Keeping up with the pathogens: Improved antimicrobial resistance detection and prediction in *Pseudomonas aeruginosa*

Danielle E. Madden[1,2], Timothy Baird[1,2,3], Scott C. Bell[4,5], Kate L. McCarthy[6,7], Erin P. Price[1,2*] and Derek S. Sarovich[1,2*#]

[1]Centre for Bioinnovation, University of the Sunshine Coast, Sippy Downs, Queensland, Australia

[2]Sunshine Coast Health Institute, Birtinya, Queensland, Australia

[3]Respiratory Department, Sunshine Coast University Hospital, Birtinya, Queensland, Australia

[4]Adult Cystic Fibrosis Centre, The Prince Charles Hospital, Chermside, Queensland, Australia

[5]Children's Health Research Centre, Faculty of Medicine, The University of Queensland, South Brisbane, Queensland, Australia

[6]University of Queensland Centre for Clinical Research, Herston, Queensland, Australia

[7]Royal Brisbane and Women's Hospital, Herston, Queensland, Australia

Keywords: Antibiotic, antimicrobial, resistance, AMR, whole-genome sequencing, high-throughput sequencing, *Pseudomonas aeruginosa*, bioinformatics, database, efflux pumps

Running title: Antibiotic resistance detection from *P. aeruginosa* genomes

#Corresponding author:

Derek S. Sarovich

dsarovich@usc.edu.au; +61 7 5456 3748

*Authors contributed equally

Abstract

Background: Antimicrobial resistance (AMR) is an intensifying threat that requires urgent mitigation to avoid a post-antibiotic era. The bacterial pathogen, *Pseudomonas aeruginosa,* represents one of the greatest AMR concerns due to ever-increasing multi- and pan-drug resistance rates. Shotgun sequencing is gaining traction for AMR profiling due to its unambiguity and data transferability; however, existing tools for *in silico* AMR prediction perform poorly for this pathogen.

Methods: We curated the most comprehensive database to date of known *P. aeruginosa* AMR determinants. We then performed comparative genomics and microbial genome-wide association (mGWAS) across a Global isolate Dataset (*n*=1391) with paired antimicrobial phenotype and genome data to identify novel AMR determinants. Finally, we assessed the predictive performance of our ARDaP-implemented database, along with AMRFinderPlus and ResFinder, for both the Global Dataset and an additional Validation Dataset of 89 isolates.

Findings: Our AMR database of 754 chromosomal variants includes 79 newly discovered variants associated with AMR phenotype/s, and 246 variants in AMR-encoding loci that do not confer AMR. Interrogation of the Global Dataset with our AMR database achieved genotype-phenotype concordance (as measured by balanced accuracy [bACC]) of 86%, compared with bACCs of 57% with AMRFinderPlus and 63% with ResFinder. For the Validation Dataset, bACCs of 80%, 54%, and 53%, respectively, were attained.

Interpretation: Our pipeline provides the most accurate tool yet for AMR prediction in *P. aeruginosa*. Implementing our ARDaP-compatible AMR determinant database for routine AMR prediction from *P. aeruginosa* (meta)genomes will vastly improve AMR identification and treatment modalities, addressing a critical facet in combatting this treatment-refractory pathogen.

Funding: This work was supported by Advance Queensland (AQIRF0362018 and AQRF13016-17RD2), the Wishlist Sunshine Coast Health Foundation (2019-14), and the National Health and Medical

49 Research Council (GNT455919). The funders had no role in study design, data acquisition, analysis,
50 interpretation, writing, or manuscript submission.

51

52 **Research in context**

53 Evidence before the study. Several bioinformatic tools have been developed with the purpose of
54 facilitating rapid and accurate AMR prediction from microbial (meta)genomic data, particularly for
55 superbugs. To this end, we searched PubMed on July 29, 2022 using ((("pseudomonas aeruginosa")
56 AND ("antimicrobial resistance")) AND ("genome")) AND ("predict*")) NOT (review) for studies
57 published in all languages. Twenty-four results were obtained, of which 20 described the use of one or
58 more bioinformatic methods to predict AMR determinants (8 x ResFinder, 6 x CARD, 4 x comparative
59 genomics, 4 x machine learning, 2 x transcriptomics, 1 x OpGen internal bioinformatic pipeline). Only 1
60 study created an AMR determinant database, which demonstrated limited ability to predict AMR
61 across a dataset of 100 *P. aeruginosa* isolates. No study has yet created a publicly available, open-
62 source tool and associated AMR determinant database for widespread use.

63 Added value of this study. Our ARDaP-compatible AMR determinant database, which was built from
64 the largest-ever paired *P. aeruginosa* AMR phenotype-genome dataset to date (*n*=1480 isolates),
65 provides a crucial advancement in the way we use WGS data to rapidly and accurately predict AMR in
66 *P. aeruginosa*. Using ARDaP, systematic analysis of large volumes of WGS data can be used to develop
67 a more detailed understanding of the specific genetic variants that underpin AMR.

68 Implications of all the available evidence. AMR prediction tools will play an essential role in the battle
69 against superbugs as they provide precise information on the genetic basis of a pathogen's resistance
70 profile – information that can be used to inform timely antimicrobial-treatment shifts, reduce the
71 reliance on broad-spectrum antibiotics, and ultimately improve patient outcomes and slow the spread
72 of AMR. Our study demonstrates that predicting AMR phenotype from genotype in *P. aeruginosa* can
73 rival the performance of phenotypic AMR sensitivity testing such as the VITEK® 2, which is routinely

74 used in clinical microbiology laboratories worldwide. The rapid detection of emerging AMR and

75 multidrug resistance, and subsequent implementation of personalised treatment regimens, serves as

76 our best chance to extend antibiotic efficacy amidst the AMR crisis.

77 **Introduction**

78 Antimicrobial-resistant (AMR) infections are predicted to cause 10 million deaths globally by 2050 and
79 cost USD$100 trillion per annum.[1] Antibiotic overuse and misuse[2] has driven the emergence of multi-
80 drug resistant (MDR) bacteria globally.[3] We are now on the verge of a 'post-antibiotic era', where
81 simple infections threaten to be untreatable with antimicrobials that once revolutionised modern
82 medicine.[4]

83 *Pseudomonas aeruginosa*, a leading cause of nosocomial infections and an important community-
84 acquired pathogen in people with chronic disease or immunosuppression, is associated with
85 significant health and economic burden.[5] Intrinsically resistant to many antibiotics, *P. aeruginosa* can
86 readily develop MDR via lateral gene transfer or chromosomal mutation.[6] This pathogen is one of the
87 most difficult bacterial species from which to accurately predict AMR phenotype from genotype,[7] even
88 using machine learning approaches,[8] with some current approaches as accurate as a coin toss.[7] Almost
89 all existing AMR prediction tools have focused on detecting AMR gene gain[9,10] and a small cache of
90 chromosomally encoded single-nucleotide polymorphisms (SNPs).[11-13] However, *P. aeruginosa* can also
91 evolve AMR through insertions-deletions (indels), loss-of-function mutations (e.g. large deletions or
92 frameshift mutations), and copy-number variations (CNVs).[14] Despite recent advances,[10,11] most AMR
93 tools remain limited in their scope and accuracy[15] – for example, loss-of-function mutations are largely
94 ignored,[13] AMR databases are often not species-specific,[13,16] resolution is limited to antibiotic class,[10]
95 and precursor mutations conferring reduced antimicrobial susceptibility are overlooked. These
96 limitations are especially problematic for accurate AMR detection in pathogens encoding complex
97 resistomes like *P. aeruginosa*.[7]

98 To address this knowledge gap, we curated and validated a comprehensive *P. aeruginosa*-specific AMR
99 determinant database, which, when used in conjunction with the Antimicrobial Resistance Detection
100 and Prediction (ARDaP) software,[13] enables high-accuracy AMR prediction from *P. aeruginosa*
101 (meta)genomes. Performance of our ARDaP-compatible AMR determinant database was first assessed
102 across 1391 geographically diverse *P. aeruginosa* strains ("Global Dataset"), and subsequently, across

103 89 analysis-naïve Australian *P. aeruginosa* strains ("Validation Dataset"). Our approach, which we
104 demonstrate exceeds the performance to current 'gold standard' AMR prediction software, provides a
105 crucial steppingstone towards the routine clinical use of (meta)genomics to inform personalised
106 treatment, such as in response to AMR evolution.

107 **Methods**

108 **Global Dataset.** To comprehensively capture geographic and phenotypic diversity, we collated a large
109 and comprehensive *P. aeruginosa* isolate collection (*n*=1391) by identifying all publicly available strains
110 with paired antimicrobial phenotype and whole-genome sequence (WGS) data (Tables S1 and S2, and
111 Figure S1). Ten clinically relevant antibiotics (amikacin, cefepime, ceftazidime, ciprofloxacin, colistin,
112 imipenem, meropenem, tobramycin, piperacillin, and piperacillin/tazobactam), representing six
113 antibiotic classes (aminoglycosides, β-lactams, carbapenems, cephalosporins, fluoroquinolones, and
114 polymyxins), were assessed (Table 1), including a substantial number of antimicrobial-susceptible
115 strains, which are essential for developing and refining high-quality AMR databases.[13,17]. Where
116 possible, isolates with minimum inhibitory concentration (MIC) data were reclassified as sensitive,
117 intermediate, or resistant using the CLSI M100S-Ed32:2022 guidelines.

118 **Validation Dataset.** To independently evaluate AMR prediction software performance, we examined a
119 collection of 89 phylogenetically diverse Australian clinical *P. aeruginosa* strains (Figure S2). Isolates
120 were obtained from people admitted to hospital with cystic fibrosis (CF [sputum]; *n*=42), bloodstream
121 infection (*n*=35), chronic obstructive pulmonary disease (COPD [sputum]; *n*=8), non-CF bronchiectasis
122 (sputum; *n*=1), ear infection (*n*=1), ulcer (*n*=1), or urinary tract infection (*n*=1), between 2008 and
123 2020. Eighty-four isolates have previously been subjected to WGS (NCBI BioProject PRJNA761496;
124 GenBank accessions NSXK00000000.1 and NSZK00000000.1) and some antimicrobial susceptibility
125 testing;[18,19] five additional COPD isolates were sequenced, and all isolates underwent antimicrobial
126 susceptibility testing across the 10 examined antibiotics (Table S3), for the current study. *P. aeruginosa*
127 strains PAO1 and LMG 6395 were included as antimicrobial-susceptible controls. Clinical specimen
128 collection was approved by The Prince Charles Hospital (HREC/13/QPCH/127 and
129 HREC/2019/QPCH/48013) and Royal Brisbane and Women's Hospital (DA: jl) Human Research Ethics
130 Committees. Participants with CF or COPD provided written consent; a waiver of informed consent
131 was approved for other specimens due to a low-to-negligible risk assessment.

132 **AMR profiling.** Antimicrobial susceptibility profiles for 8 of 10 clinically relevant antibiotics were
133 determined across the Validation Dataset (Table S3) using disc diffusions. Isolates were classed as
134 sensitive, intermediate or resistance using CLSI M100S-Ed32:2022 guidelines. Meropenem and
135 ciprofloxacin MICs were determined using ETESTs (bioMérieux, Murarrie, Australia) using sensitive,
136 intermediate, and resistant MIC cut-offs of ≤4, 8, and ≥16 µg/mL for meropenem, and ≤0.5, 1, and ≥2
137 µg/mL for ciprofloxacin.

138 ***P. aeruginosa* ARDaP database construction.** Gene names, locus tags and genomic coordinates in our
139 AMR determinant database were based on the archetypal genome, PAO1 (NC_002516.2). To identify
140 AMR determinants, a thorough literature search was conducted, and an ARDaP v1.9[13] compatible
141 SQLite database was populated. We then augmented this database by performing a microbial
142 genome-wide association study (mGWAS), along with a comprehensive comparative genomic analysis,
143 on the Global Dataset, to identify previously unidentified AMR determinants in *P. aeruginosa*. Our
144 AMR database also contains a modified version of the ResFinder database,[9] in which ubiquitous genes
145 in *P. aeruginosa* have been removed, and additional genes included, to reduce noise and improve
146 prediction accuracy. Detailed descriptions of each of these steps are described in the Supplemental
147 Methods. All additional scripts used in this study, including those used for mGWAS analysis, are
148 available at: github.com/dsarov/P_aeruginosa_ARDaP_manuscript.

149 **AMR prediction analysis.** AMR prediction in the Global and Validation Datasets was performed using
150 our *P. aeruginosa* AMR database (v1.0), implemented in ARDaP. ARDaP was chosen as it is the only
151 AMR software that can detect all mutation types (i.e. SNPs, indels, gene loss, frameshift mutations,
152 structural variants, and CNVs).[13] ARDaP performance was compared against three well-known, 'gold
153 standard' AMR detection tools: RGI v5.1.0 and CARD v3.0.9,[10] ResFinder v4.1,[9] and AMRFinderPlus
154 v3.8.28.[20] ARDaP software and the complete *P. aeruginosa* AMR database are available at:
155 github.com/dsarov/ARDaP.

156 **AMR software predictive performance in *P. aeruginosa*.** ARDaP, AMRFinderPlus, CARD, and
157 ResFinder predictions from isolate genomes were compared against their corresponding phenotypic
158 AMR profiles. Due to its inability to predict AMR towards individual antibiotics, and a very high rate of
159 false-positive predictions in the Global Dataset, CARD was deemed unsuitable for *P. aeruginosa* AMR
160 analysis and was therefore excluded from the Validation Dataset analysis. As most AMR prediction
161 software does not predict intermediate resistance, we excluded strains with this classification, and
162 removed intermediate AMR determinants from our ARDaP-compatible database, when performing
163 software comparisons. Predictive performance was determined using balanced accuracy (bACC),[21]
164 which averages sensitivity [i.e. true positives/(true positives + true negatives)], and specificity [i.e. true
165 negatives/(true negatives + false positives)]. This metric was chosen as it accounts for dataset
166 imbalance; that is, it minimises over- or under- representation of antimicrobial-sensitive or AMR
167 strains that may otherwise make certain tools appear better or worse due to dataset bias.

168 **Intermediate resistance prediction.** The capacity of our ARDaP-compatible AMR determinant
169 database to predict intermediate resistance phenotypes was examined in the Global Dataset. The
170 following rules were applied during software evaluation: 1) A true-positive prediction occurred when
171 either: i) an AMR strain was classed as AMR, or ii) an intermediate strain was classed as intermediate.
172 2) A true-negative prediction occurred when an antimicrobial sensitive strain was classed as
173 antimicrobial sensitive. 3) A false-positive prediction occurred when either: i) a sensitive strain was
174 classed as intermediate or AMR, or ii) an intermediate strain was classed as AMR. 4) A false-negative
175 prediction occurred when either: i) an intermediate strain was classed as sensitive, or when an AMR
176 strain was predicted to be antimicrobial sensitive or intermediate. These rules provided the strictest
177 evaluation criteria for the assessment of ARDaP's ability to identify intermediate strains.

178 **Results**

179 **Novel *P. aeruginosa* AMR determinant identification and refinement.** Among the 675 published
180 chromosomal AMR variants that we identified from the literature, we classified 246 (36%) as 'natural
181 variation' as they were common in both antimicrobial-sensitive and AMR strains, and thus not
182 associated with conferring an AMR phenotype towards any of the 10 tested antibiotics (Table S4).
183 Next, we identified novel AMR variants by examining all chromosomal variants occurring in known
184 AMR loci in the Global Dataset using mGWAS and comparative genomics. This analysis identified 79
185 previously unreported putative AMR determinants, each of which were significantly associated with
186 one or more AMR phenotypes (Table 2). In total, we identified 754 variants in known AMR loci, of
187 which 508 were confirmed or predicted to be associated with AMR. Upon retesting of our refined
188 ARDaP-compatible AMR database with the additional 79 novel AMR determinants, our database's
189 average bACC value across the 10 antibiotics increased from 82% to 86%.

190 **Predictive performance of AMR software across the Global Dataset.** The predictive performance of
191 AMRFinderPlus, CARD, and ResFinder was evaluated in the Global Dataset. Although superior to CARD,
192 ResFinder and AMRFinderPlus still showed relatively poor bACCs for most antibiotics. Of these two
193 tools, ResFinder was more accurate, with an average bACC of 63%, compared with AMRFinderPlus,
194 with an average bACC of just 57%. However, both tools performed well below ARDaP's average bACC
195 of 86% (Figure 1A).

196 The best ResFinder and AMRFinderPlus predictions were achieved for the aminoglycosides, with
197 average bACCs of 85% and 73%, respectively, although these rates were lower than ARDaP at an
198 average aminoglycoside bACC of 89%. AMR prediction for all other antibiotics using ResFinder and
199 AMRFinderPlus showed poor to very poor bACCs. AMRFinderPlus had a bACC of just 50% for the
200 penicillins, cephalosporins, and colistin – the same predictive capacity as a coin flip – and had only
201 slightly better bACCs for ciprofloxacin (53%) and carbapenems (61%). ResFinder also had a bACC of
202 just 50% for cephalosporins and colistin, and performed worse for carbapenems (average bACC of
203 56%) than AMRFinderPlus, although it was better for penicillins (average bACC of 71%) and

204 ciprofloxacin (bACC of 60%). In contrast, ARDaP surpassed ResFinder and AMRFinderPlus across the
205 penicillins, cephalosporins, carbapenems, ciprofloxacin, and colistin, with bACCs ranging from 70% for
206 colistin to 90% for ciprofloxacin (Figure 1A).

207 **Predictive performance of AMR software across the Validation Dataset.** We next tested
208 AMRFinderPlus, ResFinder, and our ARDaP-compatible AMR database across the Validation Dataset of
209 89 Australian clinical *P. aeruginosa* strains (Table S2) to determine each software's performance in an
210 analysis-naïve dataset. These strains exhibited similar AMR rates to the Global Dataset, ranging from
211 0% for colistin to 61% for piperacillin (Figure S2; Table S3). As no strains in the Validation Dataset
212 displayed colistin AMR, the bACC for this antibiotic could not be assessed.

213 Overall, our ARDaP-compatible database had high predictive accuracy (average bACC of 80%) across all
214 antibiotics (Figure 1B), vastly outperforming AMRFinderPlus (average bACC of 54%) and ResFinder
215 (average bACC of 53%), which, once again, were only marginally better in their predictive capacity
216 than a coin flip. Of note is the ResFinder meropenem bACC value, which, at just 45%, was worse than a
217 coin flip.

218 **ARDaP AMR prediction comparison between the Global and Validation Datasets.** Although the
219 ARDaP bACCs between the Validation and Global Datasets were similar, there was a greater
220 proportion of false-positive/false-negative determinants encoding piperacillin (19% difference),
221 amikacin and tobramycin (16% difference), cefepime (11% difference), and meropenem (8%
222 difference) AMR in the Validation Dataset. In contrast, there was a greater proportion of false-
223 positive/false-negative determinants encoding ciprofloxacin AMR (4% difference) in the Global Dataset
224 (Figure 1).

225 Comparative genomic analysis of Validation Dataset isolates with false-negative predictions for
226 aminoglycoside resistance identified clade-specific SNPs in all sequence type (ST) 801 isolates, which
227 occurred in elongation factor G (*fusA1*; D427G, S459F, G632A and R658H). Inclusion of these SNPs

228 significantly increased the ARDaP bACC values by 15% for both amikacin and tobramycin, raising them
229 to 87% and 94%, respectively.

230 **Precision and recall metrics among AMR software.** Our ARDaP-compatible database demonstrated
231 excellent average precision and recall for predicting antimicrobial-sensitive and AMR phenotypes,
232 ranging from 79% (average AMR recall) to 92% (average sensitivity precision and sensitivity recall)
233 (Figure S3A). In contrast, ResFinder's metrics ranged from only 49% (average sensitivity precision) to
234 65% (average AMR precision) (Figure S3B), and AMRFinderPlus's metrics ranged from 58% (average
235 sensitivity precision) to 67% (average sensitivity recall) (Figure S3C).

236 For colistin, AMRFinderPlus and ResFinder both yielded AMR precision and recall values of 0%; in
237 other words, neither of these tools identified colistin AMR determinants in strains exhibiting a colistin
238 AMR phenotype, whereas our ARDaP-compatible database identified colistin AMR strains with 41%
239 recall and 100% precision. Similarly, AMRFinderPlus and ResFinder failed to predict cefepime
240 sensitivity in any cefepime-sensitive strains, instead erroneously classing every strain as cefepime-
241 resistant, whereas ARDaP correctly identified cefepime-sensitive strains 91% precision and 93% recall
242 (Figure S3A). AMRFinderPlus also failed to identify piperacillin or piperacillin/tazobactam sensitivity in
243 any of the tested strains (Figure S3C).

244 **Predictive performance for intermediate resistance.** The inclusion of isolates with intermediate
245 resistance (i.e. strains with decreased susceptibility but not classed as full AMR; Table S5), reduced our
246 ARDaP-compatible database's bACC values by between 1 to 14%, depending on the antibiotic.
247 Intermediate resistance inclusion was most detrimental to cefepime AMR prediction (-14%), followed
248 by meropenem (-10%), ceftazidime, ciprofloxacin, and piperacillin (-9%), amikacin and
249 piperacillin/tazobactam (-7%), colistin (-5%), imipenem (-3%), and tobramycin (-1%).

250 **Discussion**

251 The increasing role of high-throughput sequencing in the delivery of clinical care has driven the
252 concomitant development of bioinformatic tools for AMR prediction.[22] However, despite much hope
253 and hype, almost all AMR prediction tools remain limited in their accuracy and performance, in large
254 part due to their inability to detect the gamut of variants that can confer AMR, along with their lack of
255 chromosomally encoded determinants.[13] This shortcoming becomes most evident when attempting to
256 predict AMR in pathogens with complex resistomes such as *P. aeruginosa*[7].

257 To address this issue, we constructed the most comprehensive database yet of known AMR
258 determinants in *P. aeruginosa*, which was augmented with 79 putative novel AMR determinants
259 identified using comparative genomics and mGWAS. In addition, we included 246 non-AMR-conferring
260 variants, which have been reported as potentially AMR-conferring in prior published studies. Whilst
261 the addition of these natural variants did not directly improve our database's predictive performance,
262 their inclusion was invaluable for reducing the legwork in identifying putative novel AMR variants, and
263 will help minimise erroneously reported putative novel AMR variants in future studies. We then tested
264 our *P. aeruginosa* AMR database across 1480 *P. aeruginosa* genomes using ARDaP, the only AMR
265 software that can detect all AMR-conferring genetic variants.[13] When measured across several
266 performance metrics (bACC, precision and recall), our ARDaP-implemented database far outstrips the
267 AMR predictive performance of 'gold standard' AMR prediction tools, AMRFinderPlus, CARD, and
268 ResFinder, across 10 clinically relevant antibiotics, and in both the Global and Validation Datasets
269 (Figure 1). Our findings concur with a prior study of 654 *P. aeruginosa* genomes, which also showed
270 that CARD and ResFinder exhibited poor performance metrics across all 11 tested anti-pseudomonal
271 antibiotics.[7]

272 Among the antibiotic classes, AMRFinderPlus and ResFinder performed best for aminoglycosides in the
273 Global Dataset (Figure 1A), which is heavily weighted towards American and European strains (Table
274 S2). However, these two tools performed substantially worse for the Australian Validation Dataset
275 (Figure 1B), with average aminoglycoside bACC values dropping by 15% for AMRFinderPlus and 27%

276 for ResFinder. The bACC was also lowered by 15% with our ARDaP-implemented database,
277 predominantly due to ST801 isolates calling as false-negatives. Inclusion of the novel *fusA1* SNPs
278 restored this performance to bACC values of 87% and 94% for amikacin and tobramycin, respectively.
279 This performance difference across isolate datasets can be attributed to two phenomena. The first is
280 the predominance of aminoglycoside-modifying enzymes in the Global (27% prevalence), but not
281 Validation (8% prevalence) Datasets, reflecting potential major differences in the geographic
282 prevalence of these enzymes that requires further exploration. The second is the enrichment of cystic
283 fibrosis-derived isolates, which comprise 93% of the aminoglycoside AMR strains, in the Validation
284 Dataset. These isolates have largely developed aminoglycoside AMR via chromosomal mutation rather
285 than aminoglycoside-modifying enzyme acquisition; as such, AMRFinderPlus and ResFinder had poor
286 aminoglycoside AMR predictive capacity in this dataset due to their limited chromosomal AMR variant
287 databases. These performance differences highlight the need for including isolates from diverse
288 sources, disease states, and locales to provide the most relevant AMR prediction software
289 benchmarking comparisons. Our results suggest that AMRFinderPlus and ResFinder are not useful for
290 predicting aminoglycoside AMR from cystic fibrosis-derived *P. aeruginosa*, although this finding
291 requires further exploration across larger and geographically diverse datasets.

292 Chromosomally-encoded loss-of-function mutations affecting the specialised porin, OprD, are the
293 most common cause of carbapenem AMR in *P. aeruginosa*, particularly in clinical isolates.[8] Although
294 *oprD* is notoriously hypervariable,[23] the ability to accurately identify functional OprD loss accounted
295 the high predictive accuracy for carbapenem AMR seen with ARDaP (average bACC of 89% vs. just 61%
296 for AMRFinderPlus and 52% for ResFinder). This outcome demonstrates the complex nature of the *P.*
297 *aeruginosa* resistome, the often-overlooked necessity of manual AMR variant curation efforts, and the
298 value of species-specific AMR variant databases to achieve the most accurate AMR predictions.

299 Mutations leading to chromosomal cephalosporinase (*ampC*) overexpression are an important cause
300 of β-lactam AMR, which includes the penicillins and cephalosporins.[14,24] To predict *ampC* upregulation,
301 our database includes function-altering mutations in genes known to regulate *ampC*, either directly or

302 indirectly. These chromosomally-encoded variants, alongside acquired cephalosporinases, accounted
303 for the high average ARDaP bACC observed for the cephalosporins (85% vs. just 50% with
304 AMRFinderPlus and ResFinder; Figure 1). The prominence of *ampC* over-expression provides further
305 support that this mechanism is a major cause of acquired cephalosporin AMR, particularly in clinical *P.*
306 *aeruginosa* isolates.[24,25] In support of this hypothesis, Khaledi and colleagues demonstrated a
307 considerably higher bACC for ceftazidime AMR prediction when using both transcriptomic and WGS
308 data (82%) compared with just WGS data (67%).[8] WGS alone therefore cannot currently identify all
309 instances of *ampC* over-expression, either because up-regulation is the result of an epigenic change, or
310 the mutation remains cryptic due to an incomplete understanding of *ampC* regulatory mechanisms.
311 Regardless, combining WGS alongside transcriptomic data will likely lead to improved AMR prediction
312 for most antibiotics.[8,24]

313 Predicting intermediate resistance is exceedingly difficult from WGS data alone, even with complex
314 machine learning algorithms that combine transcriptomic and WGS data.[8] We also encountered
315 difficulties with predicting intermediate resistance, with the inclusion of intermediate strains dropping
316 the bACC value by up to 14% (Table S5). Possible explanations for this prediction difficulty include the
317 need to understand the contribution of stepwise variants in conferring decreased antibiotic
318 susceptibility,[14,26,27] subtle and rapidly reversible gene expression alterations[8] caused by methylation[28]
319 or dynamic environmental stimuli, and undetected strain mixtures. Further refinement of our ARDaP-
320 compatible database, such as including capacity to interpret RNA-seq data, will continue to improve
321 this critical yet understudied area. Nevertheless, the pioneering capacity of ARDaP to predict
322 intermediate resistance, including stepwise mutations that lower the barrier to full AMR development,
323 has important implications for detecting emerging AMR in *P. aeruginosa* isolates. The eventual goal
324 will be to use this information to inform early treatment shifts to improve patient survival.[13]

325 We were unable to predict AMR with a 100% bACC value for any antibiotic. Errors introduced during
326 sample collection, metadata curation, specimen processing, or sequencing may partially account for
327 these discrepancies. For example, 11 strains in the Khaledi *et al*. dataset possessed determinants

328 known to confer ceftazidime AMR (e.g. $bla_{VIM-[2,4,47]}$, bla_{OXA-2}, $bla_{GES-[1,5]}$) yet were reported as
329 ceftazidime sensitive, and 22 strains in the Kos *et al*. dataset were amikacin sensitive, yet possessed
330 the aminoglycoside-modifying enzyme *aac*(6')-Ib-cr, known to cause amikacin AMR and reduced
331 ciprofloxacin susceptibility.[29] Due to the presence of these determinants, ARDaP, AMRFinderPlus, and
332 ResFinder all identified these strains as AMR, contributing to imperfect bACC values (Figure 1) and
333 poor precision (Figure S3) for amikacin and ceftazidime across all tools. As we did not have access to
334 these strains, it was not possible to retest their AMR phenotypes or to repeat WGS; however, we
335 hypothesise that these strains would generate different results upon phenotypic retesting.
336 Alternatively, these AMR determinants may be present in the genome but are functionally or
337 transcriptionally inactive, resulting in false-positive predictions for these isolates that must be factored
338 into AMR prediction estimates.

339 We recognise several study limitations. Some false-positive predictions were identified across all
340 antibiotic classes with our ARDaP-compatible database; however, these rates were significantly lower
341 than those reported by other software (Figure 1). Whilst not ideal, we have included some AMR
342 determinants that result in false-positive predictions at a relatively low frequency. These were not
343 removed from the final database because: i) false positives may occur as some strains revert to the
344 sensitive phenotype despite encoding a known AMR determinant, and ii) we relied on phenotypic data
345 provided by others, which may harbour inaccuracies. Our study would have benefitted from the
346 inclusion of transcriptomic data to augment our WGS data.[8] Although our understanding of the
347 molecular mechanisms of AMR in *P. aeruginosa* is improving rapidly in the genomics era, false-
348 negative predictions still occur across all antibiotic classes. Excitingly, these false-negative strains can
349 now be rapidly identified and targeted for future molecular work to identify novel AMR determinants
350 and mechanisms.

351

352 **Acknowledgements**

353 We wish to thank Jane Neill (Sunshine Coast Hospital and Health Service, Sunshine Coast University
354 Hospital) for providing the chronic obstructive pulmonary samples used in this study, and Kay Ramsay
355 (University of Queensland) and Tamieka Fraser (University of the Sunshine Coast) for laboratory
356 support.

357 **Author contributions**

358 DEM drafted the initial manuscript, compiled the original *P. aeruginosa* AMR determinant database,
359 performed data analyses and laboratory work, and assisted with ARDaP software refinement; TB
360 provided chronic obstructive pulmonary disease respiratory specimens that were used to retrieve *P.*
361 *aeruginosa* isolates, assisted with obtaining ethics approvals for sputum and bronchial washing
362 collection, and obtained study funding; SCB provided cystic fibrosis sputa that were used to retrieve *P.*
363 *aeruginosa* isolates, and obtained ethics approval for sputum collection; KM provided bloodstream *P.*
364 *aeruginosa* isolates, and obtained ethics approval for blood collection; EPP assisted with study
365 conception, obtaining and overseeing ethics approvals, manuscript preparation, data analyses and
366 laboratory work, provided student supervision, obtained study funding, and assisted with curation of
367 the natural variant database; and DSS conceived of the study, wrote the ARDaP software, refined the
368 *P. aeruginosa* AMR determinant database, assisted with manuscript preparation, data analyses and
369 laboratory work, provided student supervision, and obtained study funding. All authors read and
370 approved of the final manuscript.

371 **Conflicts of interest**

372 The authors declare no conflicts of interest.

373 **References**

374 1. O'Neill JC. Antimicrobial resistance: Tackling a crisis for the health and wealth of nations 2014
375 [Available from: https://amr-review.org/Publications.html.
376 2. Mancuso G, Midiri A, Gerace E, Biondo C. Bacterial antibiotic resistance: The most critical
377 pathogens. *Pathogens* 2021;**10**: 1310.
378 3. Bassetti M, Merelli M, Temperoni C, Astilean A. New antibiotics for bad bugs: where are we?
379 *Ann Clin Microbiol Antimicrob* 2013;**12**: 22.
380 4. Alanis AJ. Resistance to antibiotics: are we in the post-antibiotic era? *Arch Med Res* 2005;**36**:
381 697-705.
382 5. Zhen X, Lundborg CS, Sun X, Hu X, Dong H. Economic burden of antibiotic resistance in ESKAPE
383 organisms: a systematic review. *Antimicrob Resist Infect Control* 2019;**8**: 137.
384 6. Breidenstein EB, de la Fuente-Nunez C, Hancock RE. *Pseudomonas aeruginosa*: All roads lead
385 to resistance. *Trends Microbiol* 2011;**19**: 419-426.
386 7. Mahfouz N, Ferreira I, Beisken S, von Haeseler A, Posch AE. Large-scale assessment of
387 antimicrobial resistance marker databases for genetic phenotype prediction: a systematic review. *J*
388 *Antimicrob Chemother* 2020;**75**: 3099-3108.
389 8. Khaledi A, Weimann A, Schniederjans M, Asgari E, Kuo T-H, Oliver A, et al. Predicting
390 antimicrobial resistance in *Pseudomonas aeruginosa* with machine learning-enabled molecular
391 diagnostics. *EMBO Mol Med* 2020;**12**: e10264.
392 9. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for
393 predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;**75**: 3491-3500.
394 10. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020:
395 antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids*
396 *Res* 2019;**48**: D517-D525.
397 11. Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: Rapid
398 antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;**3**: e000131.
399 12. Zankari E, Allesoe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: A novel web
400 tool for WGS-based detection of antimicrobial resistance associated with chromosomal point
401 mutations in bacterial pathogens. *J Antimicrob Chemother* 2017;**72**: 2764-2768.
402 13. Madden DE, Webb JR, Steinig EJ, Currie BJ, Price EP, Sarovich DS. Taking the next-gen step:
403 Comprehensive antimicrobial resistance detection from *Burkholderia pseudomallei*. *EBioMedicine*
404 2021;**63**: 103152.
405 14. Lister PD, Wolter DJ, Hanson ND. Antibacterial-resistant *Pseudomonas aeruginosa:* Clinical
406 impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin Microbiol Rev*
407 2009;**22**: 582-610.
408 15. Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham CD, et al. Evaluation of
409 machine learning and rules-based approaches for predicting antimicrobial resistance profiles in Gram-
410 negative Bacilli from whole genome sequence data. *Front Microbiol* 2016;**7**: 1887.
411 16. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-
412 ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes.
413 *Antimicrob Agents Chemother* 2014;**58**: 212-220.
414 17. Su M, Satola SW, Read TD. Genome-based prediction of bacterial antibiotic resistance. *J Clin*
415 *Microbiol* 2019;**57**: e01405-18.
416 18. Madden DE, McCarthy KL, Bell SC, Olagoke O, Baird T, Neill J, et al. Rapid fluoroquinolone
417 resistance detection in *Pseudomonas aeruginosa* using mismatch amplification mutation assay-based
418 real-time PCR. *J Med Microbiol* 2021;**In Press**.
419 19. Madden DE, Olagoke O, Baird T, Neill J, Ramsay KA, Fraser TA, et al. Express yourself:
420 Quantitative real-time PCR assays for rapid chromosomal antimicrobial resistance detection in
421 *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 2022;**66**: e0020422.

422 20. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the
423 AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype
424 correlations in a collection of isolates. *Antimicrob Agents Chemother* 2019;**63**: e00483-19.
425 21. Bekkar M, Djemaa HK, Alitouche TA. Evaluation measures for models assessment over
426 imbalanced data sets. *J Inf Eng Appl* 2013;**3**: 27-38.
427 22. Van Goethem N, Descamps T, Devleesschauwer B, Roosens NHC, Boon NAM, Van Oyen H, et
428 al. Status and potential of bacterial genomics for public health practice: a scoping review. *Implement*
429 *Sci* 2019;**14**: 79.
430 23. Li H, Luo YF, Williams BJ, Blackwell TS, Xie CM. Structure and function of OprD protein in
431 *Pseudomonas aeruginosa*: from antibiotic resistance to novel therapies. *Int J Med Microbiol* 2012;**302**:
432 63-68.
433 24. Cabot G, Ocampo-Sosa AA, Tubau F, Macia MD, Rodriguez C, Moya B, et al. Overexpression of
434 *AmpC* and efflux pumps in *Pseudomonas aeruginosa* isolates from bloodstream infections: prevalence
435 and impact on resistance in a Spanish multicenter study. *Antimicrob Agents Chemother* 2011;**55**: 1906-
436 1911.
437 25. Lee JY, Ko KS. OprD mutations and inactivation, expression of efflux pumps and AmpC, and
438 metallo-β-lactamases in carbapenem-resistant *Pseudomonas aeruginosa* isolates from South Korea.
439 *Int J Antimicrob Agents* 2012;**40**: 168-172.
440 26. Levy SB, Bonnie M. Antibacterial resistance worldwide: Causes, challenges and responses.
441 *Nature Medicine* 2004;**10**: S122-S129.
442 27. Rehman A, Jeukens J, Levesque RC, Lamont IL. Gene-gene interactions dictate ciprofloxacin
443 resistance in *Pseudomonas aeruginosa* and facilitate prediction of resistance phenotype from genome
444 sequence data. *Antimicrob Agents Chemother* 2021;**65**: e0269620.
445 28. Law COK, Huang C, Pan Q, Lee J, Hao Q, Chan T-F, et al. A small RNA transforms the multidrug
446 resistance of *Pseudomonas aeruginosa* to drug susceptibility. *Mol Ther Nucleic Acids* 2019;**16**: 218-
447 228.
448 29. Robicsek A, Strahilevitz J, Jacoby GA, Macielag M, Abbanat D, Park CH, et al. Fluoroquinolone-
449 modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nat Med*
450 2006;**12**: 83-88.
451 30. Dötsch A, Becker T, Pommerenke C, Magnowska Z, Jänsch L, Häussler S. Genomewide
452 identification of genetic determinants of antimicrobial drug resistance in *Pseudomonas aeruginosa*.
453 *Antimicrob Agents Chemother* 2009;**53**: 2522-2531.

454

455 **Tables**

456 **Table 1. Summary of antimicrobial resistance prevalence across 10 clinically relevant antibiotics in**

457 **the *Pseudomonas aeruginosa* Global Dataset**

Antibiotic Class	Antibiotic	Res (%)	Int (%)	Sen (%)	No. isolates
Carbapenems	MEM	637 (46)	119 (9)	632 (46)	1388
	IPM	126 (76)	6 (4)	33 (20)	165
Polymyxins	CST	27 (4)	0 (0)	665 (96)	692
Fluoroquinolones	CIP	446 (61)	59 (8)	228 (31)	733
Cephalosporins	FEP	146 (33)	75 (17)	221 (50)	442
	CAZ	297 (43)	108 (16)	279 (41)	684
Penicillins	TZP	142 (52)	25 (9)	104 (38)	271
	PIP	179 (57)	23 (7)	113 (36)	315
Aminoglycosides	TOB	269 (38)	10 (1)	433 (61)	712
	AMK	166 (18)	52 (6)	721 (77)	939

458

459 Abbreviations: Res, resistant; Int, intermediate; Sen, sensitive; MEM, meropenem; IPM, imipenem;

460 CST, colistin; CIP, ciprofloxacin; FEP, cefepime; CAZ, ceftazidime, TZP, piperacillin/tazobactam; PIP,

461 piperacillin; TOB, tobramycin; AMK, amikacin

462 **Table 2. Novel antimicrobial resistance (AMR) determinants identified in *Pseudomonas aeruginosa***

463 **by microbial genome-wide association study (mGWAS) and comparative genomic analyses**

Locus	Type of mutation	Associated AMR phenotype	Fisher's Exact *p* value
PA0027	Lys34Asn	PIPr	< 0.0001
PA0295	Ala155Thr	MEMi	< 0.0001
PA0746	Lys106Asn	FEPr	< 0.0001
PA0799	Ala275Thr	PIPr	< 0.0001
PA0807 (*ampDh3*)	Ser44Arg	IPMr	0.5545
PA0946	Arg287Trp	CAZr	< 0.0001
PA0958 (*oprD*)	Gln340Pro	MEMr	1
PA0958 (*oprD*)	Leu136Gln	MEMr	1
PA0958 (*oprD*)	Leu409Pro	MEMr	1
PA1017 (*pauA*)	LOF	CAZr, MEMi	NA
PA1045	Lys322Arg	FEPr	< 0.0001
PA1760	LOF	MEMr, TZPi	NA
PA1798 (*parS*)	Ala149Thr	MEMr	< 0.0001
PA1798 (*parS*)	Arg383Cys	CSTr	0.039
PA1798 (*parS*)	Arg7His	PIPr	1
PA1798 (*parS*)	Asp249Asn	TZPr, AMKi	0.1399, 0.0003
PA1799 (*parR*)	Leu165Phe	MEMr	1
PA1799 (*parR*)	Val68Ala	CAZr	0.0309
PA1923	Ala350Val	TZPr	< 0.0001
*PA1942**	LOF	MEMr	NA
PA1997	Ile54Val	TZPr	< 0.0001
PA2020	Leu199Arg	FEPr	0.3978
PA2020	Ser198Ile	MEMr	1
PA2020	Thr32Asn	MEMr	0.4996
PA2064 (*pcoB*)	Gln59Lys	FEPr	< 0.0001
PA2252	LOF	TZPi	NA
PA2402	Val3040Ala	AMKr	< 0.0001
PA2519 (xylS)	Thr242Pro	AMKr	0.0349
PA2520 (*czcA*)	Gly1051Asp	CSTr	0.039
PA2567	Ile508Val	TZPr	< 0.0001
PA2727	Arg383Gln	TZPr	< 0.0001
PA2736	Arg30Gly	AMKr	< 0.0001
PA3083 (*pepN*)	LOF	MEMr, AMKr	NA
PA3384 (*phnC*)	LOF	MEMr, TOBr	NA
PA3418 (*ldh*)	LOF	MEMr	NA
PA3574 (*nalD*)	Arg164Pro	PIPr, FEPr	1, 0.3978
PA3574 (*nalD*)	Leu76Pro	MEMr	1
PA3721 (*nalC*)	Lys58Glu	MEMr	1
PA4020 (*mpl*)	Gly137Ser	CAZr	1
PA4020 (*mpl*)	Thr115Asn	MEMr, IPMr	0.4896, 1
PA4109 (*ampR*)	Asp135Gly	PIPr	0.2859

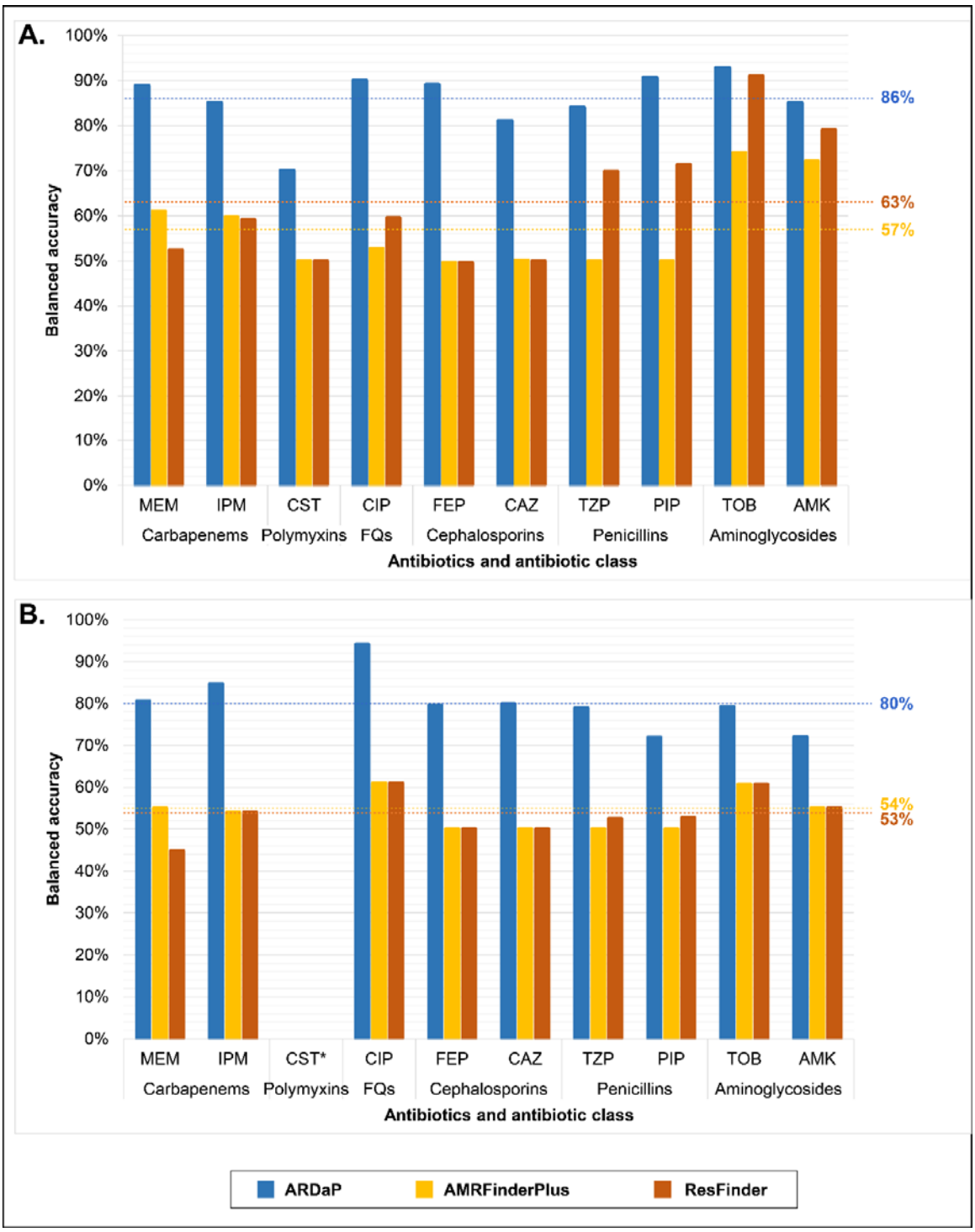
PA4109 (ampR)	Ter297Tyrext*?	PIPr	0.0006
PA4110 (*ampC*)	Asp272Asn	MEMr	1
PA4110 (*ampC*)	Leu320Pro	MEMr	1
PA4260 (*rplB*)	Gly138Asp	AMKr	0.0223
PA4260 (*rplB*)	Gly138Ser	TOBr, AMKr	1, 0.1871
PA4266 (*fusA1*)[#]	Arg658His	TOBr, AMKr	< 0.0001, < 0.0001
PA4296 (*pprB*)	Asn253His	CSTr	0.039
PA4318	Pro243Ser	PIPr	< 0.0001
PA4418 (*ftsI*)	Gly63Asp	MEMr, IPMr, CAZr, FEPr, TZPr, PIPr	1, 1, 1, 1, 1, 1
PA4464 (*ptsN*)	Pro10_Gly11dup	IPMr	1
PA4522 (*ampD*)	Gly100Glu	MEMr	1
PA4522 (*ampD*)	Gly116Val	MEMr	1
PA4522 (*ampD*)	Gly121Glu	MEMr	0.374
PA4522 (*ampD*)	Ile117Thr	CAZr	1
PA4522 (*ampD*)	Ile69Thr	MEMr	1
PA4522 (*ampD*)	Pro41Leu	PIPr	1
PA4522 (*ampD*)	Val10Gly	CAZr, FEPi	0.0114, 0.0003
PA4522 (*ampD*)	Pro41Leu	FEPr	0.3978
PA4522 (*ampD*)	Gly169Cys	FEPr	0.3978
PA4522 (*ampD*)	Cys110Gly	FEPr	0.1576
PA4522 (*ampD*)	Gly116Val	CAZr	1
PA4522 (*ampD*)	Gln88Leu	CAZr	0.4996
PA4522 (*ampD*)	His157Tyr	CAZr	0.2494
PA4531	Lys17Arg	PIPr	< 0.0001
PA4622	Ile380Val	FEPr	< 0.0001
PA4677	Asp138Val	FEPr	< 0.0001
PA4692	Ala180Val	PIPr	< 0.0001
PA4776 (*pmrA*)	Leu72Phe	CSTr	0.039
PA4777 (*pmrB*)	Arg155His	CSTr	0.039
PA4777 (*pmrB*)	Thr253Met	CSTr	0.039
PA4777 (pmrB)	Thr158Ile	CSTr	0.039
PA4777 (pmrB)	Arg10Leu	CSTr	0.039
PA4778 (*cueR*)	Gly57Ala	CSTr	0.039
PA5000 (*wapR*)	Gly163Asp	CSTr	0.039
PA5192 (*pckA*)	Ala461Thr	FEPr	< 0.0001
PA5199 (*amgS*)	Pro435Ala	CSTr	0.039
PA5199 (*amgS*)	Trp120Gly	CSTr	0.039
PA5338 (*spoT*)	Gly496Ser	CAZr	1
PA5493 (polA)	Lys395Arg	AMKr	< 0.0001

Abbreviations: AMK, amikacin; CAZ, ceftazidime; CIP, ciprofloxacin; CST, colistin; FEP, cefepime; i, intermediate resistant; IPM, imipenem; LOF, loss of function; MEM, meropenem; PIP, piperacillin; r, resistant; TOB, tobramycin; TZP, piperacillin/tazobactam

*Previously predicted to cause PIPr;[30] however, mGWAS did not identify a significant association with this phenotype. Instead, mGWAS showed that this AMR determinant was significantly associated with MEMr

[#]Identified in the Validation Dataset only.

471 **Figures**



472

473 **Figure 1. Balanced accuracy (bACC) of ARDaP, ResFinder, and AMRFinderPlus for antimicrobial**

474 **resistance (AMR) prediction in the (A) Global *Pseudomonas aeruginosa* Dataset (n=1391 isolates),**

475 **and (B) Validation *P. aeruginosa* Dataset (*n*=89 isolates), across ten clinically relevant antibiotics.** To

476 enable comparison with existing AMR prediction software, isolates with intermediate AMR were
477 removed prior to analysis. Abbreviations: AMK, amikacin; CAZ, ceftazidime; CIP, ciprofloxacin; CST,
478 colistin; FEP, cefepime; FQs, fluoroquinolones; IPM, imipenem; MEM, meropenem; PIP, piperacillin;
479 TZP, piperacillin/tazobactam; TOB, tobramycin. *No strains in the Validation Dataset exhibited CST
480 AMR; as such, bACC could not be calculated for this antibiotic.