefore we did not assess this metric.

Condensing Results

All tools provide results in which the detected AMR genes are overlapping, where one gene starts between the start and stop codon of another. If we remove overlapping genes so that only the first detected gene was included, and all genes that started before its stop codon were removed, the counts for all tools decrease (**Table 4**). However, this process does not necessarily improve metrics or counts, and it is unclear that such a tactic is useful for real life uses as there is no simple way to determine which detected AMR genes to include and which should be filtered out.

Concordance between tools

An analysis of the agreement between tools of detected AMR genes within drug classes revealed that overall, there was low agreement (<0.50) between tools at all coverage levels (**Table 6**).

Discussion

Development of a framework for assessing AMR prediction software performance using synthetic data

There is a considerable research effort to develop new software for predicting AMR using DNA sequence alone. In this dynamic environment, there is a need for researchers and epidemiologists to understand the relative performance of open source software tools within the types of sample they may encounter. While some tools currently exist for compiling the results of several AMR tools together (hAMRonizer and chARMedDb[40]), this study was motivated by the lack of an open-source pipeline for comparing the results once compiled. hAMRoaster was built so that several metrics can easily be compared across tools.

The central challenge in developing this software was to compare detected AMR genes to resistance phenotypes. Detected AMR genes needed to be classified by their corresponding drug class(es) so they could be matched to the known phenotypically resistant drug classes. One hurdle in this translation is that tools use different databases, and some databases classify genes differently. For example, shortBRED classifies gene families, while CARD classifies specific genes. While this analysis checked the drug classification via the DNA/Protein Accession value in CARD, only around 300 of the >1,000 genes detected could directly map to genes in CARD by accession value. The hAMRonization tool overcomes this challenge by providing a drug class column and filling in the values from ChEBI ontology[41] when possible. The hAMRoaster strategy is to assign a CARD drug class value to every detected AMR gene first by accession number, then by gene name. If neither of these methods assign a drug class for an AMR gene, then the drug class provided by hAMRonization is used. Another challenge in converting detected AMR genes to drug classes is that some drugs are only administered in combination, for example clavulanic acid with amoxicillin. For these instances, resistance to the drug only used in combination (e.g. clavulanic acid) is treated as an “other” drug class and excluded from analysis. In these cases, we used the experience of practicing clinicians to identify combination antibiotics.

The analysis presented here used synthetic data to compare tool performance. Synthetic data has the benefit of allowing controlled input with known ground truth. Therefore users can focus on the types of organisms and phenotypes they need to to detect in their own datasets, perform experiments with real samples, and manipulate a range of factors such as relative abundance and sequencing error. The NCBI BioSample repository (used in this study) is an invaluable resource for creating such datasets as it contains many samples with AMR phenotypes determined by international standards. Researchers could also sequence and phenotype culturable organisms in their own laboratories to provide testing standards to evaluate software. Here, we exclusively examined synthetic short read Illumina data, but this analysis strategy could be adapted to understand the effect of using data generated on long read technologies such as the Pacific Bioscience and Oxford Nanopore platforms.

Overall trends in performance and reasons for variability between tools

Tools used one of two basic strategies, either aligning reads to a database of AMR genes or using a more complex model of sequenced-based AMR detection (**Table 2**). The methods appear to lead to the different AMR genes detected across tools, as demonstrated in **Figure 1** and summarized in **Table 3.**

We found the sensitivity of almost all tools to be very good (>0.80), with the exception of sraX, which had a proportionally high number of false negatives compared to true positives. All tools except shortBRED and starAMR detected a large number of genes that were not associated with a lab-determined phenotype in our mock community. This is a feature of the approach of limiting focus to a specific set of phenotypes in the testing process. In practice, researchers and epidemiologists may be only interested in a narrow range of AMR phenotypes.

hAMRoaster calculates specificity, precision, accuracy, recall , and F1 (**Table 3)**. However, all of these measures are dependent on false positives and/or true negatives in their calculations. As these values are inherently low in our mock community due to the robust resistance profile, these metrics are not particularly informative for understanding how well these tools detect resistance in this phenotypically resistant sample. Similarly, we calculated all effective metrics when the results of all tools are combined. While sensitivity in the combined data was very high (>0.99), there was a very high number of overall detected AMR genes, including overlapping results between genes, thus,.it would be difficult for researchers to meaningfully use this type of result to understand the AMR profile. We calculated Cohen's kappa to capture the agreement at the drug class level between AMR tools to see if all AMR tools detected resistance to the same drug classes. We found that agreement was surprisingly low across all tools (**Table 6**), indicating that some tools may be better suited for detecting different types of resistance. As such, hAMRoaster provides a table with the number of genes detected per drug class for each tool.

Finally, this research supports the need for the further development of software tools for the detection of AMR genes in the human microbiome. It is increasingly recognized that the confined location and genetic diversity of this microbial population provides ideal conditions for genetic exchange among residential microbes and between residential and transient, including pathogenic microbes. Notably, rates of horizontal gene transfer among bacteria in the human microbiome (especially the gastrointestinal tract) are estimated to be many times higher than among bacteria in other diverse ecosystems, such as soil.[42] Refined tools appropriate for use in shotgun metagenomic data will be important for tracking the spread of AMR genes from diverse environmental sources to the human microbiome and across sites in the human body and understanding whether AMR genes are derived from vertical inheritance or via horizontal gene transfer, for example.

In conclusion, this study compared bioinformatics tools for detecting AMR genes in a simulated short read metagenomic sample at three coverage levels at one time point. While tools use slightly different methods and databases, these tools overall had high sensitivity for detection of AMR genes. Moreover agreement between tools was low, indicating the importance of tool selection. In our test set we found starAMR had the highest sensitivity value with fewer than 20% unknown detected genes at all coverage levels. We advocate that researchers should test these software tools using pipelines such as hAMRoaster with a synthetic community that highlights the resistance profiles and sample of interest. In particular, this assessment of performance of available tools should take place before the commencement of the study as the set of tools for detecting AMR genes are actively maintained and undergoing further improvements.

Acknowledgements

We thank Jon Moller for helping to create the hAMRoaster name.

Funding

EFW is supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1937971.

Author Contributions

EFW and TDR conceptualized and planned the initial project. TDF, VH, AD, and RAP provided ongoing support in study design and analysis. EFW and BMT processed the data. EFW, BAJ, and BMT analyzed the data. EFW, BAJ, and RAP created the hAMRoaster software. EFW and TDR drafted the initial manuscript. All authors reviewed the final manuscript.

References

(1) Centers for Disease Control and Prevention (U.S.). *Antibiotic Resistance Threats in the United States, 2019*; Centers for Disease Control and Prevention (U.S.), 2019. https://doi.org/10.15620/cdc:82532.

(2) Nji, E.; Kazibwe, J.; Hambridge, T.; Joko, C. A.; Larbi, A. A.; Damptey, L. A. O.; Nkansa-Gyamfi, N. A.; Stålsby Lundborg, C.; Lien, L. T. Q. High Prevalence of Antibiotic Resistance in Commensal Escherichia Coli from Healthy Human Sources in Community Settings. *Sci. Rep.* **2021**, *11* (1), 3372. https://doi.org/10.1038/s41598-021-82693-4.

(3) Brinkac, L.; Voorhies, A.; Gomez, A.; Nelson, K. E. The Threat of Antimicrobial Resistance on the Human Microbiome. *Microb. Ecol.* **2017**, *74* (4), 1001–1008. https://doi.org/10.1007/s00248-017-0985-z.

(4) Anjum, M. F.; Zankari, E.; Hasman, H. Molecular Methods for Detection of Antimicrobial Resistance. *Microbiol. Spectr.* **2017**, *5* (6). https://doi.org/10.1128/microbiolspec.ARBA-0011-2017.

(5) Nutrition, C. for F. S. and A. GenomeTrakr Network. *FDA* **2021**.

(6) Porter, T. M.; Hajibabaei, M. Over 2.5 Million COI Sequences in GenBank and Growing. *PLOS ONE* **2018**, *13* (9), e0200177. https://doi.org/10.1371/journal.pone.0200177.

(7) Oniciuc, E. A.; Likotrafiti, E.; Alvarez-Molina, A.; Prieto, M.; Santos, J. A.; Alvarez-Ordóñez, A. The Present and Future of Whole Genome Sequencing (WGS) and Whole Metagenome Sequencing (WMS) for Surveillance of Antimicrobial Resistant Microorganisms and Antimicrobial Resistance Genes across the Food Chain. *Genes* **2018**, *9* (5), 268. https://doi.org/10.3390/genes9050268.

(8) Dulanto Chiang, A.; Dekker, J. P. From the Pipeline to the Bedside: Advances and Challenges in Clinical Metagenomics. *J. Infect. Dis.* **2020**, *221* (Supplement_3), S331–S340. https://doi.org/10.1093/infdis/jiz151.

(9) Issues · pha4ge/hAMRonization https://github.com/pha4ge/hAMRonization (accessed 2021 -10 -12).

(10) Grüning, B.; Dale, R.; Sjödin, A.; Chapman, B. A.; Rowe, J.; Tomkins-Tinch, C. H.; Valieris, R.; Köster, J. Bioconda: Sustainable and Comprehensive Software Distribution for the Life Sciences. *Nat. Methods* **2018**, *15* (7), 475–476. https://doi.org/10.1038/s41592-018-0046-7.

(11) Kaminski, J.; Gibson, M. K.; Franzosa, E. A.; Segata, N.; Dantas, G.; Huttenhower, C. High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. *PLOS Comput. Biol.* **2015**, *11* (12), e1004557. https://doi.org/10.1371/journal.pcbi.1004557.

(12) fannyhb. *FARGene*; 2021.

(13) Alcock, B. P.; Raphenya, A. R.; Lau, T. T. Y.; Tsang, K. K.; Bouchard, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.-L. V.; Cheng, A. A.; Liu, S.; Min, S. Y.; Miroshnichenko, A.; Tran, H.-K.; Werfalli, R. E.; Nasir, J. A.; Oloni, M.; Speicher, D. J.; Florescu, A.; Singh, B.; Faltyn, M.; Hernandez-Koutoucheva, A.; Sharma, A. N.; Bordeleau, E.; Pawlowski, A. C.; Zubyk, H. L.; Dooley, D.; Griffiths, E.; Maguire, F.; Winsor, G. L.; Beiko, R. G.; Brinkman, F. S. L.; Hsiao, W. W. L.; Domselaar, G. V.; McArthur, A. G. CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **2020**, *48* (D1), D517–D525.

https://doi.org/10.1093/nar/gkz935.

(14) Barrett, T.; Clark, K.; Gevorgyan, R.; Gorelenkov, V.; Gribov, E.; Karsch-Mizrachi, I.; Kimelman, M.; Pruitt, K. D.; Resenchuk, S.; Tatusova, T.; Yaschenko, E.; Ostell, J. BioProject and BioSample Databases at NCBI: Facilitating Capture and Organization of Metadata. *Nucleic Acids Res.* **2012**, *40* (D1), D57–D63. https://doi.org/10.1093/nar/gkr1163.

(15) Huang, W.; Li, L.; Myers, J. R.; Marth, G. T. ART: A next-Generation Sequencing Read Simulator. *Bioinforma. Oxf. Engl.* **2012**, *28* (4), 593–594. https://doi.org/10.1093/bioinformatics/btr708.

(16) Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P. A. MetaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res.* **2017**, *27* (5), 824–834. https://doi.org/10.1101/gr.213959.116.

(17) Seemann, T. *ABRicate*; 2021.

(18) Berglund, F.; Österlund, T.; Boulund, F.; Marathe, N. P.; Larsson, D. G. J.; Kristiansson, E. Identification and Reconstruction of Novel Antibiotic Resistance Genes from Metagenomes. *Microbiome* **2019**, *7* (1), 52. https://doi.org/10.1186/s40168-019-0670-1.

(19) Bortolaia, V.; Kaas, R. S.; Ruppe, E.; Roberts, M. C.; Schwarz, S.; Cattoir, V.; Philippon, A.; Allesoe, R. L.; Rebelo, A. R.; Florensa, A. F.; Fagelhauer, L.; Chakraborty, T.; Neumann, B.; Werner, G.; Bender, J. K.; Stingl, K.; Nguyen, M.; Coppens, J.; Xavier, B. B.; Malhotra-Kumar, S.; Westh, H.; Pinholt, M.; Anjum, M. F.; Duggett, N. A.; Kempf, I.; Nykäsenoja, S.; Olkkola, S.; Wieczorek, K.; Amaro, A.; Clemente, L.; Mossong, J.; Losch, S.; Ragimbeau, C.; Lund, O.; Aarestrup, F. M. ResFinder 4.0 for Predictions of Phenotypes from Genotypes. *J. Antimicrob. Chemother.* **2020**, *75* (12), 3491–3500. https://doi.org/10.1093/jac/dkaa345.

(20) Alcock, B. P.; Raphenya, A. R.; Lau, T. T. Y.; Tsang, K. K.; Bouchard, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.-L. V.; Cheng, A. A.; Liu, S.; Min, S. Y.; Miroshnichenko, A.; Tran, H.-K.; Werfalli, R. E.; Nasir, J. A.; Oloni, M.; Speicher, D. J.; Florescu, A.; Singh, B.; Faltyn, M.; Hernandez, A., Koutoucheva; Sharma, A. N.; Bordeleau, E.; Pawlowski, A. C.; Zubyk, H. L.; Dooley, D.; Griffiths, E.; Maguire, F.; Winsor, G. L.; Beiko, R. G.; Brinkman, F. S. L.; Hsiao, W. W. L.; Domselaar, G. V.; McArthur, A. G. *CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database*; 2020. https://doi.org/10.1093/nar/gkz935.

(21) *NCBI Antimicrobial Resistance Gene Finder (AMRFinderPlus)*; NCBI - National Center for Biotechnology Information/NLM/NIH, 2021.

(22) *Staramr*; National Microbiology Laboratory, 2021.

(23) Panunzi, L. G. SraX: A Novel Comprehensive Resistome Analysis Tool. *Front. Microbiol.* **2020**, *11*, 52. https://doi.org/10.3389/fmicb.2020.00052.

(24) Arango-Argoty, G.; Garner, E.; Pruden, A.; Heath, L. S.; Vikesland, P.; Zhang, L. DeepARG: A Deep Learning Approach for Predicting Antibiotic Resistance Genes from Metagenomic Data. *Microbiome* **2018**, *6* (1), 23. https://doi.org/10.1186/s40168-018-0401-z.

(25) PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens | Journal of Antimicrobial Chemotherapy | Oxford Academic https://academic.oup.com/jac/article/72/10/2764/3979530?login=true (accessed 2021 -10 -12).

(26) Gupta, S. K.; Padmanabhan, B. R.; Diene, S. M.; Lopez-Rojas, R.; Kempf, M.; Landraud, L.; Rolain, J.-M. ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrob. Agents Chemother.* **2014**, *58* (1), 212–220. https://doi.org/10.1128/AAC.01310-13.

(27) Doster, E.; Lakin, S. M.; Dean, C. J.; Wolfe, C.; Young, J. G.; Boucher, C.; Belk, K. E.; Noyes, N. R.; Morley, P. S. MEGARes 2.0: A Database for Classification of Antimicrobial Drug, Biocide and Metal Resistance Determinants in Metagenomic Sequence Data. *Nucleic Acids Res.* **2020**, *48* (D1), D561–D569. https://doi.org/10.1093/nar/gkz1010.

(28) Ingle, D. J.; Valcanis, M.; Kuzevski, A.; Tauschek, M.; Inouye, M.; Stinear, T.; Levine, M. M.; Robins-Browne, R. M.; Holt, K. E. In Silico Serotyping of E. Coli from Short Read Data Identifies Limited Novel O-Loci but Extensive Diversity of O:H Serotype Combinations within and between Pathogenic Lineages. *Microb. Genomics* **2016**, *2* (7), e000064. https://doi.org/10.1099/mgen.0.000064.

(29) Carattoli, A.; Hasman, H. PlasmidFinder and In Silico PMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). *Methods Mol. Biol. Clifton NJ* **2020**, *2075*, 285–294. https://doi.org/10.1007/978-1-4939-9877-7_20.

(30) Chen, L.; Zheng, D.; Liu, B.; Yang, J.; Jin, Q. VFDB 2016: Hierarchical and Refined Dataset for Big Data Analysis—10 Years On. *Nucleic Acids Res.* **2016**, *44* (D1), D694–D697. https://doi.org/10.1093/nar/gkv1239.

(31) *Escherichia Coli Virulence Factors*; National Microbiology Laboratory, 2021.

(32) McIver, L. J.; Abu-Ali, G.; Franzosa, E. A.; Schwager, R.; Morgan, X. C.; Waldron, L.; Segata, N.; Huttenhower, C. BioBakery: A Meta'omic Analysis Environment. *Bioinformatics* **2018**, *34* (7), 1235–1237. https://doi.org/10.1093/bioinformatics/btx754.

(33) Liu, B.; Pop, M. ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **2009**, *37* (Database), D443–D447. https://doi.org/10.1093/nar/gkn656.

(34) Hyatt, D.; Chen, G.-L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* **2010**, *11* (1), 119. https://doi.org/10.1186/1471-2105-11-119.

(35) Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND. *Nat. Methods* **2021**, *18* (4), 366–368. https://doi.org/10.1038/s41592-021-01101-x.

(36) Zankari, E.; Hasman, H.; Cosentino, S.; Vestergaard, M.; Rasmussen, S.; Lund, O.; Aarestrup, F. M.; Larsen, M. V. Identification of Acquired Antimicrobial Resistance Genes. *J. Antimicrob. Chemother.* **2012**, *67* (11), 2640–2644. https://doi.org/10.1093/jac/dks261.

(37) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: Architecture and Applications. *BMC Bioinformatics* **2009**, *10* (1), 421. https://doi.org/10.1186/1471-2105-10-421.

(38) McGinnis, S.; Madden, T. L. BLAST: At the Core of a Powerful and Diverse Set of Sequence Analysis Tools. *Nucleic Acids Res.* **2004**, *32* (suppl_2), W20–W25. https://doi.org/10.1093/nar/gkh435.

(39) Reference Gene Catalog - Pathogen Detection - NCBI https://www.ncbi.nlm.nih.gov/pathogens/refgene/# (accessed 2021 -09 -13).

(40) Anthony Underwood / chAMReDb https://gitlab.com/antunderwood/chamredb (accessed 2021 -10 -12).

(41) Hastings, J.; Owen, G.; Dekker, A.; Ennis, M.; Kale, N.; Muthukrishnan, V.; Turner, S.; Swainston, N.; Mendes, P.; Steinbeck, C. ChEBI in 2016: Improved Services and an Expanding Collection of Metabolites. *Nucleic Acids Res.* **2016**, *44* (D1), D1214–D1219. https://doi.org/10.1093/nar/gkv1031.

(42) Huttenhower, C.; Gevers, D.; Knight, R.; Abubucker, S.; Badger, J. H.; Chinwalla, A. T.; Creasy, H. H.; Earl, A. M.; FitzGerald, M. G.; Fulton, R. S.; Giglio, M. G.; Hallsworth-Pepin, K.; Lobos, E. A.; Madupu, R.; Magrini, V.; Martin, J. C.; Mitreva, M.; Muzny, D. M.; Sodergren, E. J.; Versalovic, J.; Wollam, A. M.; Worley, K. C.; Wortman, J. R.; Young, S. K.; Zeng, Q.; Aagaard, K. M.; Abolude, O. O.; Allen-Vercoe, E.; Alm, E. J.; Alvarado, L.; Andersen, G. L.; Anderson, S.; Appelbaum, E.; Arachchi, H. M.; Armitage, G.; Arze, C. A.; Ayvaz, T.; Baker, C. C.; Begg, L.; Belachew, T.; Bhonagiri, V.; Bihan, M.; Blaser, M. J.; Bloom, T.; Bonazzi, V.; Paul Brooks, J.; Buck, G. A.; Buhay, C. J.; Busam, D. A.; Campbell, J. L.; Canon, S. R.; Cantarel, B. L.; Chain, P. S. G.; Chen, I.-M. A.; Chen, L.; Chhibba, S.; Chu, K.; Ciulla, D. M.; Clemente, J. C.; Clifton, S. W.; Conlan, S.; Crabtree, J.; Cutting, M. A.; Davidovics, N. J.; Davis, C. C.; DeSantis, T. Z.; Deal, C.; Delehaunty, K. D.; Dewhirst, F. E.; Deych, E.; Ding, Y.; Dooling, D. J.; Dugan, S. P.; Michael Dunne, W.; Scott Durkin, A.; Edgar, R. C.; Erlich, R. L.; Farmer, C. N.; Farrell, R. M.; Faust, K.; Feldgarden, M.; Felix, V. M.; Fisher, S.; Fodor, A. A.; Forney, L. J.; Foster, L.; Di Francesco, V.; Friedman, J.; Friedrich, D. C.; Fronick, C. C.; Fulton, L. L.; Gao, H.; Garcia, N.; Giannoukos, G.; Giblin, C.; Giovanni, M. Y.; Goldberg, J. M.; Goll, J.; Gonzalez, A.; Griggs, A.; Gujja, S.; Kinder Haake, S.; Haas, B. J.; Hamilton, H. A.; Harris, E. L.; Hepburn, T. A.; Herter, B.; Hoffmann, D. E.; Holder, M. E.; Howarth, C.; Huang, K. H.; Huse, S. M.; Izard, J.; Jansson, J. K.; Jiang, H.; Jordan, C.; Joshi, V.; Katancik, J. A.; Keitel, W. A.; Kelley, S. T.; Kells, C.; King, N. B.; Knights, D.; Kong, H. H.; Koren, O.; Koren, S.; Kota, K. C.; Kovar, C. L.; Kyrpides, N. C.; La Rosa, P. S.; Lee, S. L.; Lemon, K. P.; Lennon, N.; Lewis, C. M.; Lewis, L.; Ley, R. E.; Li, K.; Liolios, K.; Liu, B.; Liu, Y.; Lo, C.-C.; Lozupone, C. A.; Dwayne Lunsford, R.; Madden, T.; Mahurkar, A. A.; Mannon, P. J.; Mardis, E. R.; Markowitz, V. M.; Mavromatis, K.; McCorrison, J. M.; McDonald, D.; McEwen, J.; McGuire, A. L.; McInnes, P.; Mehta, T.; Mihindukulasuriya, K. A.; Miller, J. R.; Minx, P. J.; Newsham, I.; Nusbaum, C.; O’Laughlin, M.; Orvis, J.; Pagani, I.; Palaniappan, K.; Patel, S. M.; Pearson, M.; Peterson, J.; Podar, M.; Pohl, C.; Pollard, K. S.; Pop, M.; Priest, M. E.; Proctor, L. M.; Qin, X.; Raes, J.; Ravel, J.; Reid, J. G.; Rho, M.; Rhodes, R.; Riehle, K. P.; Rivera, M. C.; Rodriguez-Mueller, B.; Rogers, Y.-H.; Ross, M. C.; Russ, C.; Sanka, R. K.; Sankar, P.; Fah Sathirapongsasuti, J.; Schloss, J. A.; Schloss, P. D.; Schmidt, T. M.; Scholz, M.; Schriml, L.; Schubert, A. M.; Segata, N.; Segre, J. A.; Shannon, W. D.; Sharp, R. R.; Sharpton, T. J.; Shenoy, N.; Sheth, N. U.; Simone, G. A.; Singh, I.; Smillie, C. S.; Sobel, J. D.; Sommer, D. D.; Spicer, P.; Sutton, G. G.; Sykes, S. M.; Tabbaa, D. G.; Thiagarajan, M.; Tomlinson, C. M.; Torralba, M.; Treangen, T. J.; Truty, R. M.; Vishnivetskaya, T. A.; Walker, J.; Wang, L.; Wang, Z.; Ward, D. V.; Warren, W.; Watson, M. A.; Wellington, C.; Wetterstrand, K. A.; White, J. R.; Wilczek-Boney, K.; Wu, Y.; Wylie, K. M.; Wylie, T.; Yandava, C.; Ye, L.; Ye, Y.; Yooseph, S.; Youmans, B. P.; Zhang, L.; Zhou, Y.; Zhu, Y.; Zoloth, L.; Zucker, J. D.; Birren, B. W.; Gibbs, R. A.; Highlander, S. K.; Methé, B. A.; Nelson, K. E.; Petrosino, J. F.; Weinstock, G. M.; Wilson, R. K.; White, O.; The Human Microbiome Project Consortium. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* **2012**, *486* (7402), 207–214. https://doi.org/10.1038/nature11234.

Figure 1: Antimicrobial Resistance (AMR) Genes Detected By Software Tools by Drug Class

AMR Genes detected by each tool across coverage levels, grouped into drug class to which the genes confer resistance with the color coding indicating whether the detection was true positive (green), false positive (purple) or unknown (yellow). Clear spaces in the plot indicate that AMR genes were not detected for the drug class on the x-axis by the tool on the y-axis.

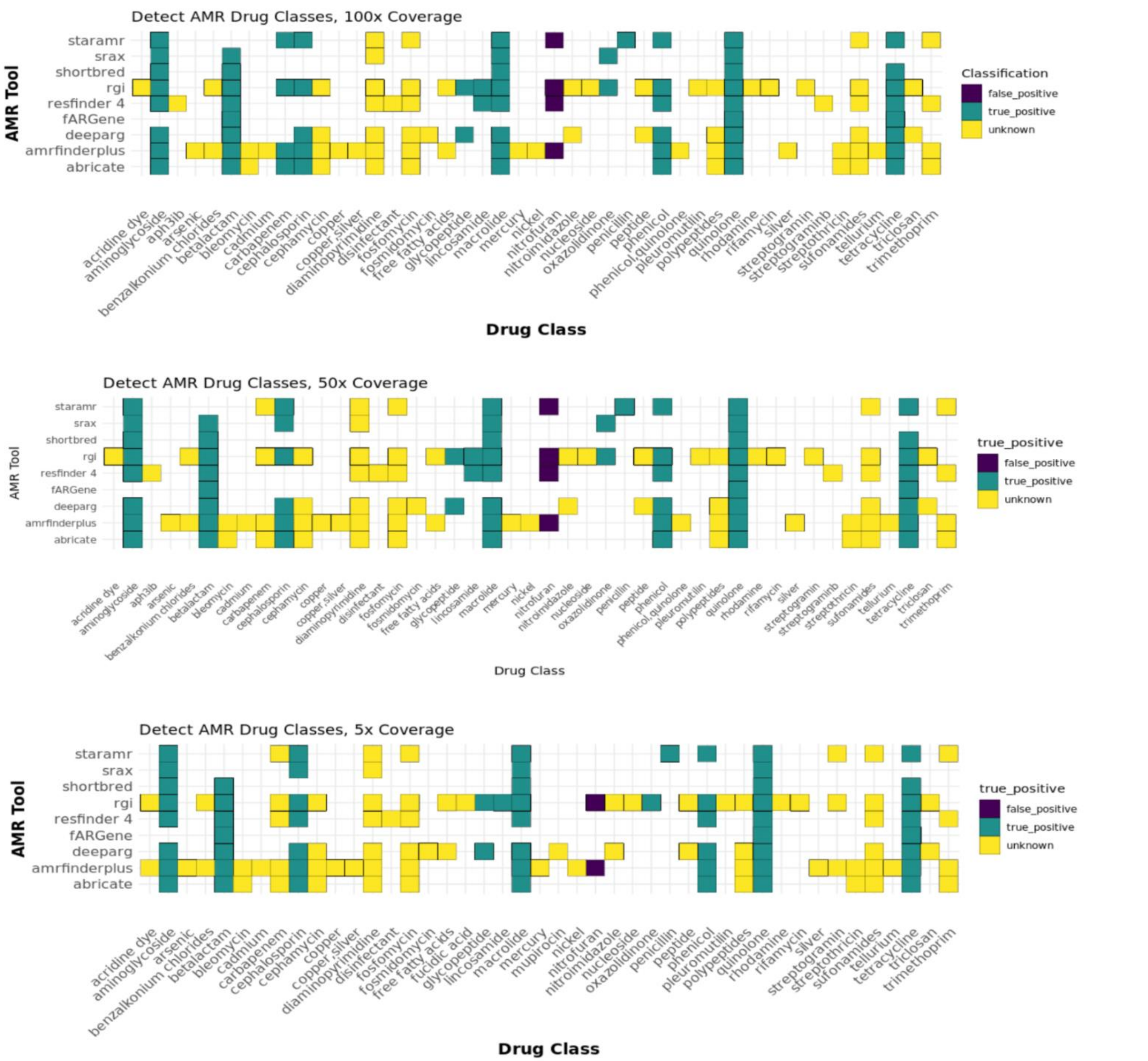
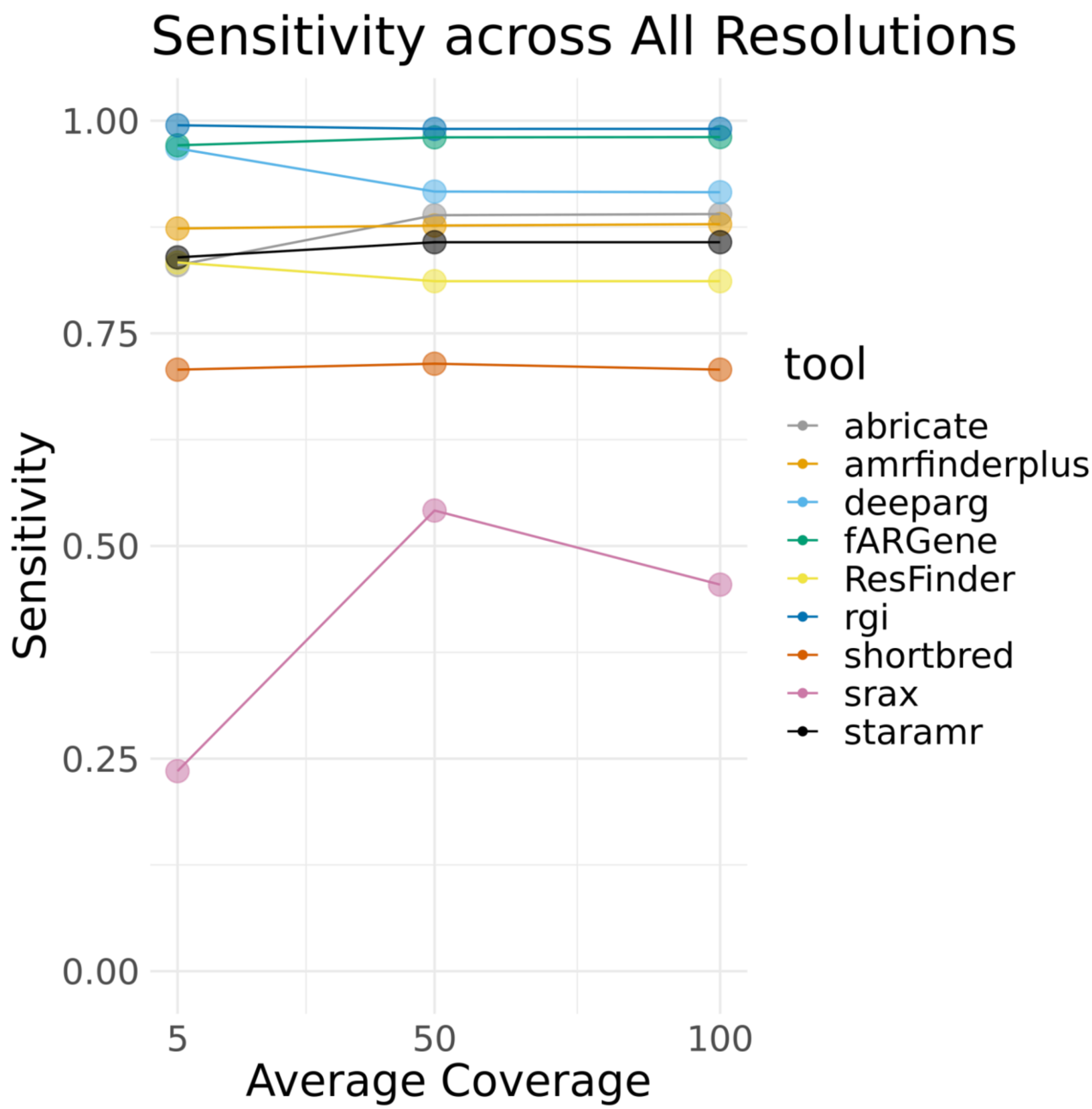
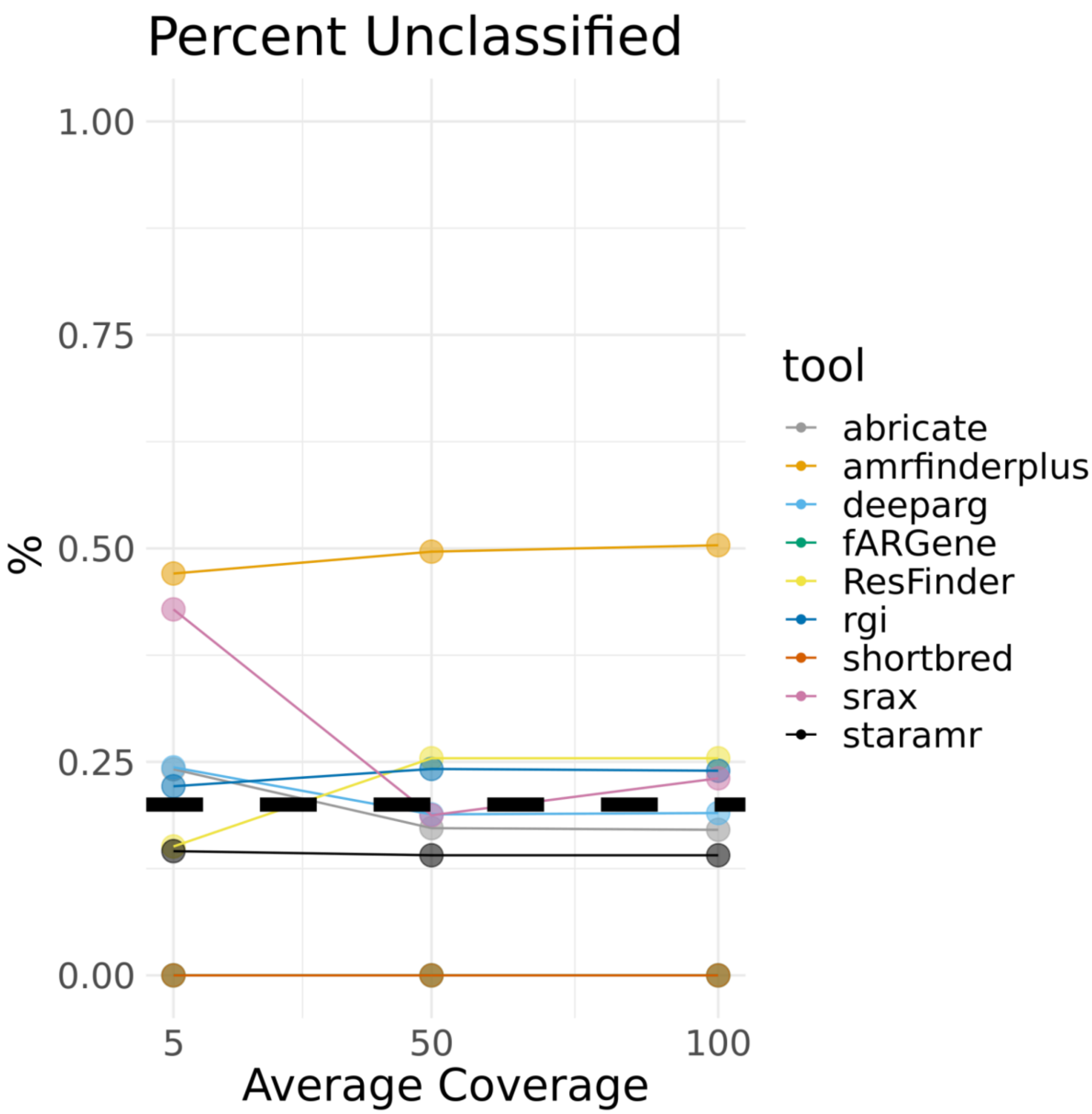


Figure 2 Sensitivity of Software Tools for Detection of Antimicrobial Resistance (AMR) Genes Across Coverage Levels



Sensitivity was calculated as (true positives) / (true positives + false negatives). Most tools were highly sensitive (greater than 0.80). All genes corresponding to "Other" or "Unknown" drug classes were not included in these calculations. Similarly, AMR genes corresponding to phenotypic resistance that was not tested in the mock community was considered "Unknown" and not included in the sensitivity analysis.

Figure 3: Percent Detection of Unknown Antimicrobial (AMR) Resistance Genes Across Coverage



The percent detection of AMR genes that could not be classified because the material the gene confers resistance to was not tested in the mock community. A black dashed line is placed at 0.20, indicating where at least 20% of the detected AMR genes could not be classified.

Supplementary Figure 1: Pictorial Methods

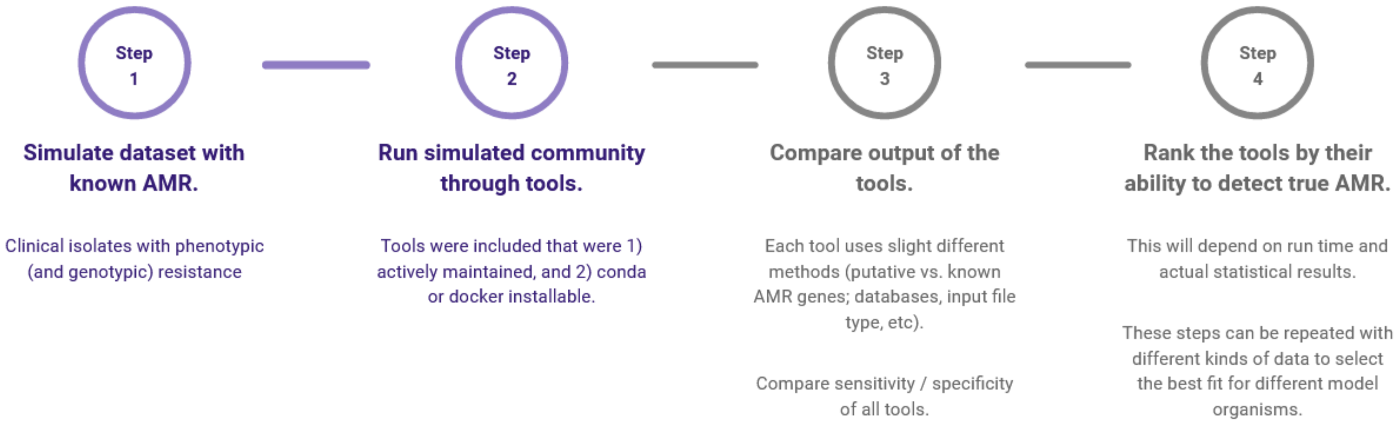


Table 1: Clinical isolates included in the simulated community. (susceptibility test is in the spreadsheet, will have to be supplemental bc so big)

Strain	Testing Standard (CLSI or EUCAST)	BioSample ID	Link
Neisseria gonorrhoeae SW0011	CLSI	SAMN15960549	https://www.ncbi.nlm.nih.gov/biosample/SAMN15960549
Klebsiella pneumoniae CCUG 70742	EUCAST	SAMN07602587	https://www.ncbi.nlm.nih.gov/biosample/SAMN07602587
Pseudomonas aeruginosa CCUG 70744	EUCAST	SAMN07602569	https://www.ncbi.nlm.nih.gov/biosample/SAMN07602569 /
Acinetobacter baumannii MRSN489669	CLSI	SAMN12087686	https://www.ncbi.nlm.nih.gov/biosample/SAMN12087686
Enterobacter cloacae 174	CLSI	SAMN04456586	https://www.ncbi.nlm.nih.gov/biosample/SAMN04456586
Citrobacter freundii	CLSI	SAMN13412315	https://www.nc

MRSN12115			bi.nlm.nih.gov/biosample/SAMN13412315
Staphylococcus aureus LAC	CLSI	SAMN08391108	https://www.ncbi.nlm.nih.gov/biosample/SAMN08391108
Escherichia coli 222	CLSI	SAMN05194390	https://www.ncbi.nlm.nih.gov/biosample/SAMN05194390

583 Table 2: Tools identified from search methods with the selection criteria and whether
584 they subsequently worked or not.

Tool	Conda / Docker Installable?	Actively Maintained?	Input format?	Included in Analysis?	Implementation Method	Database
ABRIcate	Yes - conda	Yes	FASTA	Yes	Align reads to database	NCBI AMRFinder Plus, CARD, ResFinder, ARG-ANNOT, MEGARES, EcOH, PlasmidFinder, VFDB, and Ecoli_VF
shortBRED	Yes - Docker & conda	Yes	FASTA	Yes	Align reads to database	AMR gene marker database from 849 AR protein families from the ARDB19 and independent curation

fARGene	Yes - conda	Yes	FASTQ	Yes	Compare to AMR model	Hidden markov models for quinolone, tetracycline, and beta lactamases
RGI	Yes - Docker (conda outdated)	Yes	FASTQ	Yes	Compare to AMR model	Prodigal predicts ORF and compared to CARD and WildCARD
ResFinder 4	Yes - Docker (conda broken)	Yes	FASTA	Yes	Align reads to database	ResFinder 4 database
DeepARG	Yes, Docker	Unclear	FASTA	Yes	Compare to AMR model	Supervised deep learning compares reads to antibiotic resistance categories created from CARD, ARDB, and UNIPROT
sraX	Yes - both	Yes	FASTA	Yes	Align reads to database	CARD by default

starAMR	Yes - conda	Yes	FASTA	Yes	Align reads to database	ResFinder, PointFinder, and PlasmidFinder
AMR Finder Plus	Yes - conda	Yes	FASTA	Yes	Align reads to database	Pathogen Detection Reference Gene Database
ResPipe	No	Yes	FASTQ or BAM	No		
PointFinder	Yes - Docker	Yes	FASTA	No		
PCM: Pairwise Comparative Modelling	No	Yes	FASTA - protein	No		
SRST2	No	No	FASTQ	No		
Arg_Ranker	Yes - conda	Yes	Requires special metadata input	No		
MetaCherchant	Yes - conda	No	FASTA -	No		

			genomic			
ARIBA	Yes - Docker	No	Paired end FASTQ	No		
ARG-ANNOT	No	No	Unclear	No		
kmerresistance	No	No	-	No		
c-sstar	No	No	Unknown	No - could not track down github		

585

586

Table 3: Summary Statistics from hAMRoaster: These are the counts and metrics as calculated by the hAMRoaster pipeline. Formulas for all metrics are as follows:

Specificity = TN / (TN + FP)

Sensitivity = TP / (TP + FN)

Precision = TP / (TP + FP)

Accuracy = (TP + TN) / (TP + FP + TN + FN)

Recall = true pos / (true pos + false neg)

F1 = 2 * (precision * recall) / (precision + recall)

Percent_unknown = unknown / (true_positives + false_positives + unknowns)

Full Results, 100x Coverage											
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**recall**	**Percent unclassified**
abricate	0	66	22	9	2	0.8800	1.0000	1.0000	0.8831	7.3333	0.2500
amrfinderplus	2	62	71	9	1	0.8732	0.3333	0.9688	0.8514	5.6364	0.5259
deeparg	0	98	23	8	2	0.9245	1.0000	1.0000	0.9259	12.2500	0.1901
fARGene	0	713	0	13	2	0.9821	1.0000	1.0000	0.9821	54.8462	0.0000
resfinder 4	1	43	15	9	1	0.8269	0.5000	0.9773	0.8148	4.3000	0.2542
rgi	4	559	255	6	1	0.9894	0.2000	0.9929	0.9825	55.9000	0.3117
shortbred	0	29	0	11	2	0.7250	1.0000	1.0000	0.7381	2.6364	0.0000
srax	0	10	3	11	2	0.4762	1.0000	1.0000	0.5217	0.9091	0.2308
staramr	1	52	11	9	1	0.8525	0.5000	0.9811	0.8413	.2000	0.1719

Full Results, 50x Coverage											
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**recall**	**Percent unclassified**

abricate	0	66	21	9	2	0.8800	1.0000	1.0000	0.8831	7.333	0.2414
amrfinderplus	2	62	67	9	1	0.8732	0.3333	0.9688	0.8514	5.636	0.5115
deeparg	0	99	23	8	2	0.9252	1.0000	1.0000	0.9266	12.375	0.1885
fARGene	0	702	0	13	2	0.9818	1.0000	1.0000	0.9819	54.000	0.0000
resfinder 4	1	43	15	9	1	0.8269	0.5000	0.9773	0.8148	4.300	0.2542
rgi	4	557	254	6	1	0.9893	0.2000	0.9929	0.9824	55.700	0.3117
shortbred	0	30	0	11	2	0.7317	1.0000	1.0000	0.7442	2.727	0.0000
srax	0	13	3	10	2	0.5652	1.0000	1.0000	0.6000	1.300	0.1875
staramr	1	52	11	9	1	0.8525	0.5000	0.9811	0.8413	5.200	0.1719
Full Results, 5x Coverage											
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**recall**	**Percent unclassified**
abricate	0	9	39	19	2	0.8125	1.0000	1.0000	0.8200	4.3333	0.3276
amrfinderplus	1	9	60	58	1	0.8696	0.5000	0.9836	0.8592	6.0000	0.4874

deeparg	0	8	267	86	2	0.9709	1.0000	1.0000	0.9711	33.3750	0.2436
fARGene	0	13	470	0	2	0.9731	1.0000	1.0000	0.9732	36.1538	0.0000
resfinder 4	0	9	43	10	2	0.8269	1.0000	1.0000	0.8333	4.7778	0.1887
rgi	12	6	1015	418	1	0.9941	0.0769	0.9883	0.9826	56.3889	0.2893
shortbred	0	11	29	0	2	0.7250	1.0000	1.0000	0.7381	2.6364	0.0000
srax	0	12	4	3	2	0.2500	1.0000	1.0000	0.3333	0.3333	0.4286
staramr	0	9	44	11	2	0.8302	1.0000	1.0000	0.8364	4.8889	0.2000

Table 4: Condensed Summary Statistics: This table contains the counts and metrics if the data were condensed so that overlapping genes are excluded from the count data (i.e. genes that start between the start and stop codon of another gene are not considered in analysis).

Condensed Results, 100x Coverage											
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**recall**	**Percent unclassified**
abricate	0	21	5	0	2	1	1	1	1	1	0.1923
amrfinderplus	0	22	23	0	2	1	1	1	1	1	0.5111
deeparg	0	2	1	0	2	1	1	1	1	1	0.3333
fARGene	0	713	0	0	2	1	1	1	1	1	0
resfinder 4	0	12	5	0	2	1	1	1	1	1	0.2941
rgi	1	77	38	0	1	1	0.9872	0.5	0.9872	1	0.32769
shortbred	0	29	0	0	2	1	1	1	1	1	0
srax	0	10	3	0	2	1	1	1	1	1	0.23078
staramr	1	36	6	0	1	1	0.9730	0.5	0.9737	1	0.1395
Condensed Results, 50x Coverage											
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**recall**	**Percent unclassified**

abricate	0	22	3	0	2	1	1	1	1	1	1
amrfinderplus	0	20	27	0	2	1	1	1	1	1	1
deeparg	0	1	1	0	2	1	1	1	1	1	1
fARGene	0	702	0	0	2	1	1	1	1	1	1
resfinder 4	0	11	7	0	2	1	1	1	1	1	1
rgi	1	75	38	0	1	1	0.9868421053	0.9868	0.5000	0.9870	1
shortbred	0	30	0	0	2	1	1	1	1	1	1
srax	0	13	3	0	2	1	1	1	1	1	1
staramr	1	29	7	0	1	1	0.9666666667	0.9667	0.5000	0.9677	1
Condensed Results, 5x Coverage											
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**recall**	**Percent unclassified**
abricate	0	4	3	0	2	1	1	1	1	1	0.4286
amrfinderplus	0	7	11	0	2	1	1	1	1	1	0.6111

deeparg	0	42	7	0	2	1	1	1	1	1	0.1429
fARGene	0	470	0	0	2	1	1	1	1	1	0.0000
resfinder 4	0	6	2	0	2	1	1	1	1	1	0.2500
rgi	1	48	30	0	1	1	0.9796	0.5000	0.9800	1	0.3797
shortbred	0	29	0	0	2	1	1	1	1	1	0.0000
srax	0	4	3	0	2	1	1	1	1	1	0.4286
staramr	0	33	8	0	2	1	1	1	1	1	0.1951

Table 5: Kappa Values: Kappa values (agreement) between tools across coverage levels calculated in R using the kappa2 function

5X Kappa	amrfinder…	deeparg	fARGene	resfinder 4	rgi	shortbred	srax	staramr
abricate	0.49231	0.48069	0.22703	0.74419	0.20532	0.36649	0.36649	0.64186
amrfinderp…	0	-0.00797	0.08027	0.30483	-0.1448	0.13652	0.13652	0.30483
deeparg	0	0	0.17564	0.42232	0.3901	0.28868	0.28868	0.32604
fARGene	0	0	0	0.29712	0.07293	0.72671	0.18012	0.15655
resfinder 4	0	0	0	0	0.19561	0.46828	0.46828	0.78164
rgi	0	0	0	0	0	0.12438	0.12438	0.19561
shortbred	0	0	0	0	0	0	0.54872	0.33535
srax	0	0	0	0	0	0	0	0.46828

50X Kappa	amrfinder…	deeparg	fARGene	resfinder 4	rgi	shortbred	srax	staramr
abricate	0.50185	0.56961	0.22925	0.45902	0.25277	0.36943	0.38964	0.64567
amrfinderp…	0	0.11929	0.08313	0.13699	-0.13445	0.14108	0.12511	0.31513
deeparg	0	0	0.21053	0.32258	0.4548	0.34146	0.35867	0.40529
fARGene	0	0	0	0.25	0.08313	0.72727	0.33824	0.15888
resfinder 4	0	0	0	0	0.13699	0.4	0.30769	0.58621
rgi	0	0	0	0	0	0.14108	0.20118	0.23456
shortbred	0	0	0	0	0	0	0.61702	0.33824
srax	0	0	0	0	0	0	0	0.37326

100X Kappa	amrfinder…	deeparg	fARGene	resfinder 4	rgi	shortbred	srax	staramr
abricate	0.50185	0.56961	0.22925	0.45902	0.25277	0.36943	0.32331	0.64567
amrfinderp…	0	0.11929	0.08313	0.13699	-0.13445	0.14108	0.09548	0.31513
deeparg	0	0	0.21053	0.32258	0.4548	0.34146	0.29603	0.40529
fARGene	0	0	0	0.25	0.08313	0.72727	0.39024	0.15888
resfinder 4	0	0	0	0	0.13699	0.4	0.35294	0.58621
rgi	0	0	0	0	0	0.14108	0.17085	0.23456
shortbred	0	0	0	0	0	0	0.68966	0.33824
srax	0	0	0	0	0	0	0	0.29185

Supplementary Table 1: Summary Statistics when results of all tools are combined.

Combined Stats			
	100x	**50x**	**5x**
true_positive	1703	1624	1971
unknown	329	394	605
false_positive	8	8	13
true_negatives	1	1	1
false_negatives	6	6	6
sensitivity	0.996	0.996	0.996
specificity	0.111	0.111	0.071

Supplementary table 2: link to tweet

https://twitter.com/emily_wissel/status/1336013892116488195

Supplementary table 3: tidy table of data

https://docs.google.com/spreadsheets/d/1bfACqEh0nkS65vCUj5DfMg4PvW0fHxbtrv0PgKt1gT4/edit#gid=53644837