

1 **Benchmarking software to predict antibiotic**
2 **resistance phenotypes in shotgun**
3 **metagenomes using simulated data**

4
5 Emily F. Wissel^A, Brooke M. Talbot^B, Bjorn A. Johnson^C, Robert A Petit III^D, Vicki
6 Hertzberg^A, Anne Dunlop^E, Timothy D. Read^{D,F}

7 *Author affiliations and ORCID*

8 **A:** Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, US

9 **B:** Population Biology, Ecology, and Evolution Program, Graduate Division of Biological and
10 Biomedical Science, Emory University, Atlanta, GA, US

11 **C:** Cockrell School of Engineering, The University of Texas at Austin, Austin, TX

12 **D:** Division of Infectious Diseases, Department of Medicine, School of Medicine, Emory
13 University, Atlanta, GA, US

14 **E:** Department of Gynecology & Obstetrics, Emory University School of Medicine

15 **F:** Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, US

16

17 **Author Info**

18 EFW: ewissel@emory.edu <https://orcid.org/0000-0003-2275-8456>

19 BMT: <https://orcid.org/0000-0001-5246-7209>

20 BAJ: <https://orcid.org/0000-0002-6460-2444>

21 RAP: <https://orcid.org/0000-0002-1350-9426>

22 VH: <https://orcid.org/0000-0002-8834-4363>

23 ALD: <https://orcid.org/0000-0002-5092-8136>

24 TDR: tread@emory.edu ORCID:0000-0001-8966-9680

BENCHMARKING AMR SOFTWARE

2

25 **Abstract**

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

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

41 This study revealed that tools were highly variable in sensitivity (0.25 - 0.99) and
42 specificity (0.2 - 1) in detection of resistance in our synthetic FASTQ files despite similar
43 databases and methods implemented. Tools performed similarly at all coverage levels (5x, 50x,
44 100x). Cohen's kappa revealed low agreement across tools.

45 **Importance**

46 Software selection for metagenomic AMR prediction should be driven by the context of
47 the clinical/research questions and tolerance for true and false negative results. As the prediction

BENCHMARKING AMR SOFTWARE

3

48 software and databases are in a state of constant refinement, the approach used here—creating
49 synthetic communities containing taxa and phenotypes of interest along with using hAMRoaster
50 to assess performance of candidate software—offers a template to aid researchers in selecting the
51 most appropriate strategy.

52

53 **Keywords:** antimicrobial resistance, bioinformatics, metagenomics

54

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

57

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

60

61

62

63 **Introduction**

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

76 AMR surveillance is a key strategy in understanding the threat of AMR. Currently, AMR
77 surveillance typically relies on phenotypic characterization through culture or genotypic
78 characterization through molecular diagnostics based on PCR and hybridization techniques.⁴
79 However, there is a move toward genome-based methods⁵ with the Illumina short-read platform
80 being the dominant platform for data generation at the present time.⁶ Direct sequencing of
81 clinical samples using shotgun metagenomic approaches is of growing interest for minimizing
82 sample processing and for fully characterizing the commensal members of the microbiome.
83 However, the bioinformatic tools that currently exist to detect AMR have typically not been
84 assessed for their performance on shotgun metagenomic sequence data. Further, as is common
85 with software developed in academic settings, tools are not always maintained or easy to install.

BENCHMARKING AMR SOFTWARE

5

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

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

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

99 **Methods**

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

101 **Development of a software pipeline, hAMRoaster, to assess results of antibiotic resistance** 102 **prediction**

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

BENCHMARKING AMR SOFTWARE

6

109 the Bioconda¹⁰ database. The Github site for the software is

110 <https://github.com/ewissel/hAMRoaster>.

111 hAMRoaster requires, as input, a formatted results table of runs by AMR detection tools.

112 This table is identical to that produced by the hAMRonization⁹ software. hAMRonization is

113 conda installable and can compile the outputs of many AMR tools into a unified format.

114 shortBRED¹¹ and fARGene¹² are not included in hAMRonization at the time of analysis, so

115 hAMRoaster can take the path to the raw output for these tools and partially match it to the

116 hAMRonization output.

117 hAMRoaster requires an input to the “known” phenotypic resistance in the mock

118 community (--AMR_key flag of hAMRoaster), such as a result of susceptibility testing tables

119 that are available from NCBI Biosamples. Antibiotics in the table of known resistances are

120 matched to their respective drug classes. Results classified as “susceptible” or “intermediate” in

121 susceptibility testing are filtered out so only resistant instances are considered. In cases where

122 susceptibility testing occurred with two or more agents, each agent is considered independently

123 (e.g. resistance to “amoxicillin-tetracycline” was treated as resistance to “amoxicillin” and

124 “tetracycline” independently). Each identified AMR gene is labeled with its corresponding drug

125 class for comparison. In instances where a gene confers resistance to multiple drug classes, the

126 detected gene is split into multiple rows so that each conferred resistance can be independently

127 compared to what is known from the susceptibility testing. Gene to drug class linkage is verified

128 using the CARD database¹³ when applicable. Any genes corresponding to ‘unknown’ or ‘other’

129 drug classes (including hypothetical resistance genes) are excluded from further analysis. Genes

130 that confer resistance to an antibiotic that was only effective in combination with another drug

BENCHMARKING AMR SOFTWARE

7

131 (e.g. clavulanic acid in amoxicillin-clavulanic acid) are classified as ‘Other’ and excluded from
132 analysis.

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

141 **Creation of a synthetic mock communities of antibiotic resistance bacteria**

142 Bacterial members of the base mock community were chosen from NCBI’s BioSample
143 Database¹⁴ and met the following criteria: (1) the strain had extensive antibiotic susceptibility
144 testing data using CLSI or EUCAST testing standards as part of the public NCBI BioSample
145 record; (2) the strain was isolated from human tissue; (3) the strain was the cause of a clinical
146 infection; (4) the FASTA was available to download from NCBI BioSample Database.¹⁴ Eight
147 bacteria, each representing a different species, with overlapping resistance to 43 antibiotics
148 across 18 drug classes, were selected for the mock community (**Table 1**). The included taxa were
149 *Acinetobacter baumannii* MRSN489669, *Citrobacter freundii* MRSN12115, *Enterobacter*
150 *cloacae* 174, *Escherichia coli* 222, *Klebsiella pneumoniae* CCUG 70742, *Pseudomonas*
151 *aeruginosa* CCUG 70744, *Neisseria gonorrhoeae* SW0011, and *Staphylococcus aureus* LAC
152 (Table 1).

BENCHMARKING AMR SOFTWARE

8

153 Paired-end FASTQs were simulated by ART¹⁵ using default parameters for HiSeq 2500
154 at three levels of average sequence coverage (5x, 50x, and 100x) and are available on FigShare
155 (<https://figshare.com/account/home#/projects/125974>). Simulated FASTQs were subsequently
156 concatenated to resemble shotgun metagenomics reads, and metaSPAdes¹⁶ was used to create
157 assembled contigs. The FASTQs were simulated with approximately equal numbers of reads of
158 each genome.

159 **Running antibiotic prediction software on mock communities**

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

163 **Statistical Analysis**

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

170 **Data Availability**

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

174 **Results**

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

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

185 In order for an identified tool to be considered eligible for comparison, it had to meet the
186 following criteria: (1) be conda or Docker installable; (2) have source code publicly available in
187 a data repository and be actively maintained (defined as tool updates or GitHub responses within
188 the last year); (3) have an open source license; and (4) take FASTQs or FASTAs as input files.
189 Nine tools met the criteria to be included in this analysis: ABRicate¹⁷, fARGene¹⁸ ResFinder¹⁹,
190 shortBRED¹¹, RGI²⁰, AMRFinderPlus²¹, starAMR²², sraX²³, and deepARG²⁴. PointFinder
191 also qualified²⁵, but was a subtool of ResFinder and only identified mutational resistance for
192 some organisms, so it was excluded from analysis. The code used to install and run all tools is
193 available on the hAMRoaster GitHub.

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

BENCHMARKING AMR SOFTWARE

10

196 **ABRIcate**

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

201 **shortBRED**

202 shortBRED¹¹ v0.9.3 used a set of marker genes to search metagenomic data for protein
203 families of interest. The bioBakery³² team published an AMR gene marker database built from
204 849 AR protein families derived from the ARDB³³ v1.1 and independent curation alongside
205 shortBRED, which is used in this study.

206 **fARGene**

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

211 **RGI**

212 RGI²⁰ v5.1.1 used protein homology and SNP models to predict 'resistomes'. It used
213 CARD's protein homolog models as a database. RGI predicts open reading frames (ORFs) using
214 Prodigal,³⁴ detects homologs with DIAMOND,³⁵ and matches to CARD's database and model
215 cut off values.

BENCHMARKING AMR SOFTWARE

11

216 **ResFinder**

217 ResFinder¹⁹ v4.0 was available both as a web-based application or the command line. We
218 used ResFinder 4 in this study, which was specifically designed for detecting genotypic
219 resistance in phenotypically resistant samples. ResFinder aligned reads directly to its own
220 curated database without need for assembly.

221 **deepARG**

222 deepARG²⁴ v.2.0 used a supervised deep learning based approach for antibiotic resistance
223 gene annotation of metagenomic sequences. It combines three databases—CARD, ARDB, and
224 UNIPROT—and categorizes them into resistance categories.

225 **sraX**

226 sraX²³ v.1.5 was built as a one step tool; in a single command, sraX downloads a
227 database and aligns contigs to this database with DIAMOND³⁵. By default, sraX uses CARD,
228 though other options can be specified. As we use default settings for all tools, only CARD was
229 used in this study for sraX.

230 **starAMR**

231 starAMR^{22,36} v.0.7.2 uses BLAST+³⁷ to compare contigs against a combined database
232 with data from ResFinder, PointFinder, and PlasmidFinder.

233 **AMR Finder Plus**

234 AMR Finder Plus²¹ v.3.9.3 uses BLASTX³⁸ translated searches and hierarchical tree of
235 gene families to detect AMR genes. The database was derived from the Pathogen Detection
236 Reference Gene Catalog³⁹ and was compiled as part of the National Database of Antibiotic
237 Resistant Organisms (NDARO).

BENCHMARKING AMR SOFTWARE

12

238 **Performance of selected tools on a mock bacterial community containing 43 laboratory** 239 **confirmed AMR phenotypes**

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

244 **Range of phenotypes detected**

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

251 **Sensitivity and Specificity**

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

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

BENCHMARKING AMR SOFTWARE

13

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

264 **Condensing Results**

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

272 **Concordance between tools**

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

276 **Discussion**

277 **Development of a framework for assessing AMR prediction software performance using** 278 **synthetic data**

279 There is a considerable research effort to develop new software for predicting AMR
280 using DNA sequence alone. In this dynamic environment, there is a need for researchers and
281 epidemiologists to understand the relative performance of open source software tools within the
282 types of sample they may encounter. While some tools currently exist for compiling the results
283 of several AMR tools together (hAMRonizer and chARMedDb⁴⁰), this study was motivated by

BENCHMARKING AMR SOFTWARE

14

284 the lack of an open-source pipeline for comparing the results once compiled. hAMRoaster was
285 built so that several metrics can easily be compared across tools.

286 The central challenge in developing this software was to compare detected AMR genes to
287 resistance phenotypes. Detected AMR genes needed to be classified by their corresponding drug
288 class(es) so they could be matched to the known phenotypically resistant drug classes. One
289 hurdle in this translation is that tools use different databases, and some databases classify genes
290 differently. For example, shortBRED classifies gene families, while CARD classifies specific
291 genes. While this analysis checked the drug classification via the DNA/Protein Accession value
292 in CARD, only around 300 of the >1,000 genes detected could directly map to genes in CARD
293 by accession value. The hAMRonization tool overcomes this challenge by providing a drug class
294 column and filling in the values from ChEBI ontology⁴¹ when possible. The hAMRoaster
295 strategy is to assign a CARD drug class value to every detected AMR gene first by accession
296 number, then by gene name. If neither of these methods assign a drug class for an AMR gene,
297 then the drug class provided by hAMRonization is used. Another challenge in converting
298 detected AMR genes to drug classes is that some drugs are only administered in combination, for
299 example clavulanic acid with amoxicillin. For these instances, resistance to the drug only used in
300 combination (e.g. clavulanic acid) is treated as an “other” drug class and excluded from analysis.
301 In these cases, we used the experience of practicing clinicians to identify combination
302 antibiotics.

303 The analysis presented here used synthetic data to compare tool performance. Synthetic
304 data has the benefit of allowing controlled input with known ground truth. Therefore users can
305 focus on the types of organisms and phenotypes they need to detect in their own datasets,
306 perform experiments with real samples, and manipulate a range of factors such as relative

BENCHMARKING AMR SOFTWARE

15

307 abundance and sequencing error. The NCBI BioSample repository (used in this study) is an
308 invaluable resource for creating such datasets as it contains many samples with AMR phenotypes
309 determined by international standards. Researchers could also sequence and phenotype
310 culturable organisms in their own laboratories to provide testing standards to evaluate software.
311 Here, we exclusively examined synthetic short read Illumina data, but this analysis strategy
312 could be adapted to understand the effect of using data generated on long read technologies such
313 as the Pacific Bioscience and Oxford Nanopore platforms.

314 **Overall trends in performance and reasons for variability between tools**

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

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

325 hAMRoaster calculates specificity, precision, accuracy, recall, and F1 (**Table 3**).
326 However, all of these measures are dependent on false positives and/or true negatives in their
327 calculations. As these values are inherently low in our mock community due to the robust
328 resistance profile, these metrics are not particularly informative for understanding how well these
329 tools detect resistance in this phenotypically resistant sample. Similarly, we calculated all

BENCHMARKING AMR SOFTWARE

16

330 effective metrics when the results of all tools are combined. While sensitivity in the combined
331 data was very high (>0.99), there was a very high number of overall detected AMR genes,
332 including overlapping results between genes, thus, it would be difficult for researchers to
333 meaningfully use this type of result to understand the AMR profile. We calculated Cohen's
334 kappa to capture the agreement at the drug class level between AMR tools to see if all AMR
335 tools detected resistance to the same drug classes. We found that agreement was surprisingly
336 low across all tools (**Table 6**), indicating that some tools may be better suited for detecting
337 different types of resistance. As such, hAMRoaster provides a table with the number of genes
338 detected per drug class for each tool.

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

350 In conclusion, this study compared bioinformatics tools for detecting AMR genes in a
351 simulated short read metagenomic sample at three coverage levels at one time point. While tools
352 use slightly different methods and databases, these tools overall had high sensitivity for detection

BENCHMARKING AMR SOFTWARE

17

353 of AMR genes. Moreover agreement between tools was low, indicating the importance of tool
354 selection. In our test set we found starAMR had the highest sensitivity value with fewer than
355 20% unknown detected genes at all coverage levels. We advocate that researchers should test
356 these software tools using pipelines such as hAMRoaster with a synthetic community that
357 highlights the resistance profiles and sample of interest. In particular, this assessment of
358 performance of available tools should take place before the commencement of the study as the
359 set of tools for detecting AMR genes are actively maintained and undergoing further
360 improvements.

361 **Acknowledgements**

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

363 **Funding**

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

366 **Author Contributions**

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

371

372 **References**

- 373 (1) Centers for Disease Control and Prevention (U.S.). *Antibiotic Resistance Threats in the*
374 *United States, 2019*; Centers for Disease Control and Prevention (U.S.), 2019.
375 <https://doi.org/10.15620/cdc:82532>.
- 376 (2) Nji, E.; Kazibwe, J.; Hambridge, T.; Joko, C. A.; Larbi, A. A.; Dampthey, L. A. O.;
377 Nkansa-Gyamfi, N. A.; Stålsby Lundborg, C.; Lien, L. T. Q. High Prevalence of
378 Antibiotic Resistance in Commensal Escherichia Coli from Healthy Human Sources in
379 Community Settings. *Sci. Rep.* **2021**, *11* (1), 3372. [https://doi.org/10.1038/s41598-021-](https://doi.org/10.1038/s41598-021-82693-4)
380 [82693-4](https://doi.org/10.1038/s41598-021-82693-4).
- 381 (3) Brinkac, L.; Voorhies, A.; Gomez, A.; Nelson, K. E. The Threat of Antimicrobial
382 Resistance on the Human Microbiome. *Microb. Ecol.* **2017**, *74* (4), 1001–1008.
383 <https://doi.org/10.1007/s00248-017-0985-z>.
- 384 (4) Anjum, M. F.; Zankari, E.; Hasman, H. Molecular Methods for Detection of Antimicrobial
385 Resistance. *Microbiol. Spectr.* **2017**, *5* (6). [https://doi.org/10.1128/microbiolspec.ARBA-](https://doi.org/10.1128/microbiolspec.ARBA-0011-2017)
386 [0011-2017](https://doi.org/10.1128/microbiolspec.ARBA-0011-2017).
- 387 (5) Nutrition, C. for F. S. and A. GenomeTrakr Network. *FDA* **2021**.
- 388 (6) Porter, T. M.; Hajibabaei, M. Over 2.5 Million COI Sequences in GenBank and Growing.
389 *PLOS ONE* **2018**, *13* (9), e0200177. <https://doi.org/10.1371/journal.pone.0200177>.
- 390 (7) Oniciuc, E. A.; Likotrafiti, E.; Alvarez-Molina, A.; Prieto, M.; Santos, J. A.; Alvarez-
391 Ordóñez, A. The Present and Future of Whole Genome Sequencing (WGS) and Whole
392 Metagenome Sequencing (WMS) for Surveillance of Antimicrobial Resistant
393 Microorganisms and Antimicrobial Resistance Genes across the Food Chain. *Genes* **2018**,
394 *9* (5), 268. <https://doi.org/10.3390/genes9050268>.
- 395 (8) Dulanto Chiang, A.; Dekker, J. P. From the Pipeline to the Bedside: Advances and
396 Challenges in Clinical Metagenomics. *J. Infect. Dis.* **2020**, *221* (Supplement_3), S331–
397 S340. <https://doi.org/10.1093/infdis/jiz151>.
- 398 (9) Issues · pha4ge/hAMRonization <https://github.com/pha4ge/hAMRonization> (accessed
399 2021 -10 -12).
- 400 (10) Grüning, B.; Dale, R.; Sjödin, A.; Chapman, B. A.; Rowe, J.; Tomkins-Tinch, C. H.;
401 Valieris, R.; Köster, J. Bioconda: Sustainable and Comprehensive Software Distribution
402 for the Life Sciences. *Nat. Methods* **2018**, *15* (7), 475–476.
403 <https://doi.org/10.1038/s41592-018-0046-7>.
- 404 (11) Kaminski, J.; Gibson, M. K.; Franzosa, E. A.; Segata, N.; Dantas, G.; Huttenhower, C.
405 High-Specificity Targeted Functional Profiling in Microbial Communities with
406 ShortBRED. *PLOS Comput. Biol.* **2015**, *11* (12), e1004557.
407 <https://doi.org/10.1371/journal.pcbi.1004557>.
- 408 (12) fannyhb. *FARGene*; 2021.
- 409 (13) Alcock, B. P.; Raphenya, A. R.; Lau, T. T. Y.; Tsang, K. K.; Bouchard, M.; Edalatmand,
410 A.; Huynh, W.; Nguyen, A.-L. V.; Cheng, A. A.; Liu, S.; Min, S. Y.; Miroshnichenko, A.;
411 Tran, H.-K.; Werfalli, R. E.; Nasir, J. A.; Oloni, M.; Speicher, D. J.; Florescu, A.; Singh,
412 B.; Faltyn, M.; Hernandez-Koutoucheva, A.; Sharma, A. N.; Bordeleau, E.; Pawlowski, A.
413 C.; Zubyk, H. L.; Dooley, D.; Griffiths, E.; Maguire, F.; Winsor, G. L.; Beiko, R. G.;
414 Brinkman, F. S. L.; Hsiao, W. W. L.; Domselaar, G. V.; McArthur, A. G. CARD 2020:
415 Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance
416 Database. *Nucleic Acids Res.* **2020**, *48* (D1), D517–D525.

BENCHMARKING AMR SOFTWARE

19

- 417 <https://doi.org/10.1093/nar/gkz935>.
- 418 (14) Barrett, T.; Clark, K.; Gevorgyan, R.; Gorelenkov, V.; Gribov, E.; Karsch-Mizrachi, I.;
419 Kimelman, M.; Pruitt, K. D.; Resenchuk, S.; Tatusova, T.; Yaschenko, E.; Ostell, J.
420 BioProject and BioSample Databases at NCBI: Facilitating Capture and Organization of
421 Metadata. *Nucleic Acids Res.* **2012**, *40* (D1), D57–D63.
422 <https://doi.org/10.1093/nar/gkr1163>.
- 423 (15) Huang, W.; Li, L.; Myers, J. R.; Marth, G. T. ART: A next-Generation Sequencing Read
424 Simulator. *Bioinforma. Oxf. Engl.* **2012**, *28* (4), 593–594.
425 <https://doi.org/10.1093/bioinformatics/btr708>.
- 426 (16) Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P. A. MetaSPAdes: A New Versatile
427 Metagenomic Assembler. *Genome Res.* **2017**, *27* (5), 824–834.
428 <https://doi.org/10.1101/gr.213959.116>.
- 429 (17) Seemann, T. *ABRicate*; 2021.
- 430 (18) Berglund, F.; Österlund, T.; Boulund, F.; Marathe, N. P.; Larsson, D. G. J.; Kristiansson,
431 E. Identification and Reconstruction of Novel Antibiotic Resistance Genes from
432 Metagenomes. *Microbiome* **2019**, *7* (1), 52. <https://doi.org/10.1186/s40168-019-0670-1>.
- 433 (19) Bortolaia, V.; Kaas, R. S.; Ruppe, E.; Roberts, M. C.; Schwarz, S.; Cattoir, V.; Philippon,
434 A.; Allesoe, R. L.; Rebelo, A. R.; Florensa, A. F.; Fagelhauer, L.; Chakraborty, T.;
435 Neumann, B.; Werner, G.; Bender, J. K.; Stingl, K.; Nguyen, M.; Coppens, J.; Xavier, B.
436 B.; Malhotra-Kumar, S.; Westh, H.; Pinholt, M.; Anjum, M. F.; Duggett, N. A.; Kempf, I.;
437 Nykäsenoja, S.; Olkkola, S.; Wiczorek, K.; Amaro, A.; Clemente, L.; Mossong, J.;
438 Losch, S.; Ragimbeau, C.; Lund, O.; Aarestrup, F. M. ResFinder 4.0 for Predictions of
439 Phenotypes from Genotypes. *J. Antimicrob. Chemother.* **2020**, *75* (12), 3491–3500.
440 <https://doi.org/10.1093/jac/dkaa345>.
- 441 (20) Alcock, B. P.; Raphenya, A. R.; Lau, T. T. Y.; Tsang, K. K.; Bouchard, M.; Edalatmand,
442 A.; Huynh, W.; Nguyen, A.-L. V.; Cheng, A. A.; Liu, S.; Min, S. Y.; Miroshnichenko, A.;
443 Tran, H.-K.; Werfalli, R. E.; Nasir, J. A.; Oloni, M.; Speicher, D. J.; Florescu, A.; Singh,
444 B.; Faltyn, M.; Hernandez, A.; Koutoucheva; Sharma, A. N.; Bordeleau, E.; Pawlowski,
445 A. C.; Zubyk, H. L.; Dooley, D.; Griffiths, E.; Maguire, F.; Winsor, G. L.; Beiko, R. G.;
446 Brinkman, F. S. L.; Hsiao, W. W. L.; Domselaar, G. V.; McArthur, A. G. *CARD 2020:*
447 *Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance*
448 *Database*; 2020. <https://doi.org/10.1093/nar/gkz935>.
- 449 (21) *NCBI Antimicrobial Resistance Gene Finder (AMRFinderPlus)*; NCBI - National Center
450 for Biotechnology Information/NLM/NIH, 2021.
- 451 (22) *Staramr*; National Microbiology Laboratory, 2021.
- 452 (23) Panunzi, L. G. SraX: A Novel Comprehensive Resistome Analysis Tool. *Front.*
453 *Microbiol.* **2020**, *11*, 52. <https://doi.org/10.3389/fmicb.2020.00052>.
- 454 (24) Arango-Argoty, G.; Garner, E.; Pruden, A.; Heath, L. S.; Vikesland, P.; Zhang, L.
455 DeepARG: A Deep Learning Approach for Predicting Antibiotic Resistance Genes from
456 Metagenomic Data. *Microbiome* **2018**, *6* (1), 23. [https://doi.org/10.1186/s40168-018-](https://doi.org/10.1186/s40168-018-0401-z)
457 [0401-z](https://doi.org/10.1186/s40168-018-0401-z).
- 458 (25) PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance
459 associated with chromosomal point mutations in bacterial pathogens | Journal of
460 Antimicrobial Chemotherapy | Oxford Academic
461 <https://academic.oup.com/jac/article/72/10/2764/3979530?login=true> (accessed 2021 -10 -
462 12).

BENCHMARKING AMR SOFTWARE

20

- 463 (26) Gupta, S. K.; Padmanabhan, B. R.; Diene, S. M.; Lopez-Rojas, R.; Kempf, M.; Landraud,
464 L.; Rolain, J.-M. ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic
465 Resistance Genes in Bacterial Genomes. *Antimicrob. Agents Chemother.* **2014**, *58* (1),
466 212–220. <https://doi.org/10.1128/AAC.01310-13>.
- 467 (27) Doster, E.; Lakin, S. M.; Dean, C. J.; Wolfe, C.; Young, J. G.; Boucher, C.; Belk, K. E.;
468 Noyes, N. R.; Morley, P. S. MEGARes 2.0: A Database for Classification of
469 Antimicrobial Drug, Biocide and Metal Resistance Determinants in Metagenomic
470 Sequence Data. *Nucleic Acids Res.* **2020**, *48* (D1), D561–D569.
471 <https://doi.org/10.1093/nar/gkz1010>.
- 472 (28) Ingle, D. J.; Valcanis, M.; Kuzevski, A.; Tauschek, M.; Inouye, M.; Stinear, T.; Levine,
473 M. M.; Robins-Browne, R. M.; Holt, K. E. In Silico Serotyping of E. Coli from Short
474 Read Data Identifies Limited Novel O-Loci but Extensive Diversity of O:H Serotype
475 Combinations within and between Pathogenic Lineages. *Microb. Genomics* **2016**, *2* (7),
476 e000064. <https://doi.org/10.1099/mgen.0.000064>.
- 477 (29) Carattoli, A.; Hasman, H. PlasmidFinder and In Silico PMLST: Identification and Typing
478 of Plasmid Replicons in Whole-Genome Sequencing (WGS). *Methods Mol. Biol. Clifton*
479 *NJ* **2020**, *2075*, 285–294. https://doi.org/10.1007/978-1-4939-9877-7_20.
- 480 (30) Chen, L.; Zheng, D.; Liu, B.; Yang, J.; Jin, Q. VFDB 2016: Hierarchical and Refined
481 Dataset for Big Data Analysis—10 Years On. *Nucleic Acids Res.* **2016**, *44* (D1), D694–
482 D697. <https://doi.org/10.1093/nar/gkv1239>.
- 483 (31) *Escherichia Coli Virulence Factors*; National Microbiology Laboratory, 2021.
- 484 (32) McIver, L. J.; Abu-Ali, G.; Franzosa, E. A.; Schwager, R.; Morgan, X. C.; Waldron, L.;
485 Segata, N.; Huttenhower, C. BioBakery: A Meta’omic Analysis Environment.
486 *Bioinformatics* **2018**, *34* (7), 1235–1237. <https://doi.org/10.1093/bioinformatics/btx754>.
- 487 (33) Liu, B.; Pop, M. ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **2009**,
488 *37* (Database), D443–D447. <https://doi.org/10.1093/nar/gkn656>.
- 489 (34) Hyatt, D.; Chen, G.-L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J.
490 Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.
491 *BMC Bioinformatics* **2010**, *11* (1), 119. <https://doi.org/10.1186/1471-2105-11-119>.
- 492 (35) Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive Protein Alignments at Tree-of-Life Scale
493 Using DIAMOND. *Nat. Methods* **2021**, *18* (4), 366–368. <https://doi.org/10.1038/s41592-021-01101-x>.
- 494
495 (36) Zankari, E.; Hasman, H.; Cosentino, S.; Vestergaard, M.; Rasmussen, S.; Lund, O.;
496 Aarestrup, F. M.; Larsen, M. V. Identification of Acquired Antimicrobial Resistance
497 Genes. *J. Antimicrob. Chemother.* **2012**, *67* (11), 2640–2644.
498 <https://doi.org/10.1093/jac/dks261>.
- 499 (37) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden,
500 T. L. BLAST+: Architecture and Applications. *BMC Bioinformatics* **2009**, *10* (1), 421.
501 <https://doi.org/10.1186/1471-2105-10-421>.
- 502 (38) McGinnis, S.; Madden, T. L. BLAST: At the Core of a Powerful and Diverse Set of
503 Sequence Analysis Tools. *Nucleic Acids Res.* **2004**, *32* (suppl_2), W20–W25.
504 <https://doi.org/10.1093/nar/gkh435>.
- 505 (39) Reference Gene Catalog - Pathogen Detection - NCBI
506 <https://www.ncbi.nlm.nih.gov/pathogens/refgene/#> (accessed 2021 -09 -13).
- 507 (40) Anthony Underwood / chAMReDb <https://gitlab.com/antunderwood/chamredb> (accessed
508 2021 -10 -12).

BENCHMARKING AMR SOFTWARE

21

- 509 (41) Hastings, J.; Owen, G.; Dekker, A.; Ennis, M.; Kale, N.; Muthukrishnan, V.; Turner, S.;
510 Swainston, N.; Mendes, P.; Steinbeck, C. ChEBI in 2016: Improved Services and an
511 Expanding Collection of Metabolites. *Nucleic Acids Res.* **2016**, *44* (D1), D1214–D1219.
512 <https://doi.org/10.1093/nar/gkv1031>.
- 513 (42) Huttenhower, C.; Gevers, D.; Knight, R.; Abubucker, S.; Badger, J. H.; Chinwalla, A. T.;
514 Creasy, H. H.; Earl, A. M.; FitzGerald, M. G.; Fulton, R. S.; Giglio, M. G.; Hallsworth-
515 Pepin, K.; Lobos, E. A.; Madupu, R.; Magrini, V.; Martin, J. C.; Mitreva, M.; Muzny, D.
516 M.; Sodergren, E. J.; Versalovic, J.; Wollam, A. M.; Worley, K. C.; Wortman, J. R.;
517 Young, S. K.; Zeng, Q.; Aagaard, K. M.; Abolude, O. O.; Allen-Vercoe, E.; Alm, E. J.;
518 Alvarado, L.; Andersen, G. L.; Anderson, S.; Appelbaum, E.; Arachchi, H. M.; Armitage,
519 G.; Arze, C. A.; Ayvaz, T.; Baker, C. C.; Begg, L.; Belachew, T.; Bhonagiri, V.; Bihan,
520 M.; Blaser, M. J.; Bloom, T.; Bonazzi, V.; Paul Brooks, J.; Buck, G. A.; Buhay, C. J.;
521 Busam, D. A.; Campbell, J. L.; Canon, S. R.; Cantarel, B. L.; Chain, P. S. G.; Chen, I.-M.
522 A.; Chen, L.; Chhibba, S.; Chu, K.; Ciulla, D. M.; Clemente, J. C.; Clifton, S. W.; Conlan,
523 S.; Crabtree, J.; Cutting, M. A.; Davidovics, N. J.; Davis, C. C.; DeSantis, T. Z.; Deal, C.;
524 Delehaunty, K. D.; Dewhirst, F. E.; Deych, E.; Ding, Y.; Dooling, D. J.; Dugan, S. P.;
525 Michael Dunne, W.; Scott Durkin, A.; Edgar, R. C.; Erlich, R. L.; Farmer, C. N.; Farrell,
526 R. M.; Faust, K.; Feldgarden, M.; Felix, V. M.; Fisher, S.; Fodor, A. A.; Forney, L. J.;
527 Foster, L.; Di Francesco, V.; Friedman, J.; Friedrich, D. C.; Fronick, C. C.; Fulton, L. L.;
528 Gao, H.; Garcia, N.; Giannoukos, G.; Giblin, C.; Giovanni, M. Y.; Goldberg, J. M.; Goll,
529 J.; Gonzalez, A.; Griggs, A.; Gujja, S.; Kinder Haake, S.; Haas, B. J.; Hamilton, H. A.;
530 Harris, E. L.; Hepburn, T. A.; Herter, B.; Hoffmann, D. E.; Holder, M. E.; Howarth, C.;
531 Huang, K. H.; Huse, S. M.; Izard, J.; Jansson, J. K.; Jiang, H.; Jordan, C.; Joshi, V.;
532 Katancik, J. A.; Keitel, W. A.; Kelley, S. T.; Kells, C.; King, N. B.; Knights, D.; Kong, H.
533 H.; Koren, O.; Koren, S.; Kota, K. C.; Kovar, C. L.; Kyrpides, N. C.; La Rosa, P. S.; Lee,
534 S. L.; Lemon, K. P.; Lennon, N.; Lewis, C. M.; Lewis, L.; Ley, R. E.; Li, K.; Liolios, K.;
535 Liu, B.; Liu, Y.; Lo, C.-C.; Lozupone, C. A.; Dwayne Lunsford, R.; Madden, T.;
536 Mahurkar, A. A.; Mannon, P. J.; Mardis, E. R.; Markowitz, V. M.; Mavromatis, K.;
537 McCorrison, J. M.; McDonald, D.; McEwen, J.; McGuire, A. L.; McInnes, P.; Mehta, T.;
538 Mihindukulasuriya, K. A.; Miller, J. R.; Minx, P. J.; Newsham, I.; Nusbaum, C.;
539 O’Laughlin, M.; Orvis, J.; Pagani, I.; Palaniappan, K.; Patel, S. M.; Pearson, M.; Peterson,
540 J.; Podar, M.; Pohl, C.; Pollard, K. S.; Pop, M.; Priest, M. E.; Proctor, L. M.; Qin, X.;
541 Raes, J.; Ravel, J.; Reid, J. G.; Rho, M.; Rhodes, R.; Riehle, K. P.; Rivera, M. C.;
542 Rodriguez-Mueller, B.; Rogers, Y.-H.; Ross, M. C.; Russ, C.; Sanka, R. K.; Sankar, P.;
543 Fah Sathirapongsasuti, J.; Schloss, J. A.; Schloss, P. D.; Schmidt, T. M.; Scholz, M.;
544 Schriml, L.; Schubert, A. M.; Segata, N.; Segre, J. A.; Shannon, W. D.; Sharp, R. R.;
545 Sharpton, T. J.; Shenoy, N.; Sheth, N. U.; Simone, G. A.; Singh, I.; Smillie, C. S.; Sobel,
546 J. D.; Sommer, D. D.; Spicer, P.; Sutton, G. G.; Sykes, S. M.; Tabbaa, D. G.; Thiagarajan,
547 M.; Tomlinson, C. M.; Torralba, M.; Treangen, T. J.; Truty, R. M.; Vishnivetskaya, T. A.;
548 Walker, J.; Wang, L.; Wang, Z.; Ward, D. V.; Warren, W.; Watson, M. A.; Wellington,
549 C.; Wetterstrand, K. A.; White, J. R.; Wilczek-Boney, K.; Wu, Y.; Wylie, K. M.; Wylie,
550 T.; Yandava, C.; Ye, L.; Ye, Y.; Yooseph, S.; Youmans, B. P.; Zhang, L.; Zhou, Y.; Zhu,
551 Y.; Zoloth, L.; Zucker, J. D.; Birren, B. W.; Gibbs, R. A.; Highlander, S. K.; Methé, B. A.;
552 Nelson, K. E.; Petrosino, J. F.; Weinstock, G. M.; Wilson, R. K.; White, O.; The Human
553 Microbiome Project Consortium. Structure, Function and Diversity of the Healthy Human
554 Microbiome. *Nature* **2012**, *486* (7402), 207–214. <https://doi.org/10.1038/nature11234>.

BENCHMARKING AMR SOFTWARE

22

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

556 **Class**

557 AMR Genes detected by each tool across coverage levels, grouped into drug class to which the

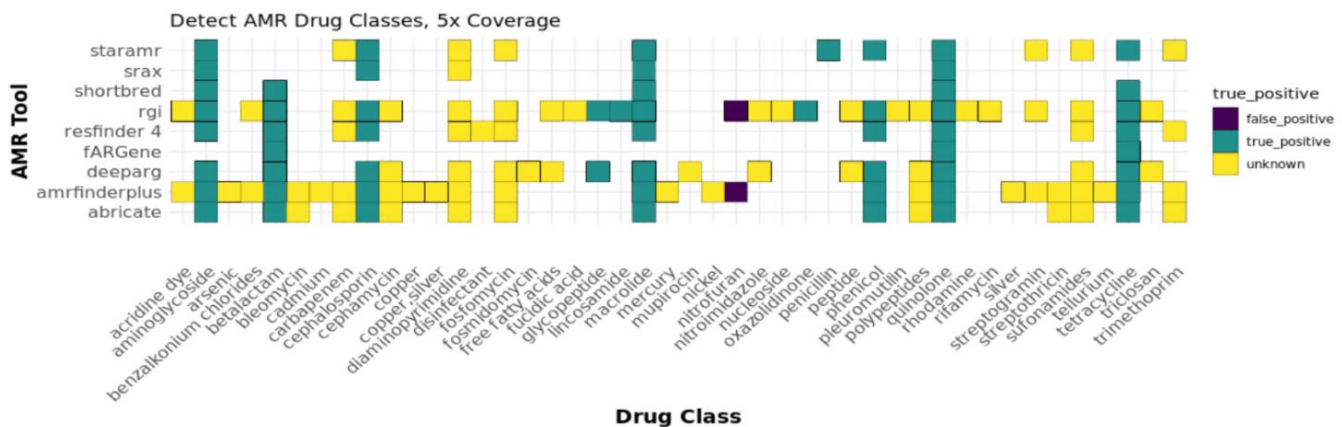
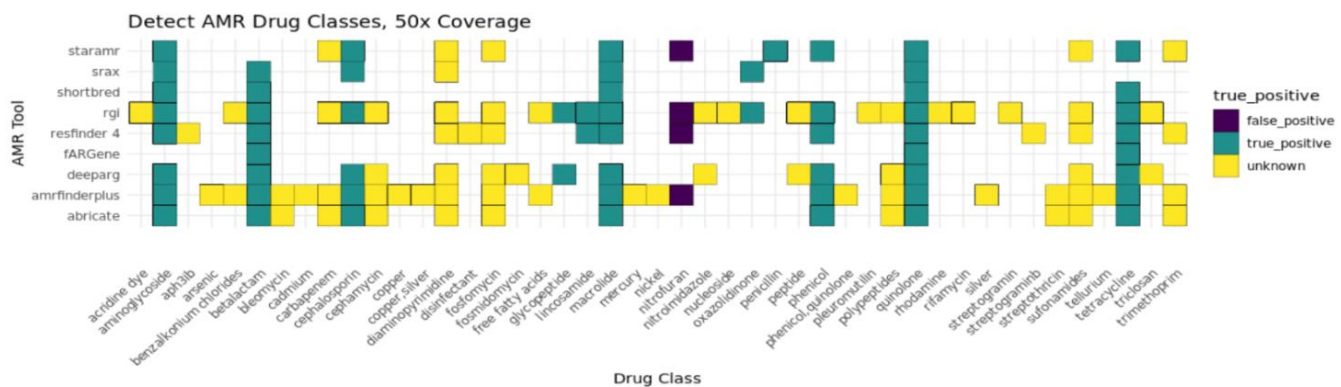
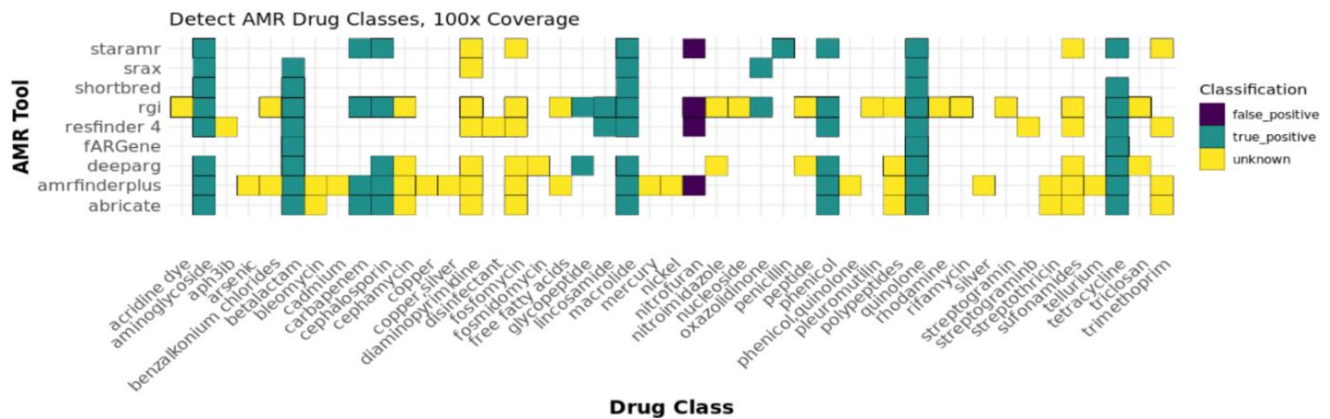
558 genes confer resistance with the color coding indicating whether the detection was true positive

BENCHMARKING AMR SOFTWARE

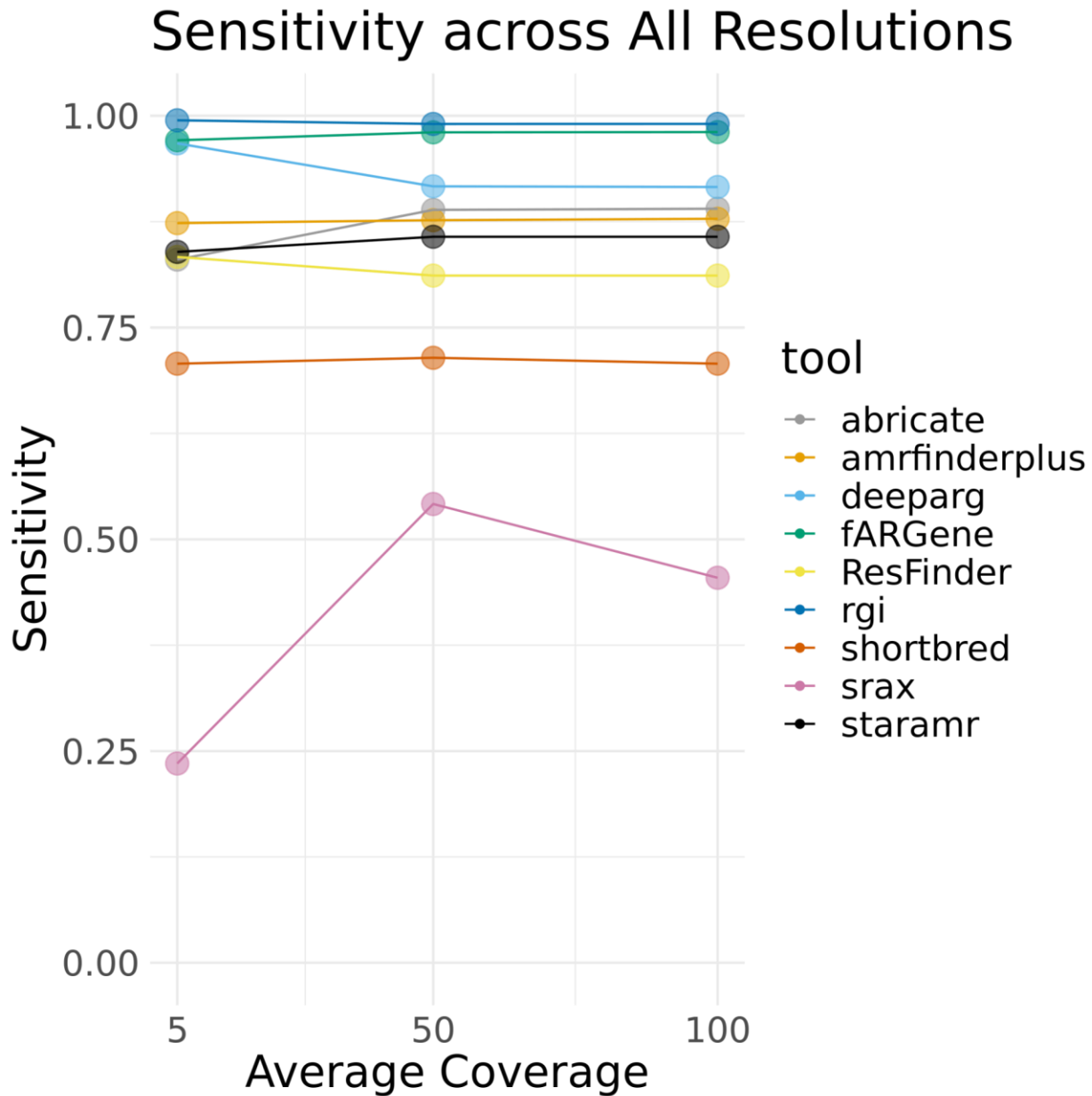
23

559 (green), false positive (purple) or unknown (yellow). Clear spaces in the plot indicate that AMR

560 genes were not detected for the drug class on the x-axis by the tool on the y-axis.



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



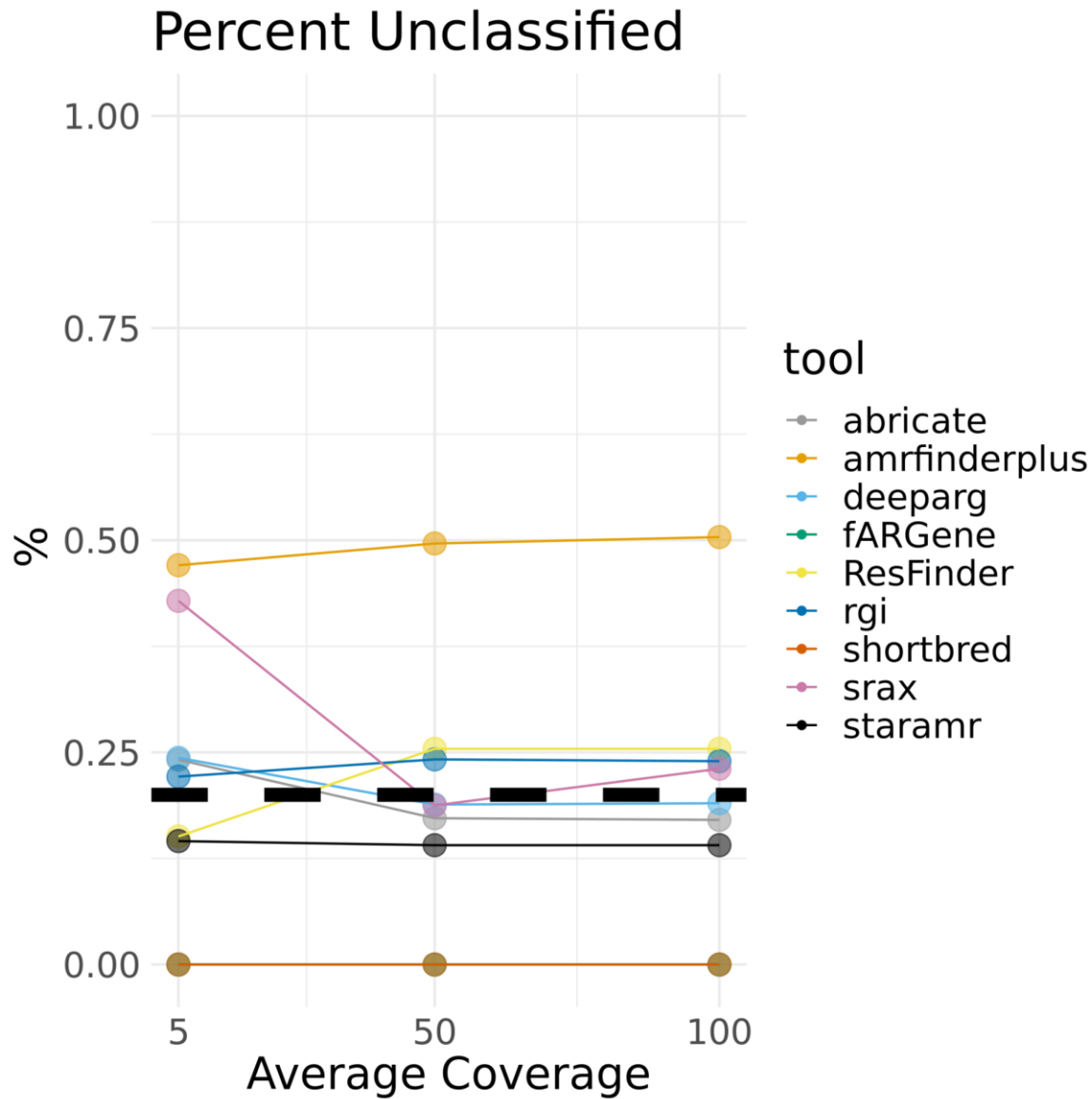
563
564 Sensitivity was calculated as (true positives) / (true positives + false negatives). Most tools were
565 highly sensitive (greater than 0.80). All genes corresponding to “Other” or “Unknown” drug
566 classes were not included in these calculations. Similarly, AMR genes corresponding to

BENCHMARKING AMR SOFTWARE

25

- 567 phenotypic resistance that was not tested in the mock community was considered “Unknown”
568 and not included in the sensitivity analysis.

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



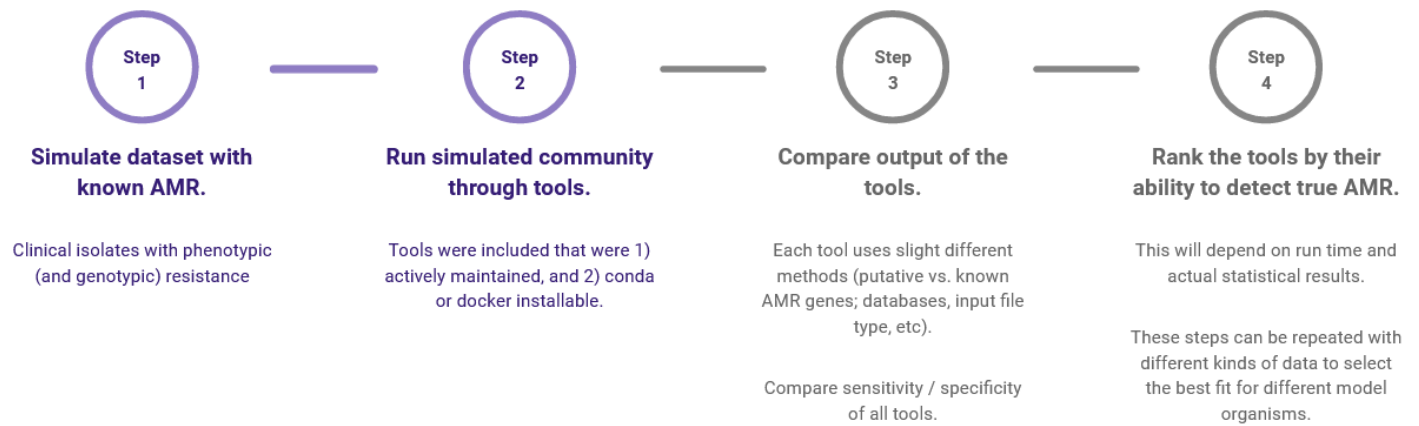
571

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

BENCHMARKING AMR SOFTWARE

27

575 **Supplementary Figure 1: Pictorial Methods**



576

577

BENCHMARKING AMR SOFTWARE

28

578

579

Table 1: Clinical isolates included in the simulated community. (susceptibility test is in

580

the spreadsheet, will have to be supplemental bc so big)

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

BENCHMARKING AMR SOFTWARE

29

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

581

582

BENCHMARKING AMR SOFTWARE

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

BENCHMARKING AMR SOFTWARE

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

BENCHMARKING AMR SOFTWARE

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	Require special metadata input	No		
MetaCherchant	Yes - conda	No	FASTA -	No		

BENCHMARKING AMR SOFTWARE

			genomic			
ARIBA	Yes - Docker	No	Paired end FASTQ	No		
ARG- ANNOT	No	No	Unclear	No		
kmerresista nce	No	No	-	No		
c-sstar	No	No	Unkno wn	No - could not track down github		

585

586

BENCHMARKING AMR SOFTWARE

34

587 **Table 3:** Summary Statistics from hAMRoaster: These are the counts and metrics as

588 calculated by the hAMRoaster pipeline. Formulas for all metrics are as follows:

589 $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$

590 $\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$

591 $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$

592 $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$

593 $\text{Recall} = \text{true pos} / (\text{true pos} + \text{false neg})$

594 $\text{F1} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$

595 $\text{Percent_unknown} = \text{unknown} / (\text{true_positives} + \text{false_positives} + \text{unknowns})$

596

597

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	0.2000	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	7.3333	0.2414
amrfinderplus	2	62	67	9	1	0.8732	0.3333	0.9688	0.8514	5.6364	0.5115
deeparg	0	99	23	8	2	0.9252	1.0000	1.0000	0.9266	12.3750	0.1885
fARGene	0	702	0	13	2	0.9818	1.0000	1.0000	0.9819	54.0000	0.0000
resfinder 4	1	43	15	9	1	0.8269	0.5000	0.9773	0.8148	4.3000	0.2542
rgi	4	557	254	6	1	0.9893	0.2000	0.9929	0.9824	55.7000	0.3117
shortbred	0	30	0	11	2	0.7317	1.0000	1.0000	0.7442	2.7273	0.0000
srax	0	13	3	10	2	0.5652	1.0000	1.0000	0.6000	1.3000	0.1875
staramr	1	52	11	9	1	0.8525	0.5000	0.9811	0.8413	5.2000	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

BENCHMARKING AMR SOFTWARE

38

599

600 **Table 4:** Condensed Summary Statistics: This table contains the counts and metrics if the

601 data were condensed so that overlapping genes are excluded from the count data (i.e. genes that

602 start between the start and stop codon of another gene are not considered in analysis).

603

604

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	0.98684210 53	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	0.96666666 67	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

BENCHMARKING AMR SOFTWARE

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

605

606

BENCHMARKING AMR SOFTWARE

42

607

608 **Table 5:** Kappa Values: Kappa values (agreement) between tools across coverage levels

609 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

610

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

611

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

612

613

614

BENCHMARKING AMR SOFTWARE

43

615

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

617

618

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

619

620

BENCHMARKING AMR SOFTWARE

44

621 Supplementary table 2: link to tweet

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

623

BENCHMARKING AMR SOFTWARE

45

624 Supplementary table 3: tidy table of data

625 <https://docs.google.com/spreadsheets/d/1bfACqEh0nkS65vCUj5DfMg4PvW0fHxbtrv0P>

626 [gKt1gT4/edit#gid=53644837](https://docs.google.com/spreadsheets/d/1bfACqEh0nkS65vCUj5DfMg4PvW0fHxbtrv0P/gKt1gT4/edit#gid=53644837)