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 IMP, NDM and SPM metallo-beta-lactamase families, and the serine beta-lactamase KPC and GES families. 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, clinically-relevant bacteria.

51

52

53 ## Author Notes

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

56

57 ## Abbreviations

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

62

63 ## Impact Statement

64 Carbapenems are commonly used to treat severe infections in humans. Resistance is
65 often mediated by carbapenemases. These enzymes degrade carbapenems and are
66 frequently present in plasmids. Here, we demonstrate that common carbapenemase-
67 encoding genes (CEGs) found in clinical isolates (e.g. bla_{KPC}, bla_{GES}, bla_{IMP}, bla_{NDM}, bla_{VIM})
68 can also be located within integrative and conjugative elements (ICEs). CEG-bearing ICEs
69 belong to three mating-pair formation families. These mobile elements may be
70 particularly important in bacteria where plasmids do not seem to play a significant role
71 in the spread of antibiotic resistance genes, as *Pseudomonas* spp. This study

72 considerably expands the knowledge of the repertoire of CEGs-bearing ICEs among
73 clinically-relevant bacterial pathogens, such as *Pseudomonas aeruginosa*, *Klebsiella*
74 *pneumoniae* and *Escherichia coli*.

75

76 **Data Summary**

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

85

86 **Introduction**

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

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

111 ICEs are self-transmissible MGEs that can integrate into and excise from the genome (as
112 transposons and phages do) and can exist as circular, sometimes replicable,
113 extrachromosomal elements and be transferred by conjugation (as some plasmids do)
114 [11–14]. ICE*clc* from *Pseudomonas knackmussii* [15], SXT from *Vibrio cholerae* [16],
115 pKLC102 from *P. aeruginosa* [17], and Tn*4371* from *Ralstonia oxalatica* [18] are among
116 the most well studied ICEs. ICEs appear to have a bipartite lifestyle that shifts between
117 vertical and horizontal transmission [12, 19, 20]. HGT by conjugation requires three main

118 components: a relaxase (MOB), a mating pair formation (MPF) system, and a type-IV
119 coupling protein, with the last two forming a spanning-membrane multi-protein
120 complex named type-IV secretion system (T4SS) [21]. Until now, eight MPF classes were
121 proposed (B, C, F, FA, FATA, G, I and T), based on the phylogeny of VirB4, the only
122 ubiquitous protein among the T4SS. The MPF_T is widely distributed both in conjugative
123 plasmids and ICEs, while MPF_F is more prevalent in plasmids and MPF_G on ICEs [11].

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

131

132 Material and methods

133 Bacterial genomes and carbapenemases search

134 In **Figure 1,** we present the workflow used in this study, from the acquisition of bacterial
135 genomes to the identification and characterization of putative ICEs. We retrieved all
136 bacterial genomes available in NCBI Reference Sequence Database (RefSeq, accessed on
137 21/03/2020), including complete and draft genome sequences, using ncbi-genome-
138 download v0.2.12 (https://github.com/kblin/ncbi-genome-download). We downloaded
139 over 6,000 curated AR protein sequences from AMRfinder database

140 (https://ftp.ncbi.nlm.nih.gov/pathogen/Antimicrobial_resistance/AMRFinderPlus/data
141 base/3.6/2020-01-22.1/) [22] and we built an in-house database only including the
142 proteins that code for a carbapenemase (n=1014, **Table S1**). We then blasted the
143 genomes against the extracted carbapenemases using diamond v0.9.29.130
144 (http://www.diamondsearch.org/index.php) [23], using minimum 100% identity and
145 subject cover and with the sensitive mode enabled.

146

147 **Tracing ICEs among the bacterial genomes**

148 The RefSeq protein files from the CEG-bearing genomes identified by diamond were
149 extracted. We used the hmmsearch function of the HMMER3 software package v3.3
150 (http://hmmer.org/) [24] to search the proteomes against the standalone version of
151 MOBfamDB, a curated hidden Markov models (HMM) relaxase database
152 (https://castillo.dicom.unican.es/mobscan_about/) [25]. We also used this function to
153 search the pfam v33.0 database for tyrosine or serine recombinase accessions numbers
154 (Pfam IDs PF00589 and PF07508). The hmmsearch command was used with default
155 parameters and an E-value threshold of 0.01. The CEG-bearing genomes with relaxase
156 and integrase hits were further analysed. We used the Find Repeats tool from Geneious
157 Prime® 2020.0.4 (https://www.geneious.com) to inspect the hits for direct repeats. To
158 delimit CEG-harbouring ICEs, we manually scanned candidate terminal regions with
159 direct repetitions of the 3' end from tRNA genes located next to the integrase-encoding
160 gene. When no tRNA gene was identified next to this gene, we scanned the presence of
161 direct repeats next to the integrase-encoding gene and next to candidate terminal
162 regions. To assist in identifying putative terminal regions, we looked for blocks of DNA

163 with variation in GC content. To predict the MPF families, the translated coding
164 sequences of delimited ICEs were analysed on the standalone CONJscan module of
165 MacSyFinder v1.0.5 (https://github.com/gem-pasteur/macsyfinder) [26, 27]. To identify
166 the multi-locus sequence type of the genomes containing CEG-bearing ICEs, we used
167 mlst v2.16.1 (https://github.com/tseemann/mlst), which scans the genomes against
168 PubMLST typing schemes (https://pubmlst.org/) [28].

169

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

171 Screening of AR genes among ICEs was performed using amrfinder v3.6.10
172 (https://github.com/ncbi/amr) [22]. The genetic platforms involved in the acquisition of
173 CEGs by ICEs were annotated using Galileo™ AMR
174 (https://galileoamr.arcbio.com/mara/) (Arc Bio, Cambridge, MA) [29]. We ran our
175 extracted ICEs against REBASE (http://rebase.neb.com/rebase/rebase.html) to look for
176 restriction-modification systems [30]. We used CRISPRCasFinder
177 (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) to look for CRISPR
178 (clustered regularly interspaced short palindromic repeats) arrays and their associated
179 (Cas) proteins within ICE sequences [31]. Secondary metabolite biosynthetic gene
180 clusters were traced using antismash v5.1.2
181 (https://antismash.secondarymetabolites.org/#!/start) [32]. We used eggNOG-mapper
182 v2 (http://eggnog-mapper.embl.de/) for functional annotation based on orthology
183 assignments of the ICE proteomes [33].

184

185 ## Results

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

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

203

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

206 We identified a total of 66 putative ICEs, including 42 newly characterized elements
207 associated with 17 different CEGs (**Table 1** and **Figure 2**). We could predict the
208 boundaries from 55 of these elements (**Table S5**). The terminal region of the remaining
209 11 putative ICEs could not be determined due to a fragmented contig or assembly gaps
210 within the sequence. Nearly half of the putative ICEs (48.5%, n=32/66) were integrated
211 at the 3' end of a $tRNA^{Gly}$ gene. Integration next to random genes was also observed
212 (**Figure 3** and **Table S5**). The bacterial hosts housing these elements belong to 23
213 sequence types (STs) (**Table S5**).

214 Using the CONJscan module of MacSyFinder we identified the MPF family for 62 hits
215 (incomplete MPF classes were predicted for the remaining 4 hits) and we noted that
216 these hits belong to three families: MPF_G (69%, n=43/62), MPF_T (26%) and MPF_F (5%)
217 (**Figure 2**, **Table 1**, **Table S5** and **Table S6**). In our results, the MPF_G class was only
218 associated with MOB_H, MPF_T with MOB_P and MPF_F with MOBF (**Figure 3**). All ICEs here
219 identified carried a tyrosine recombinase, with the majority of them (56%, n=37/66)
220 belonging to the P4 integrase, C-terminal catalytic domain family (INT_P4_C). The
221 shufflon-specific DNA recombinase Rci and bacteriophage Hp1_like integrase, C-
222 terminal catalytic domain family (INT_Rci_Hp1_C) and the DNA breaking-rejoining
223 enzymes, C-terminal catalytic domain family (DNA_BRE_C) integrases were also
224 identified in our collection (36% and 8%, respectively) (**Table 1** and **Table S5**). INT_P4_C
225 and INT_Rci_Hp1_C integrases were associated with ICEs belonging to the MPF_G and
226 MPF_T classes, while DNA_BRE_C was found on MPF_F ICEs (**Figure 3**). ICEs from the MPF_G
227 and MPF_T classes were particularly promiscuous, being responsible for the spread of
228 several CEGs of the metallo-beta-lactamase family such as bla_{NDM-1}, bla_{SPM-1} and bla_{IMP}
229 variants among clinically-relevant pathogens as *P. aeruginosa*, *K. pneumoniae* and *E.*

230 *coli*. ICEs of the MPF_F class carrying blaIMP$_{-6}$ or *bla*$_{KPC-2}$ were restricted to *E. coli* and *K.*
231 *pneumoniae*. The *bla*$_{SPM-1}$ gene was exclusively identified in *P. aeruginosa* and in ICEs of
232 the MPF_T class (**Table 1**).

233 We analysed the types of integrase, relaxase and MPF classes present among four model
234 ICEs: ICE*clc*, pKLC102, SXT and Tn*4371*. The MPF_G-INT_P4_C ICEs here identified are
235 related to ICE*clc*, since this ICE belongs to the same class, carries a MOB_H relaxase and
236 also an INT_P4_C integrase. The MPF_T-INT_Rci_Hp1_C ICEs belong to the Tn*4371* family,
237 which also carries the MOB_P relaxase. No conserved domain family could be attributed
238 for SXT; however, the MPF_F ICEs here reported should be related to this model ICE that
239 also uses a MOB_H relaxase.

240 Additionally, we also identified 386 hits encoding an integrase and a relaxase in the
241 vicinity of CEGs (**Table S7**). For these hits, however, we could not predict if the CEG is
242 located on a plasmid or an ICE, since the contig is fragmented and tracing the boundaries
243 of the element is not possible, or the sequences have assembly gaps that make this
244 prediction challenging. Also, some plasmids may encode a tyrosine or serine
245 recombinase, and some ICEs may encode replicases and partition systems that are
246 typical of plasmids [34], which can hinder the accurate prediction of genetic platforms
247 when the sequence has poor quality or is highly fragmented due to short-read
248 sequencing approaches.

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

MPF family	**CEG**	**Integrase type**	**Relaxase type**	**Bacterial Species**	**MGEs flanking the CEGs**	**ICE length (kb)[1]**	**References**
MPF_G (n=43)	AFM-1 (n=1)	INT_P4_C	MOB_H	*Bordetella trematum*	Flanked by IS*91* family ISs	130	This study
	DIM-1 (n=1)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa*	Class I In flanked by IS*6100*	89	This study, [6]
	GES-5 (n=5)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	93-116	This study, [6]
	GES-24 (n=1)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	63	This study
	IMP-1 (n=4)	INT_P4_C, INT_Rci_Hp1_C	MOB_H	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	76-109	This study, [6]

	IMP-13 (n=9)	INT_P4_C, INT_Rci_Hp1_C	MOB_H	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	65-89	This study, [6]
	IMP-14 (n=2)	INT_P4_C	MOB_H	*Achromobacter xylosoxidans*	Class I In within Tn*3*-like Tn	106-122	This study
	IMP-16 (n=1)	INT_P4_C	MOB_H	*Pseudomonas monteilii*	Class I In within Tn*3*-like Tn	86	This study
	IMP-54 (n=1)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	91	This study
	KPC-2 (n=1)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa*	Tn*4401*	115	This study, [35]
	NDM-1 (n=12)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa, Pseudomonas asiatica, Morganella morganii*	Next to (or flanked by) IS*91* family ISs	97-167	This study, [36, 37]

	VIM-2 (n=1)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn	65	This study
	VIM-4 (n=4)	INT_P4_C	MOB_H	*Pseudomonas aeruginosa, Klebsiella pneumoniae, Alcaligenes faecalis*	Class I In within Tn*3*-like Tn	88-102	This study
MPF_T (n=16)	AFM-1 (n=1)	INT_P4_C	MOB_P	*Stenotrophomonas maltophilia*	Flanked by IS*91* family ISs	77	This study
	KPC-2 (n=1)	INT_Rci_Hp1_C	MOB_P	*Pseudomonas* sp.	Next to IS*Kpn6* and IS*26*	62	This study, [6]
	NDM-1 (n=3)	INT_P4_C, INT_Rci_Hp1_C	MOB_P	*Pseudomonas aeruginosa, Pseudomonas asiatica*	Flanked by IS*91* family ISs	73-74	This study, [38]
	SPM-1 (n=11)	INT_Rci_Hp1_C	MOB_P	*Pseudomonas aeruginosa*	Flanked by IS*CR3*-like elements	44-58	This study, [39]

MPF_F (n=3)	IMP-6 (n=1) KPC-2 (n=2)	DNA_BRE_C DNA_BRE_C	MOB_F MOB_F	*Escherichia coli* *Klebsiella pneumoniae*	Class I In within Tn*3*-like Tn Tn*4401*	118 75	This study This study
Incomplete MPF_G (n=2)	IMP-8 (n=1) NDM-1 (n=1)	DNA_BRE_C INT_P4_C	MOB_H MOB_H	*Enterobacter cloacae* *Pseudomonas aeruginosa*	Class I In within Tn*3*-like Tn Flanked by IS*CR3*-like elements	124 98	This study, [40] This study
Incomplete MPF_T (n=1)	KPC-2 (n=1)	INT_Rci_Hp1_C	NA	*Achromobacter* sp.	Next to IS*Kpn6* and IS*26*	43	This study
Incomplete MPF_F (n=1)	NDM-5 (n=1)	DNA_BRE_C	NA	*Escherichia coli*	Flanked by IS*CR1* and IS*26*	54	This study, [41]

Abbreviations: CEG, carbapenemase-encoding gene; DNA_BRE_C, DNA breaking-rejoining enzymes, C-terminal catalytic domain; ICE, Integrative and Conjugative Elements; In, Integron; INT_P4_C, P4 integrase, C-terminal catalytic domain; INT_Rci_Hp1_C, Shufflon-specific DNA recombinase Rci and Bacteriophage Hp1_like integrase, C-terminal catalytic domain family; MPF, Mating Pair Formation; NA, no hit; ST, Sequence Type.

253 [1]Some sequences are not complete. For more details, please refer to **Table S5**.

254

255 **A variable repertoire of CEG-bearing integrons and transposons target**
256 **ICEs**

257 We identified 17 CEG variants among the 66 putative ICEs, dominated by *bla*$_{NDM-1}$ and
258 *bla*$_{SPM-1}$ (**Table 1**). Insertion sequences (ISs, e.g. ISCR3-like elements) were frequently
259 linked to the acquisition of *bla*$_{SPM-1}$ and *bla*$_{NDM-1}$ [38, 39], while *bla*$_{IMP}$, *bla*$_{VIM}$, and *bla*$_{GES}$
260 were found on class I integrons frequently integrated into Tn*3*-family transposons [6].
261 The *bla*$_{KPC-2}$ gene was typically found within Tn*4401*-like transposons, which are capable
262 of conferring a high frequency of transposition [42]. The recently identified AFM-1
263 metallo-beta-lactamase (Genbank accession number MK143105.1) was here identified
264 in two ICEs inserted in *Bordetella trematum* and *Stenotrophomonas maltophilia*
265 genomes (**Table 1**). We also found *bla*$_{NDM-1}$ genes in ICEs integrated into the genomes of
266 a recently proposed *Pseudomonas asiatica* species, which is spreading in hospital
267 settings in Myanmar [43]. Besides CEGs, the ICEs identified in this study also harbour
268 genes conferring resistance to other antibiotics, such as aminoglycosides,
269 fluoroquinolones, macrolides, and tetracyclines (**Table S8**), widening the spectrum of
270 transmissible AR genes selectable by carbapenems due to linkage.

271

272 **Acquisition of additional traits by ICEs include competitive weapons as**
273 **bacteriocins and siderophores**

274 Besides genes conferring antibiotic resistance, the CEG-bearing ICEs here identified
275 harbour other cargo genes that may confer a selective advantage to the ICE host. We
276 found DUF692 domains typical of bacteriocin production genes among the 6 MPF$_G$ ICEs

277 from *P. asiatica* strains and the *P. aeruginosa* N15-01092, 1334/14 and ST773 strains
278 (**Table S5**). All these ICEs carry a bla_{NDM-1} gene and a similar copy of the bacteriocin-
279 encoding gene. Curiously, the bacteriocin production gene was not identified in the
280 MPF_T ICE from *P. asiatica* strain MY569. Additionally, we found the siderophore
281 aerobactin operon within an ICE in *E. coli* strain E302 (**Table S5**). This operon is usually
282 found in enterobacterial plasmids, and was also identified in a pathogenicity island in
283 uropathogenic *E. coli* strain CFT073 [44]. We identified no CRISPR-Cas systems among
284 the ICEs here identified. Nearly half of them (47.0%, n=31/66) carried complete or
285 incomplete restriction-modification systems belonging to type II, III and IV (**Table S9**).

286 The majority of the proteins encoded within the 66 ICEs refer to replication,
287 recombination, transcription and intracellular trafficking functions (**Figure 4**). Several
288 proteins, however, encoded for unknown functions (34.0%, n=1,615/4,754, **Table S10**),
289 highlighting the lack of knowledge on the ICE proteome.

290

291 ## Discussion

292 We have set out to comprehensively identify the CEGs among all bacterial genomes
293 deposited in NCBI and the CEG-bearing ICE sequences. Our study considerably expands
294 the knowledge of the repertoire of CEGs-bearing ICEs. We uncovered 66 putative ICEs
295 that may be involved in HGT of CEGs amongst bacterial genomes. To expand our
296 predictions, we also used the CONJscan module of Macsyfinder to trace the MPF families
297 likely involved in HGT. Our analysis on the co-occurrence of relaxases with MPF families
298 (**Table 1** and **Table S5**) is in agreement with the combinations observed by Guglielmini
299 and colleagues [11]; all ICEs belonging to the MPF_G class carried a MOB_H relaxase, and
300 the MOB_P and MOB_F relaxases were linked to MPF_T and MPF_F, respectively. All MPF_G
301 class ICEs here described present a MPF class-relaxase-integrase profile that resembles
302 that from ICE*clc*. Even though pKLC102 is also a representative of the MPF_G class and
303 carries a MOB_H relaxase, it uses a DNA_BRE_C integrase instead of a INT_P4_C. The
304 absence of CEG-carrying ICEs from the pKLC102 family was already reported [6].

305 The scenario observed for the acquisition of the most important CEGs by ICEs (ISs for
306 bla_{NDM-1} and bla_{SPM-1}, class I integrons for bla_{IMP} and bla_{VIM} and Tn*4401* for bla_{KPC-2})
307 resembles that of plasmids [3] and provide additional support to the notion that the line
308 separating these elements is blurred [13, 14]. We now show that besides plasmids, this
309 promiscuous repertoire of integrons and transposons frequently targets ICEs of different
310 MPF families. Even though CEGs might rapidly spread worldwide, local selection is likely
311 required for them to reach fixation, as can be seen for the clonal expansion of *P.*
312 *aeruginosa* ST277 harbouring bla_{SPM-1} [39]. Surprisingly, we noted that this gene was not
313 detected beyond the same clonal lineage. Indeed, all hits were identified in *P.*

314 *aeruginosa* ST277 strains from Brazil and within a MPF_T ICE family, indicating that these
315 ICEs are transferring vertically and/or horizontally within this ST. Understanding the
316 limitations on HGT of this family of ICE may be translatable to other, more transferable,
317 ICEs and could underpin a control strategy to prevent the spread of these elements in
318 the future.

319 Although bla_{KPC} and bla_{OXA} were the most frequently identified CEGs within the analysed
320 genomes (**Table S2**), we only found 5 ICEs carrying bla_{KPC-2} and 2 ICEs with bla_{OXA} genes
321 (**Table S8**). The bla_{KPC} genes are mostly located in Tn*4401* transposons that target *K.*
322 *pneumoniae* plasmids, while the bla_{OXA} genes are frequently associated with ISs that
323 tend to target *Acinetobacter* spp. chromosomes and plasmids [3]. These genes may take
324 advantage of the copy number and the higher genetic plasticity of plasmids [34]. This
325 plasticity may increase the rate of appearance of novel mutations and the high copy
326 number may amplify the effect due to the increased gene dosage [45].

327 In addition to AR, ICEs can be involved in other adaptive traits such as carbon-source
328 utilization, symbiosis, restriction-modification, siderophore and bacteriocin synthesis
329 [14]. Since bacteria commonly inhabit highly competitive environments, the production
330 of specific secondary metabolites (such as bacteriocins and siderophores) may confer a
331 selective advantage to the host [46]. We speculate that the presence of these
332 metabolites within the ICEs here characterized may promote their stability by
333 preferentially selecting for cells harbouring the ICE. CRISPR-Cas systems are rarely found
334 on MGEs, and the type I-C systems carried within pKLC102-related ICEs are one of these
335 few examples [9]. Since none of the ICEs here described belong to the pKLC102 family,
336 the absence of CRISPR-Cas systems within our dataset was expected.

337 One caveat of our studies is that these results do not yet expose the complete set of
338 CEG-bearing ICEs present in all bacteria. There is an inherent bias in the number of times
339 we detect a particular CEG in certain bacterial genomes as some are over represented
340 in the database compared to others. It is possible that certain observations will flatten
341 out as more genomes are analysed. Less than 10% of the bacterial genomes currently
342 present on NCBI are complete. This is a major drawback since putative ICEs present in
343 draft genomes tend to be fragmented due to the presence of repetitive regions that are
344 not resolved using short-read sequencing. Also, the relaxase and the T4SS encoded by
345 ICEs resemble those of plasmids [11, 13, 47]. Plus, it is possible that ICEs and plasmids
346 have swapped conjugation modules throughout their evolutionary history [11]. We
347 believe that a more thorough exploration of this issue, especially regarding the precise
348 delimitation of ICEs, will be an important further step toward an improved
349 understanding of the contribution of these elements to bacterial adaptation and
350 evolution of AR.

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

357

358 **Data bibliography**

359 The DNA sequences used in this study are available in NCBI, under the accession
360 numbers provided in supplementary tables S1-S9.
361

362 Funding information

363 This work was supported by the Applied Molecular Biosciences Unit - UCIBIO which is
364 financed by national funds from FCT (UIDB/04378/2020).
365

366 Author contributions

367 J. B. conceptualized the project. J. B. ran the analyses. J.B. and J. M. analysed the data.
368 J.B. wrote the manuscript. J. M., A. P. R. and L. P. edited and revised the manuscript. All
369 authors read, commented on and approved the final manuscript.
370

371 Conflicts of interest

372 The authors declare that there are no conflicts of interest.
373

374 References

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

377 2. World Health Organization. GLOBAL PRIORITY LIST OF ANTIBIOTIC-RESISTANT BACTERIA
378 TO GUIDE RESEARCH, DISCOVERY, AND DEVELOPMENT OF NEW ANTIBIOTICS.
379 http://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-

380 which-new-antibiotics-are-urgently-needed. Accessed 11 Aug 2018.

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

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

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

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

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

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

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

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

402 11. Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. The Repertoire of

403 ICE in Prokaryotes Underscores the Unity, Diversity, and Ubiquity of Conjugation. *PLoS*
404 *Genet* 2011; **7**: e1002222.

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

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

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

411 15. Gaillard M, Vallaeys T, Vorhölter FJ, Minoia M, Werlen C, Sentchilo V, et al. The clc
412 element of Pseudomonas sp. strain B13, a genomic island with various catabolic
413 properties. *J Bacteriol* 2006; **188**: 1999–2013.

414 16. Waldor MK, Tschäpe H, Mekalanos JJ. A new type of conjugative transposon encodes
415 resistance to sulfamethoxazole, trimethoprim, and streptomycin in Vibrio cholerae
416 O139. *J Bacteriol* 1996; **178**: 4157–4165.

417 17. Klockgether J, Reva O, Larbig K, Tümmler B. Sequence Analysis of the Mobile Genome
418 Island pKLC102 of Pseudomonas aeruginosa C. *J Bacteriol* 2004; **186**: 518–534.

419 18. Springael D, Kreps S, Mergeay M. Identification of a catabolic transposon, Tn4371,
420 carrying biphenyl and 4- chlorobiphenyl degradation genes in Alcaligenes eutrophus A5.
421 *J Bacteriol* 1993; **175**: 1674–1681.

422 19. Botelho J, Schulenburg H. The Role of Integrative and Conjugative Elements in
423 Antibiotic Resistance Evolution. *Trends Microbiol* 2020; **0**.

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

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

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

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

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

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

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

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

443 28. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb
444 software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018; **3**:
445 124.

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

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

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

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

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

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

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

469 36. Urbanowicz P, Izdebski R, Baraniak A, Żabicka D, Ziółkowski G, Hryniewicz W, et al.
470 Pseudomonas aeruginosa with NDM-1, DIM-1 and PME-1 β-lactamases, and RmtD3 16S
471 rRNA methylase, encoded by new genomic islands. *J Antimicrob Chemother* 2019.

472 37. Khan A, Shropshire WC, Hanson B, Dinh AQ, Wanger A, Ostrosky-Zeichner L, et al.

473 Simultaneous infection with enterobacteriaceae and pseudomonas aeruginosa
474 harboring multiple carbapenemases in a returning traveler colonized with candida
475 auris. *Antimicrob Agents Chemother* 2020; **64**.

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

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

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

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

489 42. Cuzon G, Naas T, Nordmann P. Functional characterization of Tn4401, a Tn3-based
490 transposon involved in blaKPC gene mobilization. *Antimicrob Agents Chemother* 2011;
491 **55**: 5370–3.

492 43. Tohya M, Tada T, Watanabe S, Kuwahara-Arai K, Zin KN, Zaw NN, et al. Emergence of
493 carbapenem-resistant pseudomonas asiatica producing NDM-1 and VIM-2 Metallo-
494 Lactamases in Myanmar. *Antimicrob Agents Chemother* 2019; **63**.

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

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

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

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

504

505

506

507

508

509

510

511

512

513

514

515

516

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

527

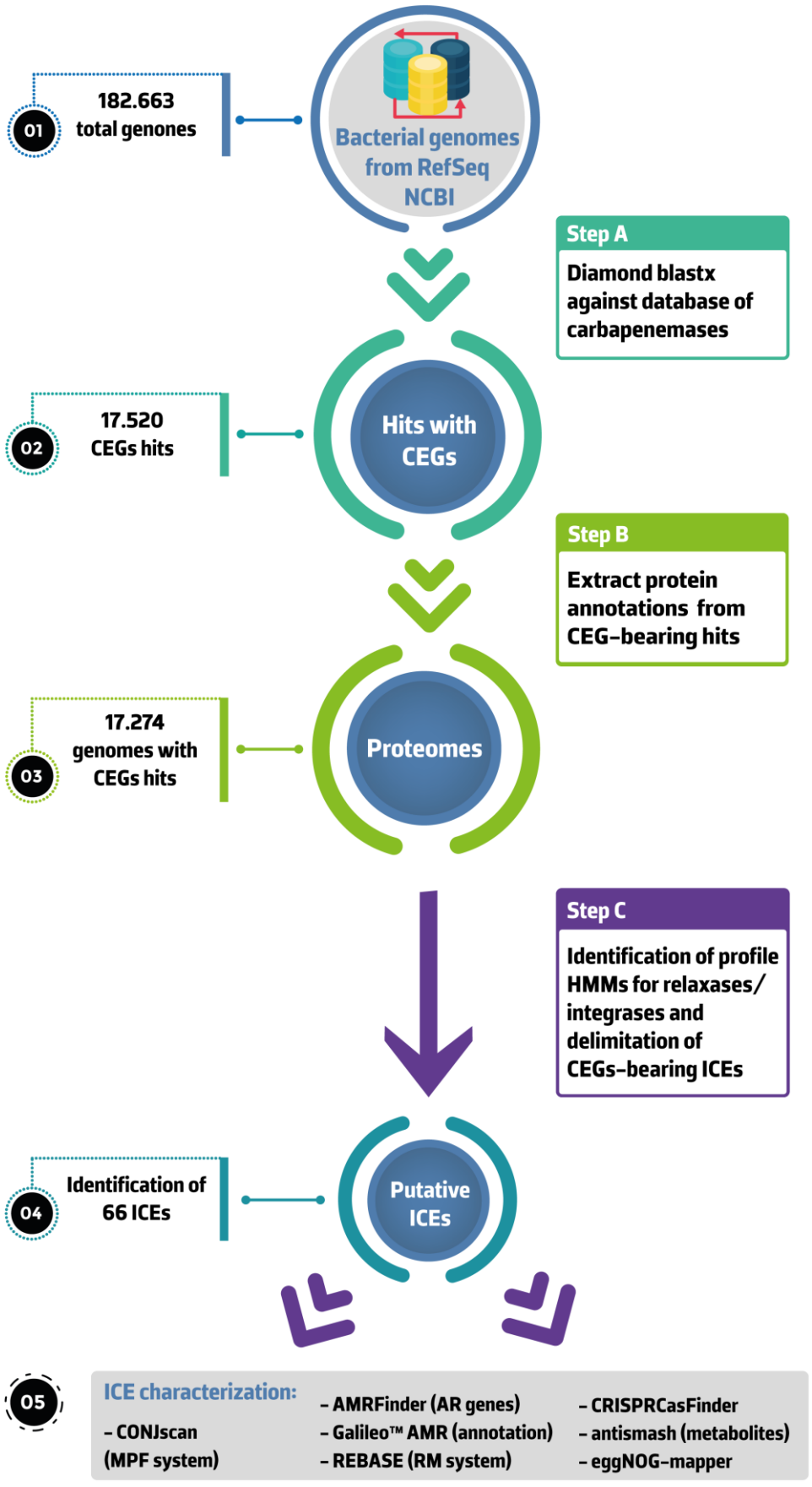
528 **Figure 2 –** Sankey diagram showing the contribution of different MPF families for the
529 spread of CEGs among several bacterial genomes. The left, centre and right axis
530 represents the association between the identified carbapenemases, MPF type and the
531 bacterial genus, respectively, while the width of each connection is proportional to the
532 number of positive hits.

533

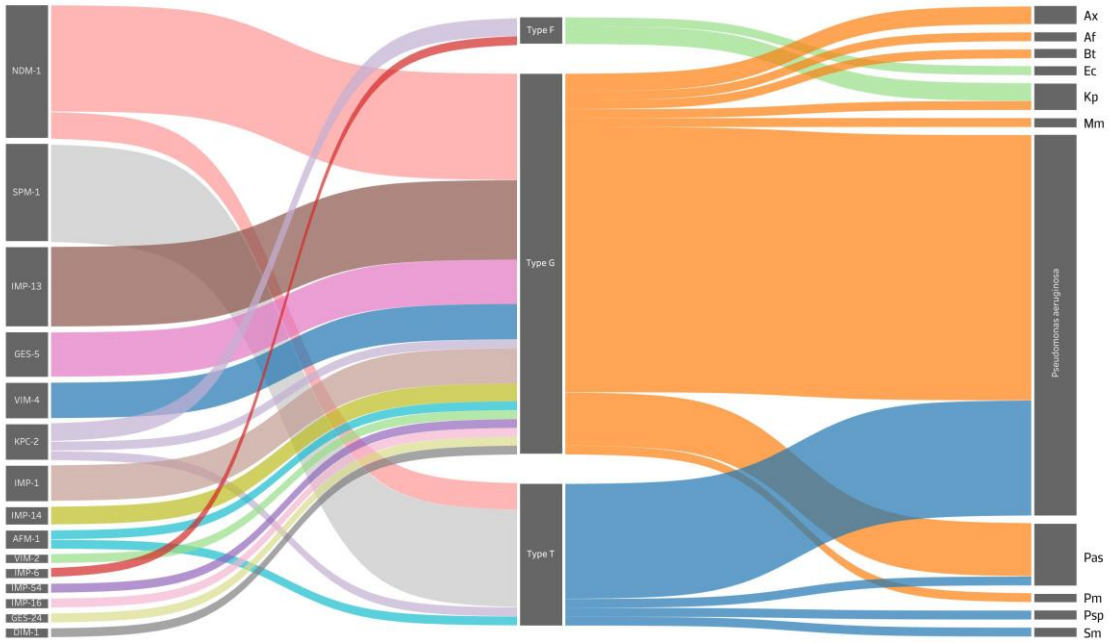
534 **Figure 3 –** Schematic representation of the different insertion site-integrase-CEGs-MPF-
535 relaxase profiles identified in this study. The figure is not to scale and the relative
536 position of the different modules (integration, conjugation, accessory CEGs) is for
537 illustrative purposes to show the various relationships observed.

538

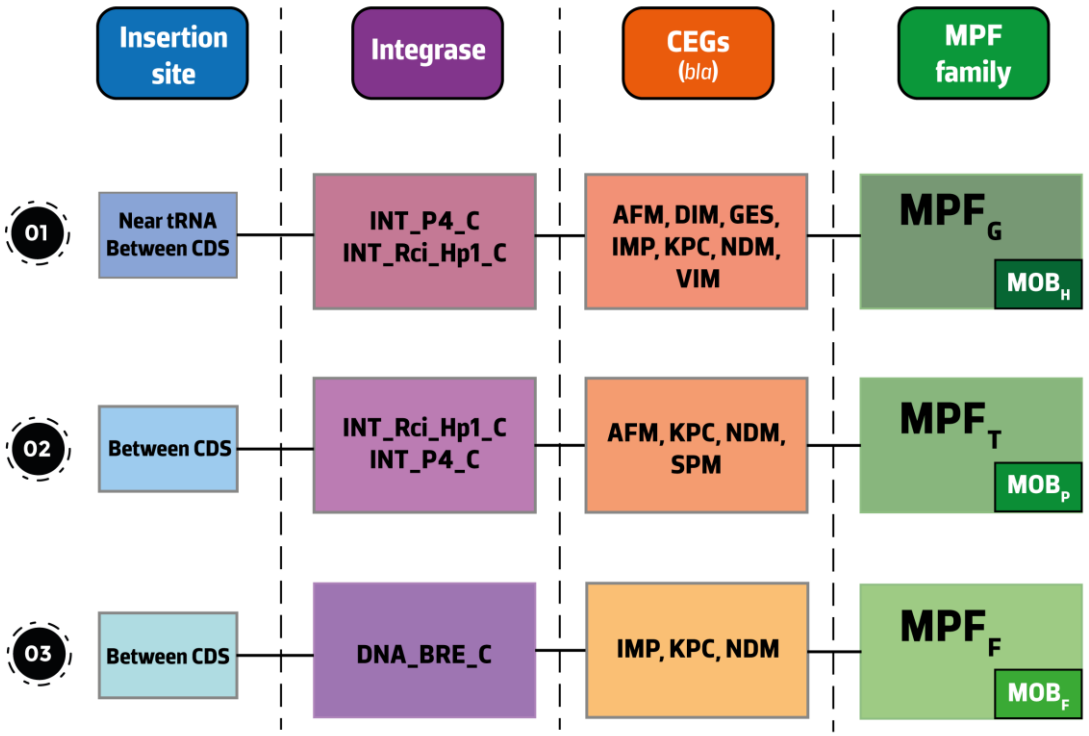
Figure 4 – Grouped bar chart representing the incidence of each eggNOG function broken down by category (represented by different colours) in the CEG-bearing ICE sequences.



Abbreviations: AR, Antibiotic Resistance; CEGs, Carbapenemase-Encoding Genes; HMMs, Hidden Markov Models; ICEs, Integrative and Conjugative Elements; MPF, Mating-Pair Formation; RM, Restriction-Modification.



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.



Abbreviations: CDS, CoDing Sequence; CEGs, Carbapenemase-Encoding Genes; DNA_BRE_C, DNA breaking-rejoining enzymes_C-terminal catalytic domain; INT_P4_C, P4 integrase_C-terminal catalytic domain; INT_Rci_Hp1_C, Shufflon-specific DNA recombinase Rci and Bacteriophage Hp1_like integrase_C-terminal catalytic domain family; MPF, Mating-Pair Formation; tRNA, transfer RNA gene.

571

572

573

574

575

576

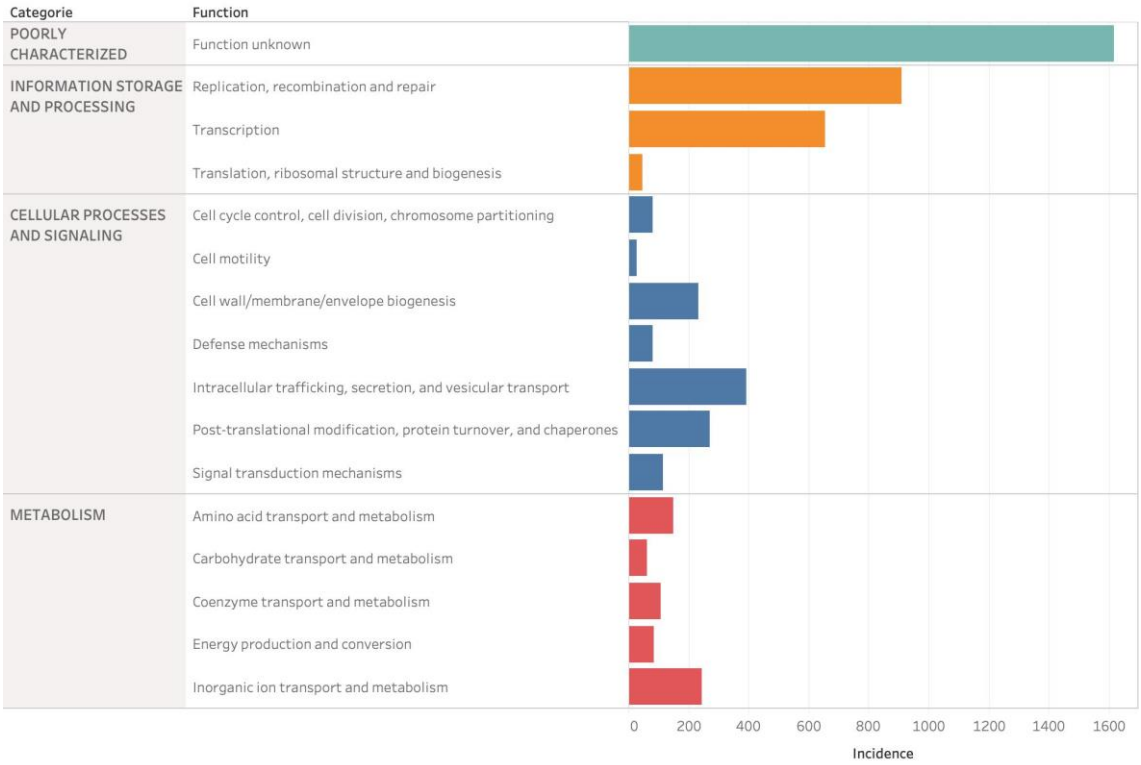
577

578

579

580

581


Categorie
Function
POORLY CHARACTERIZED
Function unknown
INFORMATION STORAGE AND PROCESSING
Replication, recombination and repair
Transcription
Translation, ribosomal structure and biogenesis
CELLULAR PROCESSES AND SIGNALING
Cell cycle control, cell division, chromosome partitioning
Cell motility
Cell wall/membrane/envelope biogenesis
Defense mechanisms
Intracellular trafficking, secretion, and vesicular transport
Post-translational modification, protein turnover, and chaperones
Signal transduction mechanisms
METABOLISM
Amino acid transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Energy production and conversion
Inorganic ion transport and metabolism
0
200
400
600
800
1000
1200
1400
1600
Incidence

582