Comprehensive genome data analysis establishes a triple whammy of carbapenemases, ICEs and multiple clinically-relevant bacteria

João Botelho[1#*], Joana Mourão[2,3,4], Adam P. Roberts[5,6], Luísa Peixe[1*]

[1]UCIBIO/REQUIMTE, Laboratório de Microbiologia, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal.

[2]University of Coimbra, Center for Innovative Biomedicine and Biotechnology, Coimbra, Portugal.

[3]University of Coimbra, Center for Neuroscience and Cell Biology, Coimbra, Portugal.

[4]University of Coimbra, Institute for Interdisciplinary Research, Coimbra, Portugal.

[5]Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom.

[6]Centre for Drugs and Diagnostics, Liverpool School of Tropical Medicine, Liverpool, United Kingdom.

[#]New affiliation: Antibiotic Resistance Evolution Group, Max-Planck-Institute for Evolutionary Biology, 24306 Plön, Germany; Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany.

*Addresses for correspondence: João Botelho, Antibiotic Resistance Evolution Group, Max-Planck-Institute for Evolutionary Biology, 24306 Plön, Germany; Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany.; e-mail: botelho@evolbio.mpg.de; Luísa Peixe, Laboratório de Microbiologia. Faculdade de Farmácia da Universidade do Porto, Rua Jorge Viterbo Ferreira nº 228, 4050-313 Porto, Portugal; phone: +351 220428500; e-mail: lpeixe@ff.up.pt.

Abstract

Carbapenemases inactivate most β-lactam antibiotics, including carbapenems and have been frequently reported among *Enterobacteriaceae*, *Acinetobacter* spp. and *Pseudomonas* spp. Traditionally, the horizontal gene transfer of carbapenemase-encoding genes (CEGs) has been linked to plasmids. However, given that integrative and conjugative elements (ICEs) are possibly the most abundant conjugative elements among prokaryotes, we conducted an *in-silico* analysis to ascertain the likely role of ICEs in the spread of CEGs among all bacterial genomes (n=182,663). We detected 17,520 CEGs, of which 66 were located within putative ICEs among several bacterial species (including clinically-relevant bacteria as *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* and *Escherichia coli*). Most CEGs detected within ICEs belong to the metallo-beta-lactamases IMP, NDM and SPM families and different mechanisms were likely responsible for acquisition of these genes. The majority of CEG-bearing ICEs belong to the MPF_G, MPF_T and MPF_F classes and often encode resistance to other antibiotics (e.g., aminoglycosides and fluoroquinolones). This study provides a snapshot of the different CEGs associated with ICEs among available bacterial genomes and sheds light on the underappreciated contribution of ICEs to the spread of carbapenem resistance globally.

Keywords

Integrative and conjugative elements, carbapenemases, antibiotic resistance.

Author Notes

All supporting data has been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article.

Abbreviations

AR, antibiotic resistance; CEG, carbapenemase-encoding gene; HGT, horizontal gene transfer; ICE, integrative and conjugative element; MGE, mobile genetic element; MPF, mating pair formation; NCBI, National Center for Biotechnology Information; T4SS, Type-IV Secretion System; WHO, World Health Organization.

Impact Statement

Carbapenems are commonly used to treat severe infections in humans. Resistance is often mediated by carbapenemases. These enzymes degrade carbapenems and are frequently present in plasmids. Here, we demonstrate that carbapenemase-encoding genes (CEGs) can also be located within integrative and conjugative elements (ICEs). CEG-bearing ICEs belong to three mating-pair formation families. These mobile elements may be particularly important in bacteria where plasmids do not seem to play a significant role in the spread of antibiotic resistance genes, as *Pseudomonas* spp. This study considerably expands the knowledge of the repertoire of CEGs-bearing ICEs

among clinically-relevant bacterial pathogens, such as *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* and *Escherichia coli*.

Data Summary

All the bacterial genomes scanned in this study have been deposited previously in the National Center for Biotechnology Information genome database and are listed on the supplementary tables. The extracted 66 ICEs in fasta format and the outputs for the profile HMMs scanned on the 386 putative MGEs identified in this study are deposited on figshare at https://figshare.com/projects/_Comprehensive_genome_data_analysis_establishes_a_triple_whammy_of_carbapenemases_ICEs_and_multiple_clinically-relevant_bacteria/78369.

Introduction

Due to the importance of carbapenems for the treatment of severe infections in humans, the WHO stated that these antibiotics should be reserved for infections caused by multidrug resistant Gram-negative bacteria in humans [1]. Recently, the same agency presented a list of bacterial pathogens for which research and development of new antibiotics are urgently required, and the top priority pathogens were the carbapenem-resistant strains of *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* [2].

92 The evolution of carbapenem-resistance in bacteria is often driven by the horizontal
93 gene transfer (HGT) of carbapenemase-encoding genes (CEGs) [3, 4]. Carbapenemases
94 are beta-lactamases able to hydrolyse carbapenems as well as most of other beta-
95 lactam antibiotics. These enzymes are members of the serine beta-lactamases classes A
96 and D, and the class B metallo-beta-lactamases [5]. The CEGs are often located on
97 integrons or transposons that themselves target mobile genetic elements (MGEs) such
98 as plasmids [3, 4], which makes the dissemination of these genes unpredictably complex
99 within bacterial communities. Recently, we demonstrated that another type of MGE,
100 the integrative and conjugative elements (ICEs), are likely to play a significant role as
101 vehicles for the dissemination of CEGs among *P. aeruginosa* [6]. Besides genes
102 conferring antibiotic resistance, ICEs may harbour additional cargo genes that provide
103 an adaptive advantage over other elements. Some of these examples include the
104 presence of the siderophore yersiniabactin encoded within the ICEKp in hypervirulent
105 clonal-group CG23 *Klebsiella pneumoniae* [7]; the Tn*5252*-related ICEs carrying
106 bacteriocin clusters in *Streptococcus suis* [8]; the type I-C CRISPR-Cas systems identified
107 within pKLC102-like ICEs in *P. aeruginosa* [9] and the type III restriction-modification
108 systems from SXT/R391-related ICEs in *Shewanella* spp. [10].

109 ICEs are self-transmissible MGEs that can integrate into and excise from the genome (as
110 transposons and phages do) and can exist as circular, sometimes replicable,
111 extrachromosomal elements and be transferred by conjugation (as some plasmids do)
112 [11–14]. ICEs appear to have a bipartite lifestyle that shifts between vertical and
113 horizontal transmission [15, 16]. HGT by conjugation requires three main components:
114 a relaxase, a mating pair formation (MPF) system, and a type-IV coupling protein, with
115 the last two forming a spanning-membrane multi-protein complex named type-IV

116 secretion system (T4SS) [17]. Until now, eight MPF classes were proposed (B, C, F, FA,
117 FATA, G, I and T), based on the phylogeny of VirB4, the only ubiquitous protein among
118 the T4SS. The MPF_T is widely distributed both in conjugative plasmids and ICEs, while
119 MPF_F is more prevalent in plasmids and MPF_G on ICEs [11].

120 Given that ICEs were identified in most bacterial clades and were proposed to be more
121 prevalent than conjugative plasmids [11], we conducted an *in silico* analysis to explore
122 the distribution of CEG-bearing ICEs among all sequenced bacterial genomes. Our results
123 demonstrate that CEG-bearing ICEs belong to three MPF families and are primarily
124 located in several clinically-relevant bacterial pathogens. Our analysis highlights the
125 importance of thoroughly investigating these elements as important vehicles for the
126 spread of antibiotic resistance (AR), particularly to carbapenems.

127

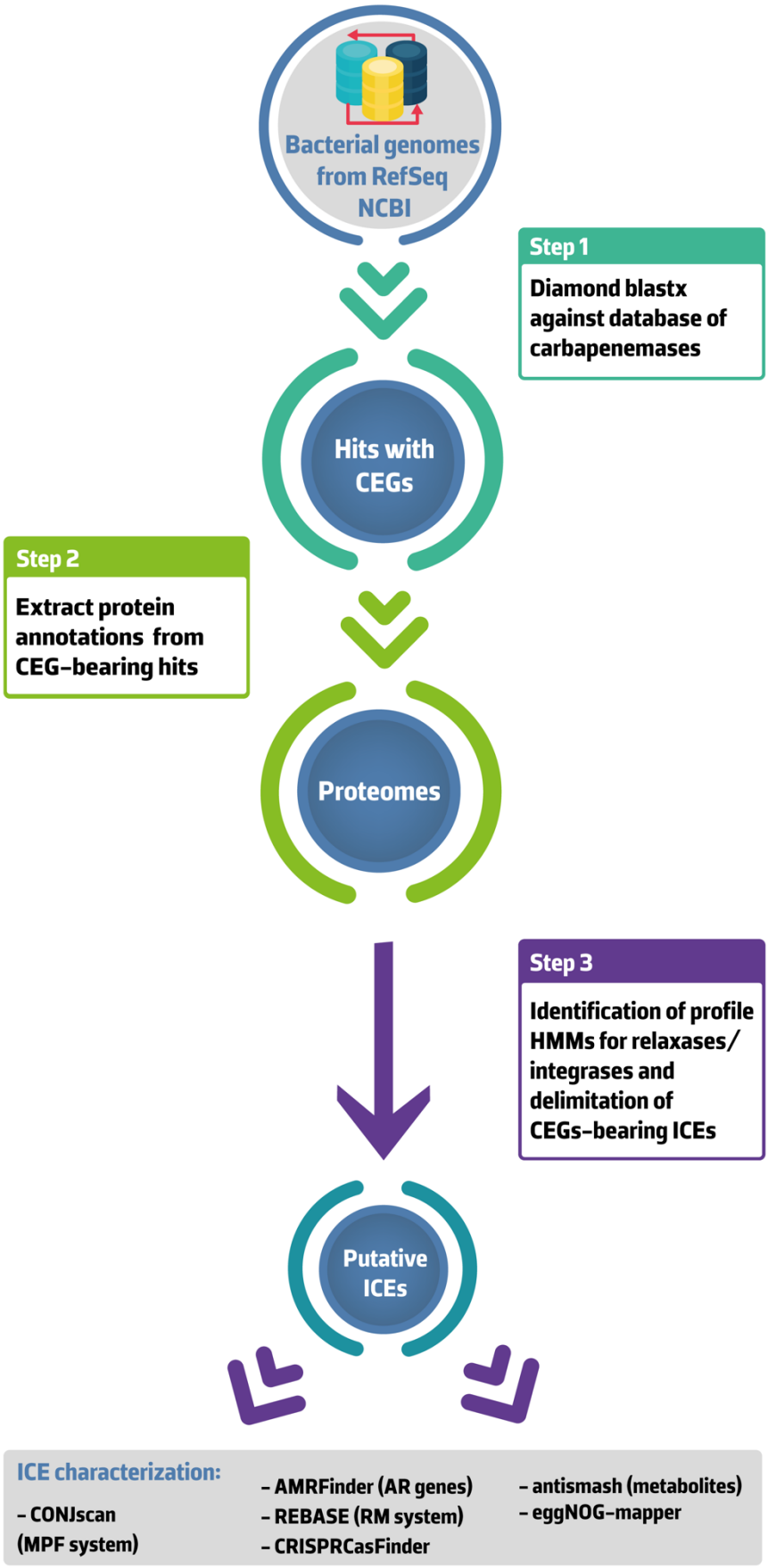
128 **Material and methods**

129 **Bacterial genomes and carbapenemases search**

130 In **Figure 1,** we present the workflow used in this study, from the acquisition of bacterial
131 genomes to the identification and characterization of putative ICEs. We retrieved all
132 bacterial genomes available in NCBI Reference Sequence Database (RefSeq, accessed on
133 21/03/2020), including complete and draft genome sequences, using ncbi-genome-
134 download v0.2.12 (https://github.com/kblin/ncbi-genome-download). We downloaded
135 over 6,000 curated AR protein sequences from AMRfinder database
136 (https://ftp.ncbi.nlm.nih.gov/pathogen/Antimicrobial_resistance/AMRFinderPlus/data
137 base/3.6/2020-01-22.1/) [18] and we built an in-house database only including the

138 proteins that code for a carbapenemase (n=1014, **Table S1**). We then blasted the
139 genomes against the extracted carbapenemases using diamond v0.9.29.130
140 (http://www.diamondsearch.org/index.php) [19], using minimum 100% identity and
141 subject cover and with the sensitive mode enabled.

142



Abbreviations: AR, Antibiotic Resistance; DB, Database; CEGs, carbapenemase-encoding genes; HMMs, Hidden Markov Models; ICEs, Integrative and Conjugative Elements; MLST, Multilocus Sequence Typing; MPF, Mating-Pair Formation; RM, Restriction-Modification

144 **Figure 1 –** Overview of the workflow followed in this study. All assemblies available in
145 NCBI RefSeq were downloaded and blasted against an in-house database of
146 carbapenemases using diamond blastx (Step 1). NCBI annotated proteins from CGE-
147 bearing genomes were then extracted (Step 2) and used for the identification of relaxase
148 and serine or tyrosine recombinase (Step 3). Search of directed repeats and delimitation
149 of putative ICEs was also performed. CONJscan was used to identify the MPF family of
150 each element. We then looked for AR genes, restriction-modification systems, CRISPR
151 arrays and their associated (Cas) proteins as well as secondary metabolites within
152 extracted ICEs. We also characterized the functional annotations of their proteomes and
153 the MLST of the genomes carrying a CEG-bearing ICE.

154

155 **Tracing ICEs among the bacterial genomes**

156 The RefSeq protein files from the CEG-bearing genomes identified by diamond were
157 extracted. We used the hmmsearch function of the HMMER3 software package v3.3
158 (http://hmmer.org/) [20] to search the proteomes against the standalone version of
159 MOBfamDB, a curated hidden Markov models (HMM) relaxase database
160 (https://castillo.dicom.unican.es/mobscan_about/) [21]. We also used this function to
161 search the pfam v33.0 database for serine or tyrosine recombinase accessions numbers
162 (Pfam IDs PF00589 and PF07508). The hmmsearch command was used with default
163 parameters and an E-value threshold of 0.01. The CEG-bearing genomes with relaxase
164 and integrase hits were further analysed. We manually inspected the hits to look for
165 direct repeats and to delimit CEG-harbouring ICEs. To predict the MPF families, the
166 translated coding sequences of delimited ICEs were analysed on the standalone

167 CONJscan module of MacSyFinder v1.0.5 (https://github.com/gem-
168 pasteur/macsyfinder) [22, 23]. To identify the multi-loci sequence type of the genomes
169 containing CEG-bearing ICEs, we used mlst v2.16.1
170 (https://github.com/tseemann/mlst), which scans the genomes against PubMLST typing
171 schemes (https://pubmlst.org/) [24].

172

173 **Characterization of the CEG-bearing ICEs**

174 Screening of AR genes among ICEs was performed using amrfinder v3.6.10
175 (https://github.com/ncbi/amr) [18]. The genetic platforms involved in the acquisition of
176 CEGs by ICEs were annotated using Galileo™ AMR
177 (https://galileoamr.arcbio.com/mara/) (Arc Bio, Cambridge, MA) [25]. We ran our
178 extracted ICEs against REBASE (http://rebase.neb.com/rebase/rebase.html) to look for
179 restriction-modification systems [26]. We also used CRISPRCasFinder
180 (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) to look for CRISPR
181 (clustered regularly interspaced short palindromic repeats) arrays and their associated
182 (Cas) proteins within ICE sequences [27]. Secondary metabolite biosynthetic gene
183 clusters were traced using antismash v5.1.2
184 (https://antismash.secondarymetabolites.org/#!/start) [28]. We used eggNOG-mapper
185 v2 (http://eggnog-mapper.embl.de/) for functional annotation based on orthology
186 assignments of the ICE proteomes [29].

187

188 **Results**

189 ## Carbapenemase-encoding genes are mainly found in Proteobacteria

190 We retrieved a total of 182,663 bacterial genomes from NCBI (16,798 complete
191 genomes and 165,865 genomes assembled at the chromosome, scaffold or contig level).
192 We identified a total of 17,520 CEGs, with 1,422 CEGs on 1,236 complete genomes
193 (including 512 chromosomes and 724 plasmids) and 16,098 CEGs on 16,038 draft
194 genomes (**Table S2**). We identified a total of 377 carbapenemase variants among the
195 17,520 hits. Our results show that CEGs are mostly located on Proteobacteria and
196 dominated by clinically-relevant pathogens as *A. baumannii*, *K. pneumoniae*, *Escherichia*
197 *coli* and *P. aeruginosa* (**Table S2**). These genomes encode a wide diversity of
198 carbapenemases, including OXA-23 (15.6%, n=2,739/17,520), KPC-2 (13.2%), KPC-3
199 (8.9%) and NDM-1 (6.7%). As we are tracing MGEs integrated in the chromosome, the
200 724 plasmid hits in complete genomes were excluded from the analysis. Among the hits
201 on the 16,038 draft genomes, we filtered out sequences with the word 'plasmid' present
202 on the fasta-header (n=131, **Table S3**). To maximise the chances of detecting entire ICEs,
203 we filtered out sequences shorter than 40kb (n=10,050). All the excluded ones are
204 available for analysis in **Table S4**. The remaining sequences from draft genomes
205 (n=5,857) were inspected for the presence of CEG-harbouring ICEs.

206

207 ## A large proportion of CEG-bearing ICEs belong to three families and target
208 ## clinically-relevant gram-negative bacteria

209 We identified a total of 66 ICEs, including 42 newly characterized elements associated
210 with 17 different CEGs (**Table 1**). We could predict the boundaries from 55 of these

211 elements. The terminal region of the remaining 11 putative ICEs could not be
212 determined due to a fragmented contig or assembly gaps within the sequence. Nearly
213 half of the putative ICEs (48.5%, n=32/66) were integrated at the 3' end of a $tRNA^{Gly}$
214 gene. The bacterial hosts housing these elements belong to 23 sequence types (STs)
215 (**Table S5**).

216 Using the CONJscan module of MacSyFinder we identified the MPF family for 62 hits (no
217 family could be determined for the remaining 4 hits) and we noted that the majority of
218 the putative ICEs belong to three families: MPF_G (69%, n=43/62), MPF_T (26%) and MPF_F
219 (5%) (**Figure 2**, **Table 1** and **Table S6**). ICEs of the MPF_G and MPF_T classes were
220 particularly promiscuous, being responsible for the spread of several CEGs of the
221 metallo-beta-lactamase family such as bla_{NDM-1}, bla_{SPM-1} and bla_{IMP} variants among
222 clinically-relevant pathogens as *P. aeruginosa*, *K. pneumoniae* and *E. coli*. ICEs of the
223 MPF_F class carrying blaIMP$_{-6}$ or bla_{KPC-2} were restricted to *E. coli* and *K. pneumoniae*. The
224 bla_{SPM-1} gene was exclusively identified in *P. aeruginosa* and in ICEs of the MPF_T class
225 (**Table 1**).

226 Additionally, we also identified 386 hits encoding an integrase and a relaxase in the
227 vicinity of CEGs (**Table S7**). For these hits, however, we could not predict if the CEG is
228 located on a plasmid or an ICE, since the contig is fragmented and tracing the boundaries
229 of the element is not possible, or the sequences have assembly gaps that make this
230 prediction challenging. Also, some plasmids may encode a tyrosine or serine
231 recombinase, and some ICEs may encode replicases and partition systems that are
232 typical of plasmids [30], which can hinder the accurate prediction of genetic platforms

233 when the sequence has poor quality or is highly fragmented due to short-read

234 sequencing approaches.

235 **Table 1 –** Diversity and characterization of carbapenemases-encoding genes in integrative and conjugative elements-associated genomes.

MPF family	**CEG**	**Bacterial Species**	**ICE genetic platforms**	**GC content (%)**	**Sequence length (kb)[1]**	**References**
MPF_G (n=43)	AFM-1 (n=1)	*Bordetella trematum*	Flanked by IS*91* family ISs	51.9	130	This study
	DIM-1 (n=1)	*Pseudomonas aeruginosa*	Class I In flanked by IS*6100*	62.8	89	This study, [6]
	GES-5 (n=5)	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	60.5-62.1	93-116	This study, [6]
	GES-24 (n=1)	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	63.0	63	This study
	IMP-1 (n=4)	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	61.3-62.3	76-109	This study, [6]
	IMP-13 (n=9)	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	61.5-62.0	65-89	This study, [6]

	IMP-14 (n=2)	*Achromobacter xylosoxidans*	Class I In within Tn*3*-like Tn	61.7-62.1	106-122	This study
	IMP-16 (n=1)	*Pseudomonas monteilii*	Class I In within Tn*3*-like Tn	62.6	86	This study
	IMP-54 (n=1)	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	61.7	91	This study
	KPC-2 (n=1)	*Pseudomonas aeruginosa*	Tn*4401*	64.2	115	This study, [31]
	NDM-1 (n=12)	*Pseudomonas aeruginosa, Pseudomonas asiatica, Morganella morganii*	Next to (or flanked by) IS*91* family ISs	58.7-64.5	97-167	This study, [32, 33]
	VIM-2 (n=1)	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	63.4	65	This study
	VIM-4 (n=4)	*Pseudomonas aeruginosa,*	Class I In within Tn*3*-like Tn	61.0-62.3	88-102	This study

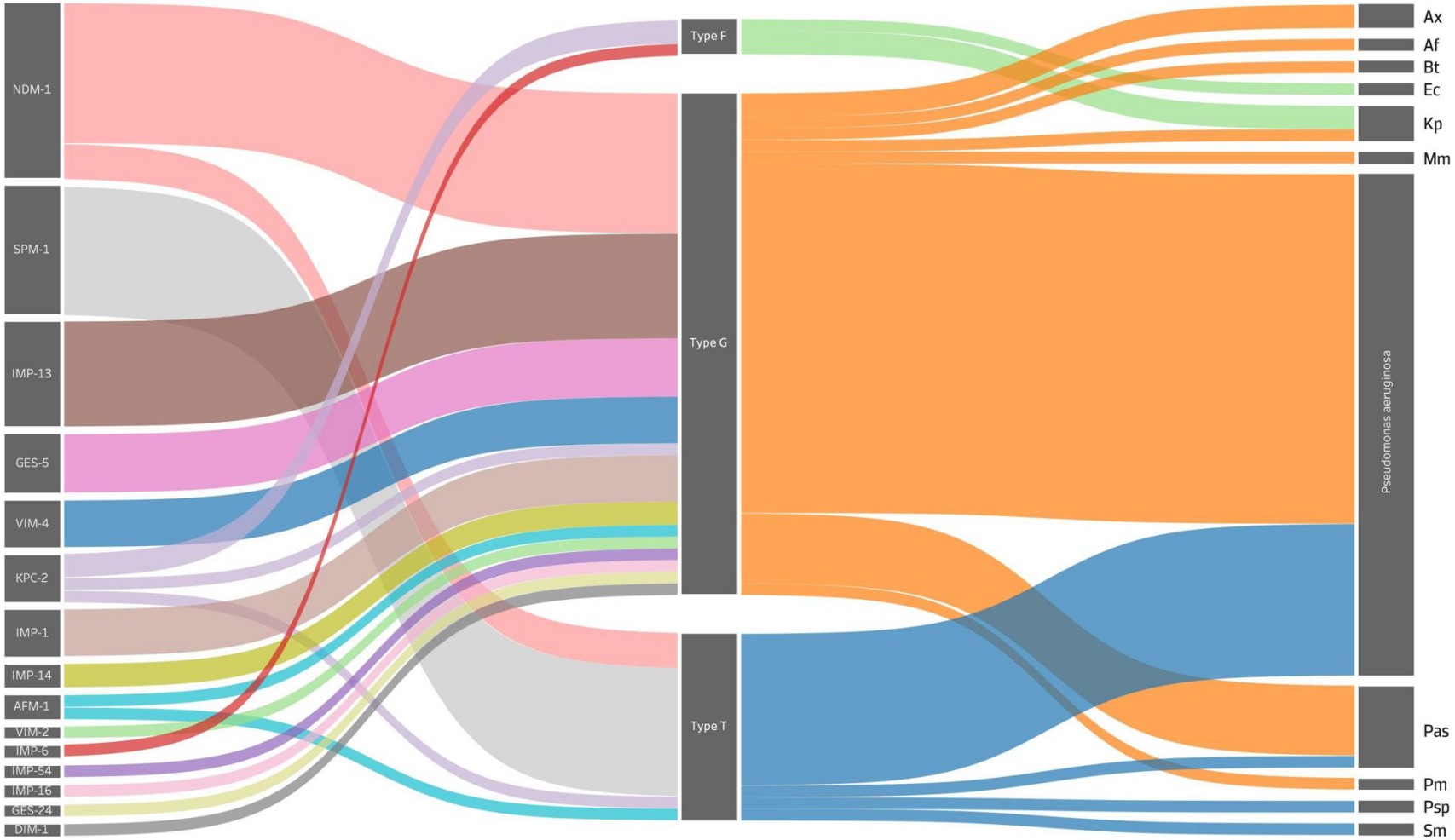
		Klebsiella pneumoniae, Alcaligenes faecalis				
MPF_T (n=16)	AFM-1 (n=1)	*Stenotrophomonas maltophilia*	Flanked by IS*91* family ISs	63.7	77	This study
	KPC-2 (n=1)	*Pseudomonas* sp.	Next to IS*Kpn6* and IS*26*	59.2	62	This study, [6]
	NDM-1 (n=3)	*Pseudomonas aeruginosa, Pseudomonas asiatica*	Flanked by IS*91* family ISs	63.6-64.6	73-74	This study, [34]
	SPM-1 (n=11)	*Pseudomonas aeruginosa*	Flanked by IS*CR3*-like elements	64.9-65.5	44-58	This study, [35]
MPF_F (n=3)	IMP-6 (n=1)	*Escherichia coli*	Class I In within Tn*3*-like Tn	51.9	118	This study
	KPC-2 (n=2)	*Klebsiella pneumoniae*	Tn*4401*	54.4	75	This study

Not assigned (n=4)	IMP-8 (n=1)	*Enterobacter cloacae*	Class I In within Tn*3*-like Tn	53.2	124	This study, [36]
	KPC-2 (n=1)	*Achromobacter* sp.	Next to IS*Kpn6* and IS*26*	60.4	43	This study
	NDM-1 (n=1)	*Pseudomonas aeruginosa*	Flanked by IS*CR3*-like elements	63.6	98	This study
	NDM-5 (n=1)	*Escherichia coli*	Flanked by IS*CR1* and IS*26*	52.4	54	This study, [37]

236 **Abbreviations:** CEG, carbapenemase-encoding gene; ICE, Integrative and Conjugative Elements; In, Integron; MPF, Mating Pair Formation; ST, Sequences
237 Type.

238 [1]Some sequences are not complete. For more details, please refer to **Table S5**.

239



Abbreviations: Ax, Achromobacter xylosoxidans; Af, Alcaligenes faecalis; Bt, Bordetella trematum; Ec, Escherichia coli, Kp, Klebsiella pneumoniae; Mm, Morganella morganii; Pas, Pseudomonas asiatica; Pm, Pseudomonas monteilii; Psp, Pseudomonas sp.; Sm, Stenotrophomonas maltophilia.

241 **Figure 2 –** Sankey diagram showing the contribution of different MPF families for the spread of CEGs among several bacterial genomes. The left,

242 centre and right axis represents the association between the identified carbapenemases, MPF type and the bacterial genus, respectively, while

243 the width of each connection is proportional to the number of positive hits.

244

245 **A variable repertoire of CEG-bearing integrons and transposons target**

246 **ICEs**

247 We identified 17 CEG variants among the 66 putative ICEs, dominated by *bla*$_{NDM-1}$ and

248 *bla*$_{SPM-1}$ (**Table 1**). Insertion sequences (ISs, e.g. ISCR3-like elements) were frequently

249 linked to the acquisition of *bla*$_{SPM-1}$ and *bla*$_{NDM-1}$ [34, 35], while *bla*$_{IMP}$, *bla*$_{VIM}$, and *bla*$_{GES}$

250 were found on class I integrons frequently integrated into Tn*3*-family transposons [6].

251 The *bla*$_{KPC-2}$ gene was typically found within Tn*4401*-like transposons, which are capable

252 of conferring a high frequency of transposition [38]. Curiously, the recently identified

253 AFM-1 metallo-beta-lactamase (Genbank accession number MK143105.1) was here

254 identified in two ICEs inserted in *Bordetella trematum* and *Stenotrophomonas*

255 *maltophilia* genomes (**Table 1**). We also found *bla*$_{NDM-1}$ genes in ICEs integrated into the

256 genomes of a recently proposed *Pseudomonas asiatica* species, which is spreading in

257 hospital settings in Myanmar [39]. Besides CEGs, the ICEs identified in this study also

258 harbour genes conferring resistance to other antibiotics, such as aminoglycosides,

259 fluoroquinolones, macrolides, and tetracyclines (**Table S8**), widening the spectrum of

260 transmissible AR genes selectable by carbapenems due to linkage.

261

262 **Acquisition of additional traits by ICEs include competitive weapons as**

263 **bacteriocins and siderophores**

264 Besides genes conferring antibiotic resistance, the CEG-bearing ICEs here identified

265 harbour other cargo genes that may confer a selective advantage to the ICE host. We

266 found DUF692 domains typical of bacteriocin production genes among the 6 MPF_G ICEs
267 from *P. asiatica* strains and the *P. aeruginosa* N15-01092, 1334/14 and ST773 strains
268 (**Table S5**). All these ICEs carry a bla_{NDM-1} gene and a similar copy of the bacteriocin-
269 encoding gene. Curiously, the bacteriocin production gene was not identified in the
270 MPF_T ICE from *P. asiatica* strain MY569. Additionally, we found the siderophore
271 aerobactin operon within an ICE in *E. coli* strain E302 (**Table S5**). This operon is usually
272 found in enterobacterial plasmids, and was also identified in a pathogenicity island in
273 uropathogenic *E. coli* strain CFT073 [40]. We identified no CRISPR-Cas systems nor
274 restriction-modification systems within the 66 putative ICEs.

275 The majority of the proteins encoded within the 66 ICEs refer to replication,
276 recombination, transcription and intracellular trafficking functions (**Figure 3**). Several
277 proteins, however, encoded for unknown functions (34.0%, n=1,615/4,754, **Table S9**),
278 highlighting the lack of knowledge on the ICE proteome.

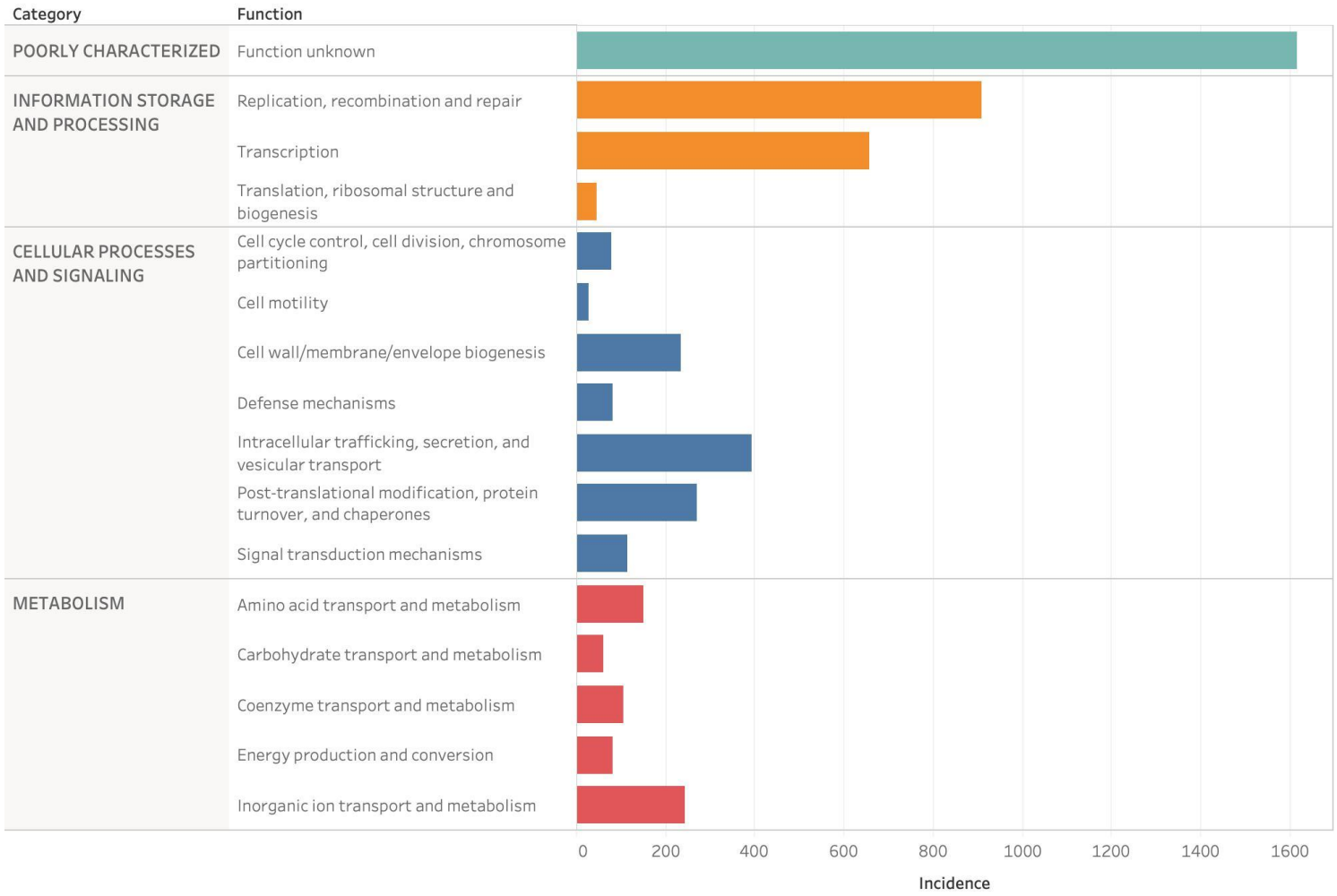


Figure 3 – Grouped bar chart representing the incidence of each eggNOG function broken down by category (represented by different colours) in the CEG-bearing ICE sequences.

282 ## Discussion

283 We have set out to comprehensively identify the CEGs among all bacterial genomes
284 deposited in NCBI and the CEG-bearing ICE sequences. Our study considerably expands
285 the knowledge of the repertoire of CEGs-bearing ICEs. We uncovered 66 putative ICEs
286 that may be involved in HGT of CEGs amongst bacterial genomes. To expand our
287 predictions, we also used the CONJscan module of Macsyfinder to trace the MPF families
288 likely involved in HGT.

289 The scenario observed for the acquisition of the most important CEGs by ICEs (ISs for
290 bla_{NDM-1} and bla_{SPM-1}, class I integrons for bla_{IMP} and bla_{VIM} and Tn*4401* for bla_{KPC-2})
291 resembles that of plasmids [3] and provide additional support to the notion that the line
292 separating these elements is blurred [13, 14]. We now show that besides plasmids, this
293 promiscuous repertoire of integrons and transposons frequently targets ICEs of different
294 MPF families. Even though CEGs might rapidly spread worldwide, local selection is likely
295 required for them to reach fixation, as can be seen for the clonal expansion of *P.*
296 *aeruginosa* ST277 harbouring bla_{SPM-1} [35]. Surprisingly, we noted that this gene was
297 unable to be detected beyond the same clonal lineage. Indeed, all hits were identified
298 in *P. aeruginosa* ST277 strains from Brazil and within a MPF_T ICE family, indicating that
299 these ICEs are only transferring vertically and/or horizontally within this ST.
300 Understanding the limitations on HGT of this family of ICE may be translatable to other,
301 more transferable, ICEs and could underpin a control strategy to prevent the spread of
302 these elements in the future.

303 Although bla_{KPC} and bla_{OXA} were the most frequently identified CEGs within the analysed
304 genomes (**Table S2**), we only found 5 ICEs carrying bla_{KPC-2} and 2 ICEs with bla_{OXA} genes

305 (**Table S8**). The *bla*$_{KPC}$ genes are mostly located in Tn*4401* transposons that target *K.*
306 *pneumoniae* plasmids, while the *bla*$_{OXA}$ genes are frequently associated with ISs that
307 tend to target the *Acinetobacter* spp. chromosomes and plasmids [3]. These genes may
308 take advantage of the copy number and the higher genetic plasticity of plasmids [30].
309 This plasticity may increase the rate of appearance of novel mutations and the high copy
310 number may amplify the effect due to the increased gene dosage [41].

311 In addition to AR, ICEs can be involved in other adaptive traits such as carbon-source
312 utilization, symbiosis, restriction-modification, siderophore and bacteriocin synthesis
313 [14]. Since bacteria commonly inhabit highly competitive environments, the production
314 of specific secondary metabolites (such as bacteriocins and siderophores) may confer a
315 selective advantage to the host [42]. We speculate that the presence of these
316 metabolites within the ICEs here characterized may promote their stability by
317 preferentially selecting for cells harbouring the ICE.

318 One caveat of our studies is that these results do not yet expose the complete set of
319 CEG-bearing ICEs present in all bacteria. There is an inherent bias in the number of times
320 we detect a particular CEG in certain pathogen genomes as some are over represented
321 in the database compared to others. It is possible that certain observations will flatten
322 out as more genomes are analysed. Less than 10% of the bacterial genomes currently
323 present on NCBI are complete. This is a major drawback since putative ICEs present in
324 draft genomes tend to be fragmented due to the presence of repetitive regions that are
325 not resolved using short-read sequencing. Also, the relaxase and the T4SS encoded by
326 ICEs resemble those of plasmids [11, 13, 43]. Plus, it is possible that ICEs and plasmids
327 have swapped conjugation modules throughout their evolutionary history [11]. We

328 believe that a more thorough exploration of this issue, especially regarding the precise
329 delimitation of ICEs, will be an important further step toward an improved
330 understanding of the contribution of these elements to bacterial adaptation and
331 evolution of AR.

332 While we have chosen to focus on CEG-bearing genomes, our computational approach
333 can be applied to trace other relevant AR genes and other cargo genes that may confer
334 a selective advantage to the ICE host. Leveraging knowledge linking the accurate
335 prediction of ICE sequences to the carriage of AR genes, will not only improve our
336 understanding of HGT, but may also uncover potential approaches to tackle the spread
337 of AR.

338

339 Data bibliography

340 The DNA sequences used in this study are available in NCBI, under the accession
341 numbers provided in supplementary tables S1-S9.

342

343 Funding information

344 This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is
345 financed by national funds from FCT (UIDB/04378/2020).

346

347 Author contributions

348 J. B. conceptualized the project. J. B. ran the analyses. J.B. and J. M. analysed the data.
349 J.B. wrote the manuscript. J. M., A. P. R. and L. P. edited and revised the manuscript. All
350 authors read, commented on and approved the final manuscript.

351

352 Conflicts of interest

353 The authors declare that there are no conflicts of interest.

354

355 References

356 1. EFSA Panel on Biological Hazards (BIOHAZ). Scientific Opinion on Carbapenem resistance
357 in food animal ecosystems. *EFSA J* 2013; **11**.

358 2. World Health Organization. GLOBAL PRIORITY LIST OF ANTIBIOTIC-RESISTANT BACTERIA
359 TO GUIDE RESEARCH, DISCOVERY, AND DEVELOPMENT OF NEW ANTIBIOTICS.
360 http://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-
361 which-new-antibiotics-are-urgently-needed. Accessed 11 Aug 2018.

362 3. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with
363 Antimicrobial Resistance. *Clin Microbiol Rev* 2018; **31**: e00088-17.

364 4. Botelho J, Grosso F, Peixe L. Antibiotic resistance in Pseudomonas aeruginosa –
365 Mechanisms, epidemiology and evolution. *Drug Resist Updat* 2019; **44**: 100640.

366 5. Queenan AM, Bush K. Carbapenemases: the versatile beta-lactamases. *Clin Microbiol Rev*
367 2007; **20**: 440–58, table of contents.

368 6. Botelho J, Roberts AP, León-Sampedro R, Grosso F, Peixe L. Carbapenemases on the
369 move: it's good to be on ICEs. *Mob DNA* 2018; **9**: 37.

370 7. Lam MMC, Wyres KL, Duchêne S, Wick RR, Judd LM, Gan YH, et al. Population genomics
371 of hypervirulent Klebsiella pneumoniae clonal-group 23 reveals early emergence and
372 rapid global dissemination. *Nat Commun* 2018; **9**.

373 8. Libante V, Nombre Y, Coluzzi C, Staub J, Guédon G, Gottschalk M, et al. Chromosomal
374 Conjugative and Mobilizable Elements in Streptococcus suis: Major Actors in the
375 Spreading of Antimicrobial Resistance and Bacteriocin Synthesis Genes. *Pathog (Basel,*
376 *Switzerland)* 2019; **9**.

377 9. van Belkum A, Soriaga LB, LaFave MC, Akella S, Veyrieras J-B, Barbu EM, et al.
378 Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant
379 Pseudomonas aeruginosa. *MBio* 2015; **6**: e01796-15.

380 10. Fang Y, Wang Y, Li Z, Liu Z, Li X, Diao B, et al. Distribution and Genetic Characteristics of
381 SXT/R391 Integrative Conjugative Elements in Shewanella spp. From China. *Front*
382 *Microbiol* 2018; **9**: 920.

383 11. Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. The Repertoire of
384 ICE in Prokaryotes Underscores the Unity, Diversity, and Ubiquity of Conjugation. *PLoS*
385 *Genet* 2011; **7**: e1002222.

386 12. Carraro N, Burrus V. The dualistic nature of integrative and conjugative elements. *Mob*
387 *Genet Elements* 2015; **5**: 98–102.

388 13. Cury J, Touchon M, Rocha EPC. Integrative and conjugative elements and their hosts:
389 composition, distribution and organization. *Nucleic Acids Res* 2017; **45**: 8943–8956.

390 14. Johnson CM, Grossman AD. Integrative and Conjugative Elements (ICEs): What They Do
391 and How They Work. *Annu Rev Genet* 2015; **49**: 577–601.

392 15. Carraro N, Poulin D, Burrus V. Replication and Active Partition of Integrative and
393 Conjugative Elements (ICEs) of the SXT/R391 Family: The Line between ICEs and

394 Conjugative Plasmids Is Getting Thinner. *PLOS Genet* 2015; **11**: e1005298.

395 16. Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. The hidden life of
396 integrative and conjugative elements. *FEMS Microbiol Rev* 2017; **41**: 512–537.

397 17. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids.
398 *Microbiol Mol Biol Rev* 2010; **74**: 434–52.

399 18. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the
400 AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance
401 Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrob Agents*
402 *Chemother* 2019; **63**.

403 19. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat*
404 *Methods* 2015; **12**: 59–60.

405 20. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011; **7**: e1002195.

406 21. Garcillán-Barcia MP, Redondo-Salvo S, Vielva L, de la Cruz F. MOBscan: Automated
407 Annotation of MOB Relaxases. *Methods in Molecular Biology*. 2020. Humana Press Inc.,
408 pp 295–308.

409 22. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. MacSyFinder: A Program to Mine
410 Genomes for Molecular Systems with an Application to CRISPR-Cas Systems. *PLoS One*
411 2014; **9**: e110726.

412 23. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. Identification of protein
413 secretion systems in bacterial genomes. *Sci Rep* 2016; **6**: 23080.

414 24. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb
415 software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018; **3**:
416 124.

417 25. Partridge SR, Tsafnat G. Automated annotation of mobile antibiotic resistance in Gram-
418 negative bacteria: the Multiple Antibiotic Resistance Annotator (MARA) and database. *J*
419 *Antimicrob Chemother* 2018; **73**: 883–890.

420 26. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and
421 modification: enzymes, genes and genomes. *Nucleic Acids Res* 2015; **43**: D298–D299.

422 27. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al.
423 CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced
424 performance and integrates search for Cas proteins. *Nucleic Acids Res* 2018; **46**: W246–
425 W251.

426 28. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline | Nucleic
427 Acids Research | Oxford Academic.
428 https://academic.oup.com/nar/article/47/W1/W81/5481154. Accessed 21 Feb 2020.

429 29. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology
430 resource based on 5090 organisms and 2502 viruses | Nucleic Acids Research | Oxford
431 Academic. https://academic.oup.com/nar/article/47/D1/D309/5173662. Accessed 24
432 Feb 2020.

433 30. Cury J, Oliveira PH, de la Cruz F, Rocha EPC. Host Range and Genetic Plasticity Explain the
434 Coexistence of Integrative and Extrachromosomal Mobile Genetic Elements. *Mol Biol*
435 *Evol* 2018; **35**: 2230–2239.

436 31. Abril D, Marquez-Ortiz RA, Castro-Cardozo B, Moncayo-Ortiz JI, Olarte Escobar NM,
437 Corredor Rozo ZL, et al. Genome plasticity favours double chromosomal Tn4401b-bla
438 KPC-2 transposon insertion in the Pseudomonas aeruginosa ST235 clone. *BMC Microbiol*
439 2019; **19**.

440 32. Urbanowicz P, Izdebski R, Baraniak A, Żabicka D, Ziółkowski G, Hryniewicz W, et al.

441 Pseudomonas aeruginosa with NDM-1, DIM-1 and PME-1 β-lactamases, and RmtD3 16S
442 rRNA methylase, encoded by new genomic islands. *J Antimicrob Chemother* 2019.

443 33. Khan A, Shropshire WC, Hanson B, Dinh AQ, Wanger A, Ostrosky-Zeichner L, et al.
444 Simultaneous infection with enterobacteriaceae and pseudomonas aeruginosa
445 harboring multiple carbapenemases in a returning traveler colonized with candida auris.
446 *Antimicrob Agents Chemother* 2020; **64**.

447 34. Ding Y, Teo JWP, Drautz-Moses DI, Schuster SC, Givskov M, Yang L. Acquisition of
448 resistance to carbapenem and macrolide-mediated quorum sensing inhibition by
449 Pseudomonas aeruginosa via ICETn43716385. *Commun Biol* 2018; **1**: 57.

450 35. Nascimento APB, Ortiz MF, Martins WMBS, Morais GL, Fehlberg LCC, Almeida LGP, et al.
451 Intraclonal Genome Stability of the Metallo-β-lactamase SPM-1-producing Pseudomonas
452 aeruginosa ST277, an Endemic Clone Disseminated in Brazilian Hospitals. *Front Microbiol*
453 2016; **7**: 1946.

454 36. Zhan Z, Hu L, Jiang X, Zeng L, Feng J, Wu W, et al. Plasmid and chromosomal integration
455 of four novel blaIMP-carrying transposons from Pseudomonas aeruginosa, Klebsiella
456 pneumoniae and an Enterobacter sp. *J Antimicrob Chemother* 2018; **73**: 3005–3015.

457 37. McCracken MG, Adam HJ, Blondeau JM, Walkty AJ, Karlowsky JA, Hoban DJ, et al.
458 Characterization of carbapenem-resistant and XDR Pseudomonas aeruginosa in Canada:
459 results of the CANWARD 2007–16 study. *J Antimicrob Chemother* 2019; **74**: iv32–iv38.

460 38. Cuzon G, Naas T, Nordmann P. Functional characterization of Tn4401, a Tn3-based
461 transposon involved in blaKPC gene mobilization. *Antimicrob Agents Chemother* 2011;
462 **55**: 5370–3.

463 39. Tohya M, Tada T, Watanabe S, Kuwahara-Arai K, Zin KN, Zaw NN, et al. Emergence of
464 carbapenem-resistant pseudomonas asiatica producing NDM-1 and VIM-2 Metallo-

465 Lactamases in Myanmar. *Antimicrob Agents Chemother* 2019; **63**.

466 40. Banerjee R, Weisenhorn E, Schwartz KJ, Myers KS, Glasner JD, Perna NT, et al. Tailoring a
467 Global Iron Regulon to a Uropathogen. *MBio* 2020; **11**.

468 41. San Millan A, Escudero JA, Gifford DR, Mazel D, MacLean RC. Multicopy plasmids
469 potentiate the evolution of antibiotic resistance in bacteria. *Nat Ecol Evol* 2016; **1**: 0010.

470 42. Inglis RF, Bayramoglu B, Gillor O, Ackermann M. The role of bacteriocins as selfish genetic
471 elements. *Biol Lett* 2013; **9**: 20121173.

472 43. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. Key
473 components of the eight classes of type IV secretion systems involved in bacterial
474 conjugation or protein secretion. *Nucleic Acids Res* 2014; **42**: 5715–27.

475