The ESKAPE mobilome contributes to the spread of antimicrobial resistance and CRISPR-mediated conflict between mobile genetic elements

João Botelho[1,2*], Adrian Cazares[3,4], Hinrich Schulenburg[1,2]

[1]Antibiotic Resistance Evolution Group, Max Planck Institute for Evolutionary Biology, Plön, Germany

[2]Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian Albrechts University, Kiel, Germany

[3]EMBL's European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, United Kingdom

[4]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, United Kingdom

[*]Correspondence: botelho@evolbio.mpg.de

Abstract

Mobile genetic elements (MGEs) mediate the shuffling of genes among organisms. They contribute to the spread of virulence and antibiotic resistance genes in human pathogens, including the particularly problematic group of ESKAPE pathogens, such as *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* sp. Here, we performed the first systematic analysis of MGEs, including plasmids, prophages, and integrative and conjugative/mobilizable elements (ICEs/IMEs), in the ESKAPE pathogens. We characterized over 1700 complete ESKAPE genomes and found that different MGE types are asymmetrically distributed across these pathogens. While some MGEs are capable of exchanging DNA beyond the genus (and phylum) barrier, horizontal gene transfer (HGT) is mainly restricted by phylum or genus. We further observed that most genes on MGEs have unknown functions and show intricate distribution patterns. Moreover, AMR genes and anti-CRISPRs are overrepresented in the ESKAPE mobilome. Our results also underscored species-specific trends shaping the number of MGEs, AMR, and virulence genes across pairs of conspecific ESKAPE genomes with and without CRISPR-Cas systems. Finally, we found that CRISPR targets vary according to MGE type: while plasmid CRISPRs almost exclusively target other plasmids, ICEs/IME CRISPRs preferentially target ICEs/IMEs and prophages. Overall, our study highlights the general importance of the ESKAPE mobilome in contributing to the spread of AMR and mediating conflict among MGEs.

Introduction

Mobile genetic elements (MGEs) are DNA entities that are capable of capturing and shuffling genes intra- and intercellularly [1]. Coevolution of bacterial hosts with these MGEs has driven the evolution of complexity [2]. Movement within the genome is often mediated by specific MGEs, such as insertion sequences and transposons [3]. Others like plasmids, prophages, and integrative and conjugative/mobilizable elements (ICEs/IMEs) are key vectors for intercellular mobility, being responsible for a large fraction of the variability observed between bacterial species [4–6]. Bacteria undergo extensive horizontal gene transfer (HGT), and some estimates suggest that more than 80% of bacterial genes were laterally transferred at some point in their evolutionary history [7]. These events are largely shaped by ecological niches, by the difference in the GC content between pairs of bacteria exchanging material, and by phylogenetic barriers [8–10]. Network-based methods are useful to trace HGT events and recover shared content between bacterial genomes [10,11], and have been recently applied to explore the population structure of thousands of plasmids [12,13]. Even though this approach has been useful to explore population structure of plasmids, network analyses of other MGEs (such as prophages and ICEs/IMEs) is understudied. More importantly, only a few studies have used network-based approaches to explore the co-evolutionary dynamics of different MGE types [14,15].

MGEs carry non-essential genes that can provide their bacterial host with adaptive traits and alter their fitness, such as antimicrobial resistance (AMR) and virulence genes. These elements apply a myriad of ecological and evolutionary strategies to promote their own replication and transmission, which allow them to persist even in the absence of positive selection for the beneficial genes they carry [16,17]. As a predicted immune response against invading MGEs, bacteria have evolved different, often complex, mechanisms such as restriction-modification and clustered, regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) genes [18]. These systems are usually clustered in 'defence islands' and are widespread in bacteria and archaea [19,20]. Spacer sequences, derived from fragments of invading MGEs (termed protospacers), are continuously incorporated at CRISPR loci, enabling partial reconstruction of recent HGT events. Curtailment of HGT through CRISPR-Cas immunity can be favourable when dealing with obligate parasites, but can be detrimental by hindering the acquisition of genetic novelty and/or beneficial traits carried by MGEs. Invasion by these MGEs can, however, be associated with fitness costs that may lead to selection against carriage [5,21]. Hence, bacteria often face a trade-off between immunity and acquisition of novel elements, which favour adaptation to different ecological niches and stressors, such as antibiotic pressure. MGEs can be equipped with inhibitors of CRISPR-Cas systems, called anti-CRISPR (Acr) proteins, which have

66 been reported mostly in prophages [22–24]. Recently, Acr proteins were identified in non-phage
67 MGEs, including plasmids and ICEs [25].

68 Bacterial pathogens belonging to the ESKAPE panel consist of five species (*Enterococcus faecium*,
69 *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *Pseudomonas*
70 *aeruginosa*) and one genus (*Enterobacter* sp.) [26,27]. These pathogens are frequently involved in
71 nosocomial infections, due to their multi-drug resistance and/or invasive phenotypes [28–33]. The
72 WHO recently published a list of pathogens for which new antibiotic development is urgently
73 required, and the ESKAPE pathogens were designated "priority status" [34]. AMR and virulence
74 genes are broadly distributed in plasmids across the ESKAPE pathogens [28,35], and also in ICEs [36,37].
75 Recently, CRISPR-Cas systems have been identified in plasmids and ICEs from several bacterial
76 species (including representatives of the ESKAPE pathogens), and may be involved in conflict
77 between MGEs [38–40].

78 In this study, we performed the first systematic analysis focusing on the ESKAPE pathogens
79 mobilome. We asked i) how prevalent are different MGEs (plasmids, ICEs/IMEs, and prophages)
80 across the ESKAPE pathogens; ii) how broad or constrained is the combined MGEs' network; iii)
81 which traits are overrepresented in these MGEs, and if AMR and virulence genes are differently
82 distributed in pairs of conspecific ESKAPE pathogens with and without CRISPR-Cas systems; iv)
83 whether the CRISPR spacers have a targeting bias towards different MGE types. We found that
84 plasmids, ICEs/IMEs, and prophages are unequally distributed across these pathogens, and
85 found signatures of DNA sharing events between different species. Uncovering the structure of
86 MGEs and masked (i.e., MGE-free) genomes allowed us to discover an overrepresentation of
87 AMR genes and anti-CRISPRs in the ESKAPE mobilome. Our results also unveiled ESKAPE-specific
88 trends of MGEs, AMR, and virulence genes promoted by the presence of CRISPR-Cas systems.
89 Finally, our work shows that CRISPR spacers found on ICEs/IMEs and plasmids across the ESKAPE
90 pathogens have different targeting biases.

91

92 **Results**

93

94 **1. MGEs are unevenly distributed among the ESKAPE pathogens**

95 We downloaded 1782 ESKAPE complete genomes from NCBI's RefSeq database. To correct for
96 species taxonomy, genomes with <95% average nucleotide identity (ANI) were removed for each
97 ESKAPE species (**Table S1)**. Since this parameter is only applied for species delineation, we also

98 built a phylogenomic tree with *Enterobacter* sp. genomes and type strains belonging to the
99 *Enterobacteriaceae* family (**Figure S1**). Our curated dataset included 1746 complete genomes,
100 belonging to multiple MLST profiles (**Table S2**) and consisting of the same number of
101 chromosomes plus 2640 plasmids (**Figures 1A and 1D**). We confirmed that plasmids were
102 abundant in *K. pneumoniae* (~3 plasmids/genome). These replicons were also prevalent in every
103 ESKAPE except *P. aeruginosa* and *S. aureus* (**Figure 1A**). The majority of plasmids carried a
104 relaxase (62.5%, 1651/2640), and were classified as mobilizable (either self-conjugative or not)
105 [41].

106 To look for regions of genome plasticity (RGPs) exclusively integrated in the chromosome, we
107 used the 1746 chromosomal replicons to generate plasmid-free pangenomes for each ESKAPE
108 taxon. We identified a total of 50482 plasmid-free RGPs in chromosomal replicons (**Figure 1E**).
109 Of these, 2685 were classified as ICEs/IMEs due to the presence of relaxase and integrase
110 domains (**Figures 1D and 1E**). At least one ICE/IME was detected in >50% of genomes for all
111 ESKAPE pathogens, and was abundant in *E. faecium* and *P. aeruginosa* (~3 elements/genome)
112 (**Figures 1B and 1D**). After masking the ICEs/IMEs identified in the ESKAPE chromosomes, we
113 performed a search for prophages. Prophages were less frequently found than ICEs/IMEs and
114 plasmids (**Figures 1C and 1D**). Prophages were however abundant in *P. aeruginosa*, *Enterobacter*
115 sp., and *K. pneumoniae*. We found ~2 elements/genome on *K. pneumoniae*, whereas these
116 elements were particularly rare in *S. aureus*.

117 When looking into the presence/absence combination of co-occurring MGEs across the ESKAPE
118 pathogens, we noticed that the most frequent combination involved the presence of the three
119 MGEs (in 387 out of the 1746 genomes, **Table S2**). Only 144 strains carried none of these
120 elements. Examples of these contrasting combinations (i.e., all MGEs vs. no MGE at all) were
121 found in all representatives of the ESKAPE pathogens. We noticed that the majority of the strains
122 with the three MGEs co-occurring in the same genome belonged to *K. pneumoniae* (268/387),
123 while the no MGEs combination was mostly represented by *S. aureus* (118/144). Non-
124 surprisingly, we observed that *S. aureus* was the only ESKAPE representative with a proportion
125 of MGE/strain lower than one. Our results show that different MGEs are asymmetrically
126 distributed across the ESKAPE pathogens, with *K. pneumoniae* genomes taking the lead for the
127 co-occurrence of ICEs/IMEs, plasmids, and prophages (**Figure 1**).

128

129 **2. DNA sharing between different vectors is common across the ESKAPE mobilome**

130 MGEs tend to have a GC content lower than the one from its host [42–44]. However, how conserved
131 is this difference when considering different MGE types is poorly studied. We confirmed that
132 the arithmetic mean GC content of plasmids, ICEs, and prophages is significantly lower when
133 compared to masked genomes across the ESKAPE pathogens (**Figure S2**, p-value < 2.2e-16). With
134 the exception of *S. aureus*, we observed that plasmids across the ESKAPE pathogens have a wider
135 variation in their size, when compared with ICEs/IMEs and prophages (**Figure S3**). Across all
136 ESKAPE pathogens, we observed a weak positive correlation between the ICEs/IMEs and
137 plasmids' GC content and sequence length (R = 0.38 and 0.35, respectively, p<2.2e-16), and a
138 weak negative correlation between the prophages' GC content and sequence length (R = -0.074,
139 p = 0.0031, **Figure S4**).

140 We used an alignment-free sequence similarity comparison of the ESKAPE mobilome to infer an
141 undirected network (**Figures 2A and 2B**). The sparse network assigned 96.7% (6715) of the MGEs
142 into 150 clusters. The majority of MGE pairs shared little similarity, with a Jaccard Index (JI) value
143 below 0.25 (**Figure S5A**), in accordance with the high diversity frequently observed across MGEs.
144 The network revealed clear structural differentiation, where the majority of clusters were
145 homogeneous for a given ESKAPE/MGE pair (**Figures 2A and 2B**). The absence of pairwise
146 distance similarities with intermediate JI (**Figure S5A**) helps to explain this clustering in discrete
147 groups, instead of a continuous genetic structure. However, the two largest clusters challenge
148 interspecies and MGE type barriers and correspond to multiple MGEs with the four
149 Proteobacteria representatives in the first (i.e., *K. pneumoniae*, *Enterobacter* sp., *P. aeruginosa*,
150 and *A. baumannii*), and *S. aureus* and *E. faecium* in the second cluster. MGEs within these
151 promiscuous clusters tend to be more dissimilar than those assigned to densely connected and
152 ESKAPE/MGE pair-restricted clusters. To plot this network, we used as a threshold the mean
153 value (0.07612963) of the estimated pairwise distances between the 6941 MGEs identified in
154 this study. We also plotted the network using the median value (0.009765625) and confirmed a
155 lower sparsity of the network and the two larger clusters physically linked to one another (**Figure**
156 **S6**). In parallel, we built a network using the masked genomes as nodes connected by edges
157 indicating pairwise distances, and we exclusively observed species-specific clustering (**Figures**
158 **S5B and S7**), which is in agreement with cohesive evolutionary forces that maintain the genomic
159 coherence of bacterial species [45,46]. Interestingly, we verified that ICEs/IMEs from *P. aeruginosa*
160 are the MGE/ESKAPE pair most frequently associated with other bacteria, sharing DNA segments
161 with plasmids from *Enterobacter* sp., *K. pneumoniae*, and *A. baumannii*, as well as ICEs/IMEs
162 from *Enterobacter* sp. and *K. pneumoniae*, and prophages from *K. pneumoniae* (**Figure S8** and
163 **Table S3**). Even though rare, we observed cross-phylum interactions, between plasmids from *P.*

164 *aeruginosa* and plasmids from *S. aureus*, as well as between an ICEs/IME from *E. faecium* and a
165 plasmid from *K. pneumoniae* (**Figure S8** and **Table S3**). Altogether, our results highlight that
166 while some elements are capable of exchanging DNA beyond the genus (and phylum) barrier,
167 for the majority of clusters host similarity and MGE type restrain DNA sharing between different
168 ESKAPE MGEs.

169

170 **3. The proteome of the ESKAPE mobilome is highly diverse in sequence and functions**

171 To gain functional insights into the proteome of the ESKAPE mobilome, we investigated the
172 diversity of clusters of orthologous groups (COGs) encoded by the MGEs identified in this study.
173 COGs are protein sets conserved across different lineages that typically share functions, and are
174 therefore used for systematic function prediction in poorly characterised genomes [47].
175 Furthermore, COGs have been assigned to curated and uniform functional categories, enabling
176 the comparison of their distribution among genome sequences [48]. We distinguished 1183
177 different COGs in ESKAPE MGEs (**Figure S9**) encompassing most of the functional categories
178 reported in the COGs scheme (**Figure 3**), thereby highlighting the diversity of functions
179 associated with the ESKAPE mobilome. Plasmids, prophages, and ICEs/IMEs contain unique
180 COGs, consistent with their distinctive features as mobile elements (**Figure S9**). However, they
181 also shared 113 COGs, indicative of a common pool of proteins and functions carried by MGEs
182 in ESKAPE pathogens. We detected some variation in the relative contribution of MGE
183 proteomes to different COG functional categories among the ESKAPE pathogens (**Figure 3**). For
184 example, proteins associated with "carbohydrate transport and metabolism" are more frequent
185 in MGE proteomes of *E. faecium* (particularly in plasmids and ICEs/IMEs) than in other ESKAPE
186 (**Figure 3**). On the other hand, proteins of the "replication, recombination and repair" category
187 are similarly frequent amongst MGE proteomes of most ESKAPE. Another interesting example is
188 the relative high frequency of "translation, ribosomal structure and biogenesis" proteins in the
189 proteome of *K. pneumoniae* prophages, which appears to be distinct across MGE/ESKAPE pairs.

190 To explore the diversity of the ESKAPE MGEs proteome further, we clustered their 474223
191 proteins based on sequence similarity, resulting in 51960 groups (**Table S4**). Around 78% of the
192 representatives of these protein groups were assigned the tag "hypothetical protein" by Prokka,
193 underlining the large proportion of uncharacterised proteins encoded by ESKAPE MGEs. Among
194 the representatives with an assigned function, transposases and integrases were the most
195 frequent protein product (1829 cluster representatives) (**Table S4**). Recombinase, repressor and
196 resistance, were also common terms across the representative products, with more than 200

197 occurrences each; the latter being mostly associated with metal or drug resistance. We looked
198 for widespread protein clusters in the MGEs proteome, i.e. those observed in the three MGE
199 types or in at least three of the ESKAPE pathogens. Our search resulted in the identification of
200 466 protein clusters widespread across MGEs and 383 present in at least three different ESKAPE
201 (**Table S4**). Interestingly, only 38 clusters were identified in common between the two
202 widespread categories. Although hypothetical proteins dominated both widespread categories
203 (79% of MGEs and 52% of ESKAPE widespread protein clusters), various protein clusters with
204 functions associated with transposition and antimicrobial resistance were also identified (**Table**
205 **S4**). Hierarchical clustering of the MGE/ESKAPE pairs and protein clusters based on the
206 distribution and relative frequency of widespread proteins uncovered structured patterns of
207 sharing (**Figure 4**). For example, we detected a large component of protein clusters present in
208 plasmids of *Enterobacter* sp., *K. pneumoniae*, and *A. baumannii*. When it comes to protein
209 clusters present in different MGE types, we observed sets predominantly occurring in ICEs/IMEs
210 and prophages of *A. baumannii* and *S. aureus*. Overall, the distribution of ESKAPE widespread
211 proteins clustered MGE/ESKAPE pairs by MGE type (**Figure 4**). The clustering observed from the
212 distribution of MGE widespread proteins was more intricate but tended towards the same
213 ESKAPE pathogen, with the exception of *Enterobacter* sp. and *K. pneumoniae*, which were
214 typically clustered together.

215

216 **4. AMR genes are overrepresented in the ESKAPE mobilome**

217 In order to explore the AMR repertoire of the ESKAPE mobilome, we only focused on genes that
218 are horizontally transferred, such as beta-lactamases and aminoglycoside-modifying enzymes
219 (that lead to antimicrobial inactivation), and those that promote target site modification (such
220 as rRNA methyltransferases, the *vanA* and *vanB* gene clusters, and the staphylococcal cassette
221 chromosome *mec*). We observed that AMR genes are broadly distributed in plasmids and
222 ICEs/IMEs across the ESKAPE pathogens (**Figure S10**). All ESKAPE pathogens have a large
223 proportion of AMR-carrying plasmids (>35% of plasmids across the different ESKAPE carry at
224 least one AMR gene), while a high proportion of AMR-harbouring ICEs (>25%) was only observed
225 for *S. aureus* and *P. aeruginosa* (**Figures S11A and S11B**). As expected from the vast repertoire
226 of MGEs present in *K. pneumoniae* (**Figure 1D**), this species presented a wider selection of
227 different AMR genes. Some AMR genes were restricted to specific ESKAPE pathogens, while
228 others were more promiscuous. For example, different beta-lactamases (*bla* genes) were
229 prevalent among proteobacterial representatives of the ESKAPE pathogens, but were mostly

230 absent from *S. aureus* and *E. faecium* (**Figure S10**). The only exception was the *blaZ* gene, which
231 was exclusively identified in plasmids and ICEs/IMEs from *S. aureus*. This gene is typically
232 embedded within the SCCmec elements of this species and may have been acquired from
233 distantly related non-*Staphylococcus* species [49]. Genes encoding resistance to aminoglycosides
234 (*aac*, *ant*, and *aph* genes), chloramphenicol (*cat* genes), trimethoprim (*dfr* genes), and
235 tetracyclines (*tet* genes) were found in all representatives of the ESKAPE pathogens. Genes
236 involved in resistance to vancomycin (the *vanHAX* and *vanHBX* gene clusters) were exclusively
237 found in *S. aureus* and *E. faecium*, while genes coding resistance for quinolones (*qnr* genes) and
238 colistin (*mcr* genes) were only found in the proteobacterial representatives (**Figure S10**).

239 We next assessed the distribution of virulence genes. These genes are more prevalent in
240 ICEs/IMEs than in plasmids (**Figure S12**). In fact, we identified no virulence genes in *E. faecium*
241 plasmids, and only 0.6% of *A. baumannii* plasmids carry these genes. More than 25% ICEs/IMEs
242 in *S. aureus*, *K. pneumoniae*, and *E. faecium* carried at least one virulence gene (**Figures S11C**
243 **and S11D**). *K. pneumoniae* is again the ESKAPE pathogen carrying a wider variety of genes in
244 both ICEs/IMEs and plasmids. As previously reported [50], we noticed that virulence and AMR
245 genes are rarely encoded by prophages. An important exception is the enterotoxins found in
246 prophages from *S. aureus* (**Figure S12**), which are also known as pyrogenic toxin superantigens
247 (PTSAgs) [51]. Polyketide synthesis loci *ybt* and *clb* encoding the iron-scavenging siderophore
248 yersiniabactin and genotoxin colibactin, respectively, were solely identified in
249 *Enterobacteriaceae* representatives of the ESKAPE pathogens (i.e., *K. pneumoniae* and
250 *Enterobacter* sp.). These virulence loci were mostly present in ICEs/IMEs, as previously reported
251 [52], but we also found these genes on plasmids (**Figure S12**). Interestingly, *S. aureus* was the
252 ESKAPE pathogen with a higher proportion of both plasmids and ICEs/IMEs carrying at least one
253 AMR or virulence genes (**Figure S11**).

254 Since chromosomes are substantially larger than MGEs and consequentially have more genes,
255 we corrected the prevalence of AMR and virulence genes to the total number of genes present
256 in MGEs and masked genomes across the different ESKAPE pathogens. Overall, we noticed that
257 AMR genes were largely overrepresented in MGEs (~14x), when compared with masked
258 genomes (**Figure 5 and Table S5**). On the other hand, virulence genes were ~2x more likely to
259 be located on masked genomes. Taken together, our results show that plasmids are important
260 vectors for AMR genes across the ESKAPE pathogens, while ICEs/IMEs play a more important
261 role for the distribution of virulence genes. When compared with masked genomes, we
262 confirmed that AMR genes are preferentially distributed in the ESKAPE mobilome.

263

264 **5. CRISPR-Cas systems shape the number of MGEs, AMR, and virulence genes across**
265 **ESKAPE pathogens**

266 CRISPR-Cas systems were identified in every ESKAPE pathogen except *E. faecium*, which was
267 therefore excluded from subsequent analyses. The proportion of genomes with CRISPR-Cas
268 systems varied across the ESKAPE pathogens, from around 45.7% for *P. aeruginosa* to around
269 0.7% for *S. aureus* (**Figure 6A** and **Table S2**). We then explored the prevalence of CRISPR-Cas
270 systems across closely related strains belonging to the same MLST profile. Since *Enterobacter*
271 sp. consists of multiple species, this ESKAPE pathogen was excluded from subsequent analyses.
272 Given the low prevalence of CRISPR-Cas systems in *S. aureus*, this species was also excluded, and
273 we focused exclusively on *P. aeruginosa*, *K. pneumoniae*, and *A. baumannii*. Interestingly, some
274 sequence types (ST) consisted entirely of either CRISPR-Cas positive or negative genomes (**Table**
275 **S2**). For example, the most frequent MLST profile in *A. baumannii* from our dataset was ST2
276 (n=101), which only included strains with no CRISPR-Cas systems. On the other hand, the second
277 most prevalent MLST profile in this species (ST1, n=14), only consisted of strains with I-F CRISPR-
278 Cas systems. The most frequently observed MLST profile from *K. pneumoniae* (ST11, n=105),
279 consists mostly of CRISPR-Cas negative strains (96.2%, 101/105). The four strains with positive
280 hits carried IV-A3 CRISPR-Cas systems on plasmids. ST258 (n=47) was the second most common
281 *K. pneumoniae* MLST profile identified in our dataset, and again, consisted entirely of strains
282 with no CRISPR-Cas systems. However, ST147 and ST15 (n=31 and n=23, respectively) carried I-
283 E CRISPR-Cas systems in all strains. Finally, looking at *P. aeruginosa*, we found that the most
284 prevalent MLST profiles in our dataset (ST235 and ST549, n=16 and n=11, respectively) carried
285 no CRISPR-Cas systems. The only exception was found in a ST235 strain, which carried an I-C
286 CRISPR-Cas system on an ICE/IME. In contrast, ST233 and ST1971 (n=8 for both) consisted
287 exclusively of strains carrying the I-F CRISPR-Cas system on masked genomes (**Table S2**). These
288 findings suggest that the presence or absence of CRISPR-Cas systems across the ESKAPE
289 pathogens is largely driven by the expansion of specific clones.

290 Our analysis revealed a large variety of MGE-encoded CRISPR-Cas subtypes, with I-C, I-E, I-F, III-
291 A, IV-A1, and IV-A3 represented across the dataset (**Figure 6B**). We found CRISPR-Cas systems
292 on plasmids (n=28 IV-A3 subtype in *K. pneumoniae* and n=1 IV-A1 in *P. aeruginosa*) and
293 ICEs/IMEs (n=7 I-C in *P. aeruginosa* and n=1 III-A in *S. aureus*) (**Figure 6B** and **Table S6**). The MGEs
294 carrying these systems were large, ranging from 102-430 kb. Additionally, all CRISPR-Cas-
295 carrying plasmids have a MOB relaxase. This is in agreement with previous findings [53], which

296 reported an enrichment of CRISPR-Cas systems across plasmids with conjugative functions and
297 of larger sizes. We also found AMR and virulence genes on these CRISPR-Cas positive MGEs, but
298 no anti-CRISPRs, suggesting that the CRISPR-Cas systems are functional.

299 When looking into the influence of GC content and sequence length in pairs of conspecific
300 ESKAPE pathogens with and without CRISPR-Cas systems, we would expect to observe smaller
301 and GC richer strains on those carrying these systems. These expectations could only be
302 confirmed for *P. aeruginosa*, for which CRISPR-Cas positive genomes were significantly smaller
303 than their counterparts (**Figure S13**, p-value 0.00037), as observed before [54–56]. Still, the non-
304 significant differences observed for *A. baumannii*, *Enterobacter* sp., *Klebsiella pneumoniae*, and
305 *S. aureus* could in part be explained by the low sample size of CRISPR-Cas positive genomes
306 (**Figure 6A**). Regarding the GC content, we observed significant differences in *A. baumannii*, *K.*
307 *pneumoniae*, and *P. aeruginosa*. CRISPR-Cas positive genomes were GC richer in *A. baumannii*
308 and *P. aeruginosa* (**Figure S14**, p-values 0.00098 and 0.00673, respectively). Curiously, we
309 noticed that CRISPR-Cas positive genomes in *K. pneumoniae* were GC poorer (**Figure S14**, p-value
310 2.5e-05). Given the known association between GC content and genome size [44], these
311 differences in CRISPR-Cas positive and negative *P. aeruginosa* genomes may be a spurious
312 correlation driven by small size of CRISPR-Cas positive genomes. So, we corrected the GC content
313 for variation in genome size, and we observed that the association was maintained (**Figure S15**,
314 p-value 0.00042), in accordance to a previous study [55].

315 Since virulence genes are overrepresented in the chromosome (**Figure 5**), we assessed the
316 distribution of these genes in pairs of conspecific ESKAPE pathogens with and without CRISPR-
317 Cas systems. Virulence genes were significantly less abundant in CRISPR-Cas positive genomes
318 from *P. aeruginosa* and *A. baumannii* (**Table S7** and **Figure S16**, p-values 1.7e-06 and 0.0016,
319 respectively). Given that *P. aeruginosa* genomes positive for these systems are significantly
320 smaller than their CRISPR-Cas negative counterparts (**Figure S13**), the lower prevalence of these
321 genes in CRISPR-Cas positive *P. aeruginosa* genomes may again be driven by a spurious
322 correlation. As so, we corrected the number of virulence genes for variation in genome size, and
323 we observed that indeed the difference was no longer significant (**Figure S17**, p-value 0.73),
324 confirming our prediction that the genome size was a confounding variable obscuring the effect
325 of CRISPR-Cas systems on the prevalence of virulence genes in *P. aeruginosa*.

326 We then explored whether CRISPR-Cas presence or absence reduced the number of MGEs
327 acquired in pairs of conspecific ESKAPE pathogens. We would expect to detect a smaller number
328 of MGEs in genomes harbouring these immune systems. Indeed, this trend was observed in *K.*

329 *pneumoniae* and *P. aeruginosa* (**Figure 7**). Conversely, CRISPR-Cas positive *A. baumannii*
330 genomes were significantly larger than genomes of the same species lacking these systems (p-
331 value 0.047). Finally, we focused on AMR and virulence genes carried exclusively by plasmids
332 and ICEs, as these were the most important vectors (**Figure S10**). For most MGE/ESKAPE pairs,
333 we observed no significant difference between pairs of conspecific genomes with and without
334 CRISPR-Cas systems. When it comes to AMR genes, we only observed significant differences in
335 MGEs from *P. aeruginosa* (**Figure S18A**, p-values 0.037). Indeed, AMR genes were more
336 prevalent on ICEs/IMEs from *P. aeruginosa* genomes with CRISPR-Cas systems (**Table S8**). A
337 similar correlation was previously reported [56]. Curiously, the less prevalent I-C CRISPR-Cas
338 subtype, which was exclusively identified in *P. aeruginosa* and mostly on ICEs/IMEs (**Table S2**),
339 was found to be positively correlated with certain AMR genes [54]. Regarding virulence genes, we
340 observed significant differences in MGEs from *K. pneumoniae*, where CRISPR-Cas-carrying
341 elements were associated with more virulence genes than their CRISPR-Cas negative
342 counterparts (**Figure S18B**, p-values 0.0054). Taken together, we observed species-specific
343 trends shaping the number of MGEs, AMR, and virulence genes across pairs of conspecific
344 ESKAPE genomes with and without CRISPR-Cas systems.

345

346 **6. Anti-CRISPRs are overrepresented in the ESKAPE mobilome**

347 Anti-CRISPR proteins (n=161) antagonising CRISPR-Cas subtypes I-E, I-F, and II-A were identified
348 in a total of 90 *P. aeruginosa*, *K. pneumoniae*, and *A. baumannii* MGEs (**Figure 6C** and **Table S9**).
349 ICEs/IMEs were the most important vectors for these counter-defence systems in *P. aeruginosa*,
350 while prophages were more frequent in *K. pneumoniae*. Even though *P. aeruginosa* was the
351 ESKAPE pathogen with a higher proportion of genomes with CRISPR-Cas systems (**Figure 6A**), we
352 observed Anti-CRISPR proteins on 28.3 % (n=34/120) CRIPR-Cas positive genomes from this
353 species (**Table S2**). We also looked for these proteins across the masked genomes, and found
354 hits in all ESKAPE except *S. aureus* (n=252, **Table S10**). After correcting their prevalence to total
355 number of genes, we verified that anti-CRISPRs are largely overrepresented in MGEs (~9x) when
356 compared with masked genomes (**Figure 5)**. When compared with masked genomes, our results
357 show that Anti-CRISPR proteins are preferentially encoded in MGEs.

358

359 **7. CRISPR spacers in ICEs/IMEs and plasmids have different targeting biases**

360 We explored the targets for all CRISPR spacers, retrieved from complete CRISPR-Cas systems,
361 but also orphan CRISPRs, identified in our collection of ESKAPE genomes. Since we provided here
362 a representative dataset of plasmids, ICEs/IMEs, and prophages, we used this collection as a
363 database and took the CRISPR spacers identified in masked genomes as a query. In parallel, we
364 used the CRISPR spacers identified in MGEs as a query and the MGEs collection as a database.
365 For the latter, and to avoid self-targeting hits, we masked all CRISPR spacers from the MGEs
366 collection used as database. We found a total of 897 CRISPR spacers across the MGEs (n=554 on
367 plasmids and n=343 on ICEs/IMEs, **Table S11**). Only a small fraction of plasmids and ICEs/IMEs
368 carried these spacers (1.3%, 34/2640 plasmids and 0.6%, 17/2685 ICEs/IMEs, respectively). No
369 CRISPR spacers were found on prophages, which is in agreement with the lack of CRISPR-Cas
370 systems in these elements from our collection. Given the large number of MGEs and CRISPR-
371 Cas-encoding plasmids in *K. pneumoniae* (**Figure 1** and **Figure 6B**), most spacers were found in
372 this species (522/897). The large majority of plasmid spacers were identified on mobilizable
373 plasmids (99.4%, 551/554).

374 We then looked for protospacers, and identified invading DNA fragments on 9.2% of total MGEs
375 (n=641/6941, **Table S12**). A substantially larger fraction of CRISPR spacers from plasmids
376 targeted mobilizable plasmids from our ESKAPE collection (92.6%, 21488/23200 of total plasmid
377 spacer's interactions). Only a small fraction of plasmid spacers targeted protospacers in non-
378 transmissible plasmids (5.2%), prophages (1.6%), and ICEs/IMEs (0.6%) (**Figure 8A**). Moreover,
379 ICEs/IMEs spacers were also biased towards other ICEs/IMEs (70.7%, 795/1124 of total
380 ICEs/IMEs interactions). However, unlike plasmids, a larger proportion of ICEs/IMEs spacers
381 targeted prophages (27.7%) (**Figure 8A**). Only a small fraction of ICEs/IMEs spacers targeted
382 plasmids (1.6%).

383 When it comes to CRISPR arrays on masked genomes, we observed a total of 13709 spacers in
384 30.8% ESKAPE genomes (n=538/1746, **Table S13**). Even though the prevalence of CRISPR-Cas
385 systems in *S. aureus* was considerably low (< 1%, **Figure 6A**), we found spacers in 31.4% of the
386 total *S. aureus* strains (n=174/553). Consistent with the absence of CRISPR-Cas systems in the *E.*
387 *faecium* strains from our dataset, no spacers were identified in masked genomes from this
388 species. The number of CRISPR spacers per genome varied considerably between the masked
389 genomes of the ESKAPE pathogens (**Figure S19**), reaching as high as 189 in *A. baumannii*. In fact,
390 only strains from this species carried more than 100 spacers per genome. We observed that
391 27.8% (3815/13709) of CRISPR spacers in ESKAPE masked genomes yielded significant matches
392 to protospacer MGE sequences, found on 25.0% MGEs (n=1735/6941, **Table S14**). Most CRISPR
393 spacers targeted plasmids (72.8%, 84256/115743 total interactions), followed by ICEs/IMEs

394 (16.9%) and prophages (10.3%). As observed for CRISPR spacers in MGEs, most plasmid targets
395 correspond to mobilizable plasmids (88.5%, 74567/84256). We then tested if MGE or masked
396 genome spacers preferentially targeted ESKAPE MGEs of variable size. We found that the CRISPR
397 spacer targets were significantly larger than those with no protospacers (p-value <2.2e-16**,**
398 **Figures 8B and 8C**). Altogether, our results show that ICEs/IMEs preferentially target ICEs/IMEs
399 and also prophages, while plasmids almost exclusively target other plasmids. Our data also
400 shows that CRISPR spacers either on MGEs or masked genomes consistently target larger MGEs.

401

402 **Discussion**

403 In this work, we performed a systematic analysis of ICEs/IMEs, plasmids, and prophages across
404 all ESKAPE pathogens. We focused on this panel because the ESKAPE group of pathogens consist
405 of clinically-relevant bacteria, for which many genomes are completely sequenced (an important
406 parameter when delineating MGEs), and which include representatives of both Proteobacteria
407 and Firmicutes, and also phylogenetically divergent bacteria (with exception of *K. pneumoniae*
408 and *Enterobacter* sp., the remaining representatives belong to different bacterial families).
409 Studying MGEs in parallel allowed us to explore their uneven distribution across a collection of
410 complete genomes from important pathogens, and to explore DNA sharing events between
411 different MGE types. By separating these elements from masked (MGE-free) genomes, we were
412 able to observe an overrepresentation of AMR genes and anti-CRISPRs across the ESKAPE
413 mobilome. Furthermore, we assessed the influence of CRISPR-Cas defence systems in shaping
414 the acquisition of MGEs and beneficial genes, and we unveiled different targeting biases for
415 CRISPR spacers identified in plasmids and ICEs/IMEs.

416 The network-based approach used here to study the ESKAPE MGEs revealed a clear structural
417 differentiation, where the majority of clusters were homogeneous for a given ESKAPE/MGE pair.
418 Using pairwise genetic distances of alignment-free *k*-mer sequences has circumvented the
419 exclusion of non-coding elements that was observed in gene content similarity networks from
420 previous work [57], providing a more comprehensive picture of plasmid population and dynamics
421 [12]. Other groups have shown that plasmids form coherent clusters [12,13], similar in concept to
422 what was observed for bacterial genomes [45,46,58]. Focusing on the ESKAPE pathogens, we
423 demonstrated here that the same happens for ICEs/IMEs and prophages for the majority of the
424 clusters. However, two large heterogeneous clusters were also observed. Unlike bacterial
425 genomes, where recombination between closely related replicons is the main force promoting
426 genomic cohesiveness [45,58], MGEs such as plasmids, ICEs/IMEs and prophages are capable of

427 exchanging DNA beyond the genus barrier (and even occasionally across phylum) [59]. Based on
428 our network-based approach, we observed that the ESKAPE mobilome appears to follow a
429 bipartite structure, with some elements being capable of shuffling DNA between distantly
430 related species.

431 The abundance of MGEs is strongly associated with the prevalence of AMR genes. This is
432 particularly evident for *K. pneumoniae*, which carries a high proportion of important vectors of
433 AMR genes, such as plasmids and ICEs/IMEs. After correcting the prevalence to total number of
434 genes, we observed that AMR genes are nearly 15 times more likely to be found on ESKAPE
435 MGEs than on masked genomes. Most likely the use of different antibiotics targeting either
436 Gram negative or positive infections may have selected for the emergence of different AMR
437 genes across the Proteobacteria/Firmicutes divide. For example, vancomycin is used as last
438 resort for the treatment of sepsis and other infections caused by Gram-positive bacteria, while
439 colistin mainly serves to target multi-drug resistant Gram-negative infections [26]. Anti-CRISPRs
440 are nearly 10 times more likely to be detected on these MGEs than on masked genomes. Unlike
441 AMR genes, where genes conferring resistance to similar antibiotics where identified in different
442 ESKAPE pathogens, the distribution of virulence genes across the ESKAPE mobilome was mostly
443 species-specific. This may be explained by different mechanisms of virulence and toxicity across
444 bacterial species. Even considering the relative proportion of these genes, we confirmed that
445 these genes are twice more likely to be identified in masked genomes than in MGEs.

446 Curiously, we found no CRISPR-Cas systems in our curated *E. faecium* dataset. In addition to our
447 main analysis, we specifically searched for the presence of these systems on the excluded
448 genomes, based on <95% genome similarity threshold defined for species delineation [58], and
449 found three strains with these systems (**Table S15**). Since the majority of the most
450 representative MLST profiles in our collection consists of genomes either with or without
451 CRISPR-Cas systems (**Table S2**), analysis of intra-ST CRISPR variability between pairs of
452 conspecific ESKAPE genomes was not performed in this study. Except for *P. aeruginosa* [54,55], we
453 found no evidence for genome length as a marker for HGT inhibition by CRISPR-Cas systems. A
454 similar observation was made before, for a different collection of bacterial pathogens [56].

455 Defence systems such as CRISPR-Cas are inherently costly to bacterial hosts, mainly due to
456 different forms of autoimmunity [60]. To offset the short-term deleterious effects, these systems
457 benefit from associating with MGEs, such as the examples observed here and elsewhere [38].
458 Although these and other defence mechanisms such as restriction-modification systems do not
459 qualify as bona fide MGEs (the systems lack mechanisms controlling their own replication), these

460 quasi-autonomous systems take advantage of MGEs to promote their own dissemination and
461 maintenance across bacterial hosts. Conversely, MGEs benefit from these systems and may
462 pervasively repurpose them for inter-MGE competition [53]. In fact, we found that spacers in the
463 ICE/IME and plasmid CRISPR arrays target competing MGEs, underscoring the genetic
464 independence of CRISPR-Cas systems in ICEs/IMEs and plasmids in MGE-MGE conflicts. Even
465 though we found no CRISPR-Cas systems in prophages, a phage-encoded system was identified
466 in *Vibrio cholerae*, and targeted phage-inducible chromosomal islands [61]. Importantly, the
467 presence of CRISPR-Cas subtypes preferentially distributed in MGEs (I-C and IV-A3) points to the
468 existence of distinct selective pressure that promote the maintenance of specific subtypes on
469 plasmids and ICEs/IMEs versus masked genomes. Unsurprisingly given the large sizes of these
470 systems, we observed a bias towards larger MGEs (both plasmids and ICEs/IMEs >100kb). Since
471 plasmids with an identifiable relaxase (conjugative and mobilizable) are larger than the so-called
472 non-transmissible plasmids [41], we also found a relaxase in all CRISPR-Cas positive plasmids.
473 Prophages seem to follow similar streamlining dynamics [62]. The high proportion of *P. aeruginosa*
474 genomes carrying both CRISPR-Cas systems and anti-CRISPRs (28.3%, **Table S2**) suggests a large
475 fraction of these systems may be non-functional in this species, as hypothesized before [54]. Since
476 most Anti-CRISPRs in *P. aeruginosa* were identified in ICES/IMEs, these results underscore a
477 major role played by the overlooked ICEs/IMEs in potentially silencing immune systems in this
478 species.

479 Our results shows that only a restricted fraction of CRISPR spacers matched protospacer MGE
480 sequences, which is in agreement with previous findings [53,55]. While the large majority of
481 plasmid-encoded spacers were predicted to target other plasmids, CRISPR arrays in ICEs/IMEs
482 preferentially targeted other ICEs/IMEs, but also a large proportion of prophages. This
483 complementary targeting preference can be explained by the different lifestyle of these MGEs.
484 Since plasmids are maintained as extra-chromosomal elements, and ICEs/IMEs and prophages
485 are integrated in the chromosome, we hypothesize that while plasmids preferentially target
486 plasmid competitors [38], ICEs/IMEs exploit CRISPR-Cas systems to protect their host against other
487 ICEs/IMEs and viral predators.

488 Several bioinformatic tools exist to look for plasmids and prophages, but currently the options
489 for ICEs/IMEs are scarce [63]. We provide here an accurate identification of these elements,
490 building upon a recently reported tool to scan RGPs [64]. However, our approach depends on the
491 availability of a pangenome for the considered taxa, which is an important limitation for species
492 with an insufficient number of completely sequenced genomes. Moreover, our work only
493 focused on three MGE types, which are likely to be of main importance for the dissemination of

494 genes involved in pathogenesis and AMR and thus critical for our understanding of the evolution
495 of the ESKAPE pathogens. Nevertheless, other MGEs should be considered in future work, for
496 example those that contribute to communication between cells (such as phage-inducible
497 chromosomal islands [65]), or intracellular transfer events (e.g., transposons and insertions
498 sequences [35]).

499 To conclude, our results indicate that ICEs/IMEs, plasmids, and prophages are asymmetrically
500 distributed across the ESKAPE pathogens. We found that these MGEs can exchange DNA beyond
501 the genus (and phylum) barrier, even though most clusters are constrained by host similarity
502 and/or the type of MGE. We observed that most ESKAPE MGE proteins have unknown functions
503 and have convoluted distribution patterns. When comparing against masked (MGE-free)
504 genomes, we observed the pervasive association of AMR genes and anti-CRISPRs with the
505 ESKAPE mobilome. We also found different targeting biases to CRISPR spacers found on plasmids
506 and ICEs/IMEs, highlighting their genetic independence. Taken together, our results illustrate
507 the potential of network-based approaches and comparative genomics to underscore the
508 composition and dynamics of gene flow across different MGEs, and sheds a new light in the role
509 of the overlooked ICEs/IMEs as important players in the MGE-MGE warfare in the ESKAPE
510 pathogens and thus the main groups of highly problematic human pathogens.

511

512 **Material and methods**

513 **ESKAPE pathogens collection**

514 We retrieved all complete ESKAPE genomes available in the NCBI Reference Sequence Database
515 (RefSeq, accessed on 12-11-2020), using ncbi-genome-download v0.3.0
516 (https://github.com/kblin/ncbi-genome-download). Genomes listed as 'unverified' were
517 removed from our dataset. We also excluded genomes classified as 'Enterobacteriaeceae'.
518 Finally, we used pyANI v0.2.10 (https://github.com/widdowquinn/pyani) to calculate the
519 average nucleotide identity based on MUMmer (ANIm), and removed genomes with an ANIm
520 value below the 95% threshold for species delineation [58,66]. To evaluate the taxonomy of the
521 *Enterobacter* species, we retrieved genomes for *Enterobacteriaceae* type strains and used them
522 together with the *Enterobacter* genomes to create a phylogenetic tree using GToTree v1.5.22
523 (https://github.com/AstrobioMike/GToTree)[67] and the IQ-TREE algorithm to estimate maximum
524 likelihood [68]. We used the pre-built set of 74 single copy gene bacterial Hidden Markov Models
525 (HMM) available in GToTree. Genomes labelled as belonging to the *Enterobacter* genus, but
526 clustered in the phylogenomic tree with type strains other than those from the *Enterobacter*

527 genus, were removed from subsequent analyses. MLST profiles were determined with mlst
528 v2.19.0 (https://github.com/tseemann/mlst). The curated genomes were automatically
529 annotated using Prokka v1.14.6 (https://github.com/tseemann/prokka) [69].

530

531 **Extraction of plasmids, ICEs/IMEs, and prophages**

532 Since the ESKAPE pathogens (as most bacteria) are haploid, we separated the large replicon (i.e.,
533 the chromosome) from the extrachromosomal replicons. For the latter, only accessions with
534 'plasmid' and 'complete sequence' on their description were kept and were used for further
535 plasmid analyses.

536 To extract ICEs from chromosomal replicons, we used the chromosomal genbank files created
537 with Prokka as input to build a pangenome for each ESKAPE pathogen. We used ppanggolin
538 v1.1.96 (https://github.com/labgem/PPanGGOLiN)[70], which uses a graphical model and a
539 statistical method to partition the pangenome in persistent, shell and cloud genomes. We used
540 the panRGP method to build the pangenomes [64]. This method predicts RGPs, which are clusters
541 of genes made of accessory genes (shell and cloud genomes) in the pangenome graph. We then
542 used bedtools v2.30.0 (https://bedtools.readthedocs.io/en/latest/) [71] to extract RGPs from the
543 chromosomal replicons. The proteomes of these extracted RGPs were scanned for relaxases
544 with hmmer v.3.3.1 (http://hmmer.org/) [72] against MOBfamDB, a curated relaxase profile HMM
545 database [73]. Simultaneously, the proteomes were screened with hmmer against integrases
546 (Pfam accession PF00589) and recombinases (PF07508). RGPs with hits both for integrases
547 (phage integrase or recombinase) and relaxases were classified as putative ICEs/IMEs and were
548 kept for further analysis.

549 To look for prophages, we masked the ICEs/IMEs locations in the chromosomal replicons using
550 bedtools. The ICE-masked chromosomal genbank files were then used as input in phispy v4.2.6
551 (https://github.com/linsalrob/PhiSpy), which combines similarity- and composition-based
552 strategies to look for prophages [74]. We also masked the prophage regions in the chromosomal
553 replicons, to build the final masked genomes, that are ICE- and prophage-masked (these masked
554 replicons are also free of plasmids, since these are part of the extrachromosomal replicons).

555

556 **Network-based approach and functional annotation**

557 To estimate pairwise distances between all ESKAPE MGEs (plasmids, ICEs and prophages), we
558 reduced the MGEs into sketches and compared the Jaccard index (JI) and mutation distances
559 between pairs of MGEs using BinDash v 0.2.1 (https://github.com/zhaoxiaofei/bindash)[75]. Each
560 MGE sequence was converted to a set of 21-bp *k*-mers. We used the mean() and median()
561 functions in R to calculate the arithmetic mean and median of the JI, respectively. Only JI equal
562 to or above the mean and median were considered, and the mutation distances were used as
563 edge attributes to plot the network with Cytoscape v3.9.0 under the prefuse force directed
564 layout (https://cytoscape.org/). We used the Analyzer function in Cytoscape to compute a
565 comprehensive set of topological parameters, such as the clustering coefficient, the network
566 density, the centralization, and the heterogeneity.

567 MGE proteins assignment to COGs was retrieved from the tab-separated annotation files
568 produced by Prokka. Lists of non-redundant COGs were extracted separately for prophages,
569 plasmids and ICEs/IMEs and compared with venny v2.1.0
570 (https://bioinfogp.cnb.csic.es/tools/venny/index.html) to identify unique and shared COGs.
571 Likewise, COGs occurrence was determined separately for the proteomes of the three MGE
572 types in the different ESKAPE. Information on the COGs classification into functional categories
573 was retrieved from https://ftp.ncbi.nih.gov/pub/COG/COG2020/data/. The relative frequency
574 of the different COG functional categories per MGE/ESKAPE pair was calculated by summing up
575 the occurrences of COGs belonging to a given functional category and dividing the resulting
576 number by the total number of proteins observed in the corresponding MGE/ESKAPE pair.

577 To explore the diversity of MGE-encoded proteins we combined their proteomes (474223
578 proteins) and clustered them using the cluster algorithm from the MMSEQseqs2 v13.45111
579 package [76]. The proteins were clustered at 80% sequence identity and 80% coverage, with
580 otherwise default settings to match the parameters used by ppanggolin when generating the
581 ESKAPE pangenomes. The relative frequency of the different protein clusters identified per
582 MGE/ESKAPE pair was calculated following the same approach used to estimate the relative
583 frequency of COG functional categories, but using the occurrence of proteins belonging to a
584 given cluster instead. Representatives of the 51960 clusters identified were annotated with
585 eggNOG-mapper v2 [77], with default settings to further explore the functions of the MGE-
586 encoded proteins.

587 We used abricate v1.0.1 (https://github.com/tseemann/abricate) to scan extracted MGEs and
588 masked genomes against antimicrobial resistance and virulence genes (using pre-downloaded
589 databases from Resfinder[78] and VFDB[79] containing 3108 and 3686 sequences, respectively, and

590 both updated on the 05-02-2021). To identify and classify CRISPR-Cas systems, we used
591 CRISPRCasTyper v1.2.3 (https://github.com/Russel88/CRISPRCasTyper)[80]. We also used this tool
592 to look for CRISPR spacers. The entire CRISPR arrays identified on MGEs were then masked using
593 bedtools, and these masked MGEs served as a local blast database, when using MGE CRISPR
594 spacers as a query. CRISPR spacers from masked genomes were also blasted against a local
595 database of our extracted (non-masked) MGEs. Hits with at least 95% nucleotide identity and
596 95% sequence coverage were considered as protospacers [81]. While a representative collection
597 of plasmids and virus is publicly available at RefSeq's NCBI database (n=33269 and 13778,
598 respectively, accessed on the 21/05/2021), a substantially smaller collection of ICEs/IMEs is
599 available at ICEberg database (n=1325), and was last updated in September 2018. Due to this
600 limitation in the number of publicly available ICEs/IMEs sequences, and to have a representative
601 collection of these three different types of MGEs, we focused on the curated dataset presented
602 in this study to look for targets of CRISPR spacers.

603 We retrieved an Anti-CRISPR collection of non-redundant proteins from Anti-CRISPRdb
604 (http://guolab.whu.edu.cn/anti-CRISPRdb/, accessed on the 19-02-2021). This collection was
605 used to build a local database with diamond v2.0.9 (https://github.com/bbuchfink/diamond)[82].
606 We used the blastp command in diamond to scan the MGEs and masked proteomes against the
607 anti-CRISPR local database. We used an amino acid-based homology approach to find Anti-
608 CRISPRs encoded in the ESKAPE mobilome. Even though recent approaches have applied a guilt-
609 by-association method to identify new Anti-CRISPRs [25], currently there is no tool available to
610 apply this method in a large dataset of bacterial genomes.

611

612 **Statistical analysis**

613 Comparisons between MGEs' GC content and sequence length were performed using the
614 Kruskal-Wallis test, and the p-values adjusted using the Holm–Bonferroni method. Comparisons
615 between pairs of conspecific genomes with and without CRISPR-Cas systems, as well as between
616 MGE targets for CRISPR spacers, were performed using the Wilcoxon test, and the p-values
617 adjusted using the Holm–Bonferroni method. Values above 0.05 were considered as non-
618 significant (ns). We used the following convention for symbols indicating statistical significance:
619 * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001.

620

621 **Code availability**

622 All code used to create the main figures are available at

623 https://gitlab.gwdg.de/botelho/eskape_mobilome).

624

625 **References**

626 1. Ghaly, T. M. & Gillings, M. R. Mobile DNAs as Ecologically and Evolutionarily
627 Independent Units of Life. *Trends Microbiol.* **26**, 904–912 (2018).

628 2. Koonin, E. V. Viruses and mobile elements as drivers of evolutionary transitions. *Philos.*
629 *Trans. R. Soc. B Biol. Sci.* **371**, (2016).

630 3. Kazazian, H. H. Mobile Elements: Drivers of Genome Evolution. *Science* **303**, 1626–1632
631 (2004).

632 4. San Millan, A. & MacLean, R. C. Fitness Costs of Plasmids: a Limit to Plasmid
633 Transmission. *Microbiol. Spectr.* **5** (2017).

634 5. Botelho, J. & Schulenburg, H. The Role of Integrative and Conjugative Elements in
635 Antibiotic Resistance Evolution. *Trends Microbiol.* **29**, 8-18 (2020).

636 6. Chen, J. *et al.* Genome hypermobility by lateral transduction. *Science* **362**, 207–212
637 (2018).

638 7. Dagan, T., Artzy-Randrup, Y. & Martin, W. Modular networks and cumulative impact of
639 lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**,
640 10039–10044 (2008).

641 8. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the
642 human microbiome. *Nature* **480**, 241–244 (2011).

643 9. Ellabaan, M. M. H., Munck, C., Porse, A., Imamovic, L. & Sommer, M. O. A. Forecasting
644 the dissemination of antibiotic resistance genes across bacterial genomes. *Nat.*
645 *Commun.* **12**, 2435 (2021).

646 10. Popa, O., Hazkani-Covo, E., Landan, G., Martin, W. & Dagan, T. Directed networks reveal
647 genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes.
648 *Genome Res.* **21**, 599–609 (2011).

649 11. Kloesges, T., Popa, O., Martin, W. & Dagan, T. Networks of Gene Sharing among 329
650 Proteobacterial Genomes Reveal Differences in Lateral Gene Transfer Frequency at
651 Different Phylogenetic Depths. *Mol. Biol. Evol.* **28**, 1057–1074 (2011).

652 12. Acman, M., van Dorp, L., Santini, J. M. & Balloux, F. Large-scale network analysis
653 captures biological features of bacterial plasmids. *Nat. Commun.* **11**, 2452 (2020).

654 13. Redondo-Salvo, S. *et al.* Pathways for horizontal gene transfer in bacteria revealed by a
655 global map of their plasmids. *Nat. Commun.* **11**, 3602 (2020).

656 14. Campbell, D. E. *et al.* Interrogation of the integrated mobile genetic elements in gut-
657 associated Bacteroidaceae with a consensus prediction approach. *bioRxiv*
658 2021.09.02.458807 (2021) doi:10.1101/2021.09.02.458807.

659 15. Croucher, N. J. *et al.* Diversification of bacterial genome content through distinct
660 mechanisms over different timescales. *Nat. Commun.* **5**, 1–12 (2014).

661 16. Wein, T., Hülter, N. F., Mizrahi, I. & Dagan, T. Emergence of plasmid stability under non-

662 selective conditions maintains antibiotic resistance. *Nat. Commun.* **10**, 2595 (2019).

663 17. Rodriguez-Beltran, J. *et al.* Multicopy plasmids allow bacteria to escape from fitness
664 trade-offs during evolutionary innovation. *Nat. Ecol. Evol.* **2**, 873–881 (2018).

665 18. Koonin, E. V., Makarova, K. S., Wolf, Y. I. & Krupovic, M. Evolutionary entanglement of
666 mobile genetic elements and host defence systems: guns for hire. *Nat. Rev. Genet.* **21**,
667 119-131 (2019).

668 19. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction
669 and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43**, D298–D299
670 (2015).

671 20. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class
672 2 and derived variants. *Nat Rev Microbiol.* **18**, 67-83 (2020).

673 21. Millan, A. S. *et al.* Positive selection and compensatory adaptation interact to stabilize
674 non-transmissible plasmids. *Nat. Commun.* **5**, 5208 (2014).

675 22. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes
676 that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).

677 23. Li, Y. & Bondy-Denomy, J. Anti-CRISPRs go viral: the infection biology of CRISPR-Cas
678 inhibitors. *Cell Host Microbe.* **29**, 704–714 (2021).

679 24. Mahendra, C. *et al.* Broad-spectrum anti-CRISPR proteins facilitate horizontal gene
680 transfer. *Nat. Microbiol.* **5**, 620–629 (2020).

681 25. Pinilla-Redondo, R. *et al.* Discovery of multiple anti-CRISPRs highlights anti-defense
682 gene clustering in mobile genetic elements. *Nat. Commun.* **11**, 5652 (2020).

683 26. De Oliveira, D. M. P. *et al.* Antimicrobial resistance in ESKAPE pathogens. *Clin. Microbiol.*
684 *Rev.* **33**, e00181-19 (2020).

685 27. Mortensen, K., Lam, T. J. & Ye, Y. Comparison of CRISPR–Cas Immune Systems in
686 Healthcare-Related Pathogens. *Front. Microbiol.* **12**, 758782 (2021).

687 28. Wyres, K. L. *et al.* Genomic surveillance for hypervirulence and multi-drug resistance in
688 invasive *Klebsiella pneumoniae* from South and Southeast Asia. *Genome Med.* **12**, 11
689 (2020).

690 29. Botelho, J., Grosso, F. & Peixe, L. Antibiotic resistance in *Pseudomonas aeruginosa* –
691 Mechanisms, epidemiology and evolution. *Drug Resist. Updat.* **44**, 100640 (2019).

692 30. Davin-Regli, A., Lavigne, J.-P. & Pagès, J.-M. *Enterobacter* spp.: Update on Taxonomy,
693 Clinical Aspects, and Emerging Antimicrobial Resistance. *Clin. Microbiol. Rev.* **32**,
694 e00002-19 (2019).

695 31. Copin, R. *et al.* Sequential evolution of virulence and resistance during clonal spread of
696 community-acquired methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci.*
697 *U. S. A.* **116**, 1745–1754 (2019).

698 32. Gao, W., Howden, B. P. & Stinear, T. P. Evolution of virulence in *Enterococcus faecium*, a
699 hospital-adapted opportunistic pathogen. *Curr Opin Microbiol.* **41**, 76–82 (2018).

700 33. Harding, C. M., Hennon, S. W. & Feldman, M. F. Uncovering the mechanisms of
701 *Acinetobacter baumannii* virulence. *Nat Rev Microbiol.* **16**, 91–102 (2018).

702 34. World Health Organization. Global priority list of antibiotic-resistant bacteria to guide

703 research, discovery, and development of new antibiotics. http://www.who.int/news-
704 room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-
705 urgently-needed.

706 35. Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile Genetic Elements
707 Associated with Antimicrobial Resistance. *Clin. Microbiol. Rev.* **31**, e00088-17 (2018).

708 36. Botelho, J., Mourão, J., Roberts, A. P. & Peixe, L. Comprehensive genome data analysis
709 establishes a triple whammy of carbapenemases, ICEs and multiple clinically relevant
710 bacteria. *Microb. genomics* **6**, mgen000424 (2020).

711 37. Paauw, A., Leverstein-van Hall, M. A., Verhoef, J. & Fluit, A. C. Evolution in quantum
712 leaps: Multiple combinatorial transfers of HPI and other genetic modules in
713 Enterobacteriaceae. *PLoS One* **5**, 8662 (2010).

714 38. Pinilla-Redondo, R. *et al.* Type IV CRISPR-Cas systems are highly diverse and involved in
715 competition between plasmids. *Nucleic Acids Res.* **48**, 2000–2012 (2020).

716 39. León, L. M., Park, A. E., Borges, A. L., Zhang, J. Y. & Bondy-Denomy, J. Mobile element
717 warfare via CRISPR and anti-CRISPR in Pseudomonas aeruginosa. *Nucleic Acids Res.* **49**,
718 2114–2125 (2021).

719 40. Moya-Beltrán, A. *et al.* Evolution of Type IV CRISPR-Cas Systems: Insights from CRISPR
720 Loci in Integrative Conjugative Elements of *Acidithiobacillia*. *CRISPR J.* **4**, 656-672
721 (2021).

722 41. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz, F. Mobility
723 of plasmids. *Microbiol. Mol. Biol. Rev.* **74**, 434–52 (2010).

724 42. Nishida, H. Comparative Analyses of Base Compositions, DNA Sizes, and Dinucleotide
725 Frequency Profiles in Archaeal and Bacterial Chromosomes and Plasmids. *Int. J. Evol.*
726 *Biol.* **2012**, 342482 (2012).

727 43. Cury, J., Touchon, M. & Rocha, E. P. C. Integrative and conjugative elements and their
728 hosts: composition, distribution and organization. *Nucleic Acids Res.* **45**, 8943–8956
729 (2017).

730 44. Almpanis, A., Swain, M., Gatherer, D. & McEwan, N. Correlation between bacterial G+C
731 content, genome size and the G+C content of associated plasmids and bacteriophages.
732 *Microb. genomics* **4**, e000168 (2018).

733 45. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial
734 species. *Nat Rev Microbiol*. **6**, 431–440 (2008).

735 46. Shapiro, B. J. *et al.* Population genomics of early events in the ecological differentiation
736 of bacteria. *Science* **335**, 48–51 (2012).

737 47. Galperin, M. Y. *et al.* COG database update: focus on microbial diversity, model
738 organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2021).

739 48. Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Microbial
740 genome analysis: the COG approach. *Brief. Bioinform.* **20**, 1063–1070 (2019).

741 49. John, J., George, S., Nori, S. R. C. & Nelson-Sathi, S. Evolutionary route of resistant genes
742 in *Staphylococcus aureus*. *Genome Biol. Evol.* **11**, 2917-2926 (2019).

743 50. Enault, F. *et al.* Phages rarely encode antibiotic resistance genes: a cautionary tale for
744 virome analyses. *ISME J.* **11**, 237–247 (2017).

745 51. MM, D., PM, O. & PM, S. Exotoxins of *Staphylococcus aureus*. *Clin. Microbiol. Rev.* **13**,
746 16–34 (2000).

747 52. Lam, M. M. C. *et al.* Genetic diversity, mobilisation and spread of the yersiniabactin-
748 encoding mobile element ICEKp in *Klebsiella pneumoniae* populations. *Microb.*
749 *Genomics* **4**, e000196 (2018).

750 53. Pinilla-Redondo, R. *et al.* CRISPR-Cas systems are widespread accessory elements across
751 bacterial and archaeal plasmids. *Nucleic Acids Res.* **1**, gkab859 (2021).

752 54. van Belkum, A. *et al.* Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-
753 Resistant *Pseudomonas aeruginosa*. *mBio* **6**, e01796-15 (2015).

754 55. Wheatley, R. M. & MacLean, R. C. CRISPR-Cas systems restrict horizontal gene transfer
755 in *Pseudomonas aeruginosa*. *ISME J.* **15**, 1420-1433 (2020).

756 56. Pursey, E., Dimitriu, T., Paganelli, F. L., Westra, E. R. & Houte, S. van. CRISPR-Cas is
757 associated with fewer antibiotic resistance genes in bacterial pathogens. *Philos. Trans.*
758 *R. Soc. B* **377**, 20200464 (2022).

759 57. Halary, S., Leigh, J. W., Cheaib, B., Lopez, P. & Bapteste, E. Network analyses structure
760 genetic diversity in independent genetic worlds. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 127–
761 132 (2010).

762 58. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High
763 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
764 *Nat. Commun.* **9**, 5114 (2018).

765 59. Cury, J., Oliveira, P. H., de la Cruz, F. & Rocha, E. P. C. Host Range and Genetic Plasticity
766 Explain the Coexistence of Integrative and Extrachromosomal Mobile Genetic Elements.
767 *Mol. Biol. Evol.* **35**, 2230–2239 (2018).

768 60. Rollie, C. *et al.* Targeting of temperate phages drives loss of type I CRISPR–Cas systems.
769 *Nature* **578**, 149-153 (2020).

770 61. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its
771 own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–
772 491 (2013).

773 62. Al-Shayeb, B. *et al.* Clades of huge phages from across Earth's ecosystems. *Nature* **578**,
774 425-431 (2020).

775 63. Liu, M. *et al.* ICEberg 2.0: an updated database of bacterial integrative and conjugative
776 elements. *Nucleic Acids Res.* **47**, D660-D665 (2018).

777 64. Bazin, A., Gautreau, G., Médigue, C., Vallenet, D. & Calteau, A. panRGP: a pangenome-
778 based method to predict genomic islands and explore their diversity. *Bioinformatics* **36**,
779 i651–i658 (2020).

780 65. Fillol-Salom, A. *et al.* Phage-inducible chromosomal islands are ubiquitous within the
781 bacterial universe. *ISME J.* **12**, 2114-2128 (2018).

782 66. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic
783 species definition. *Proc. Natl. Acad. Sci.* **106**, 19126–19131 (2009).

784 67. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics **35**, 4162-4164
785 (2019).

786 68. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and

787 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.*
788 *Biol. Evol.* **32**, 268–274 (2015).

789 69. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
790 2069 (2014).

791 70. Gautreau, G. *et al.* PPanGGOLiN: Depicting microbial diversity via a partitioned
792 pangenome graph. *PLOS Comput. Biol.* **16**, e1007732 (2020).

793 71. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic
794 features. *Bioinformatics* **26**, 841–842 (2010).

795 72. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

796 73. Garcillán-Barcia, M. P., Redondo-Salvo, S., Vielva, L. & de la Cruz, F. MOBscan:
797 Automated Annotation of MOB Relaxases. in *Methods in Molecular Biology* vol. 2075
798 295–308 (Humana Press Inc., 2020).

799 74. Akhter, S., Aziz, R. K. & Edwards, R. A. PhiSpy: A novel algorithm for finding prophages
800 in bacterial genomes that combines similarity-and composition-based strategies.
801 *Nucleic Acids Res.* **40**, e126–e126 (2012).

802 75. Zhao, X. BinDash, software for fast genome distance estimation on a typical personal
803 laptop. *Bioinformatics* **35**, 671–673 (2019).

804 76. Steinegger, M. & Söding, J. Clustering huge protein sequence sets in linear time. *Nat.*
805 *Commun. 2018 91* **9**, 1–8 (2018).

806 77. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J.
807 eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
808 Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825-5829 (2021).

809 78. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob.*
810 *Chemother.* **67**, 2640–2644 (2012).

811 79. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: Hierarchical and refined dataset
812 for big data analysis - 10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).

813 80. Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A. & Sørensen, S. J.
814 CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-
815 Cas Loci. *Cris. J.* **3**, 462–469 (2020).

816 81. Shmakov, S. A. *et al.* The CRISPR spacer space is dominated by sequences from species-
817 specific mobilomes. *mBio* **8**, (2017).

818 82. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale
819 using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

820

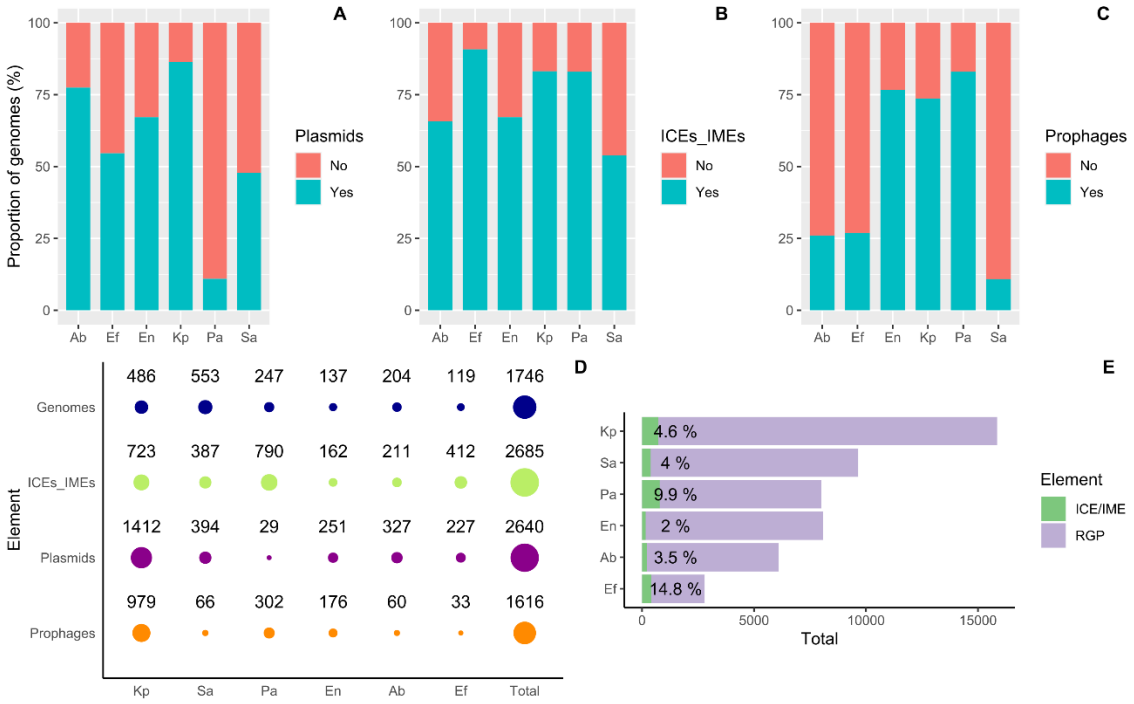
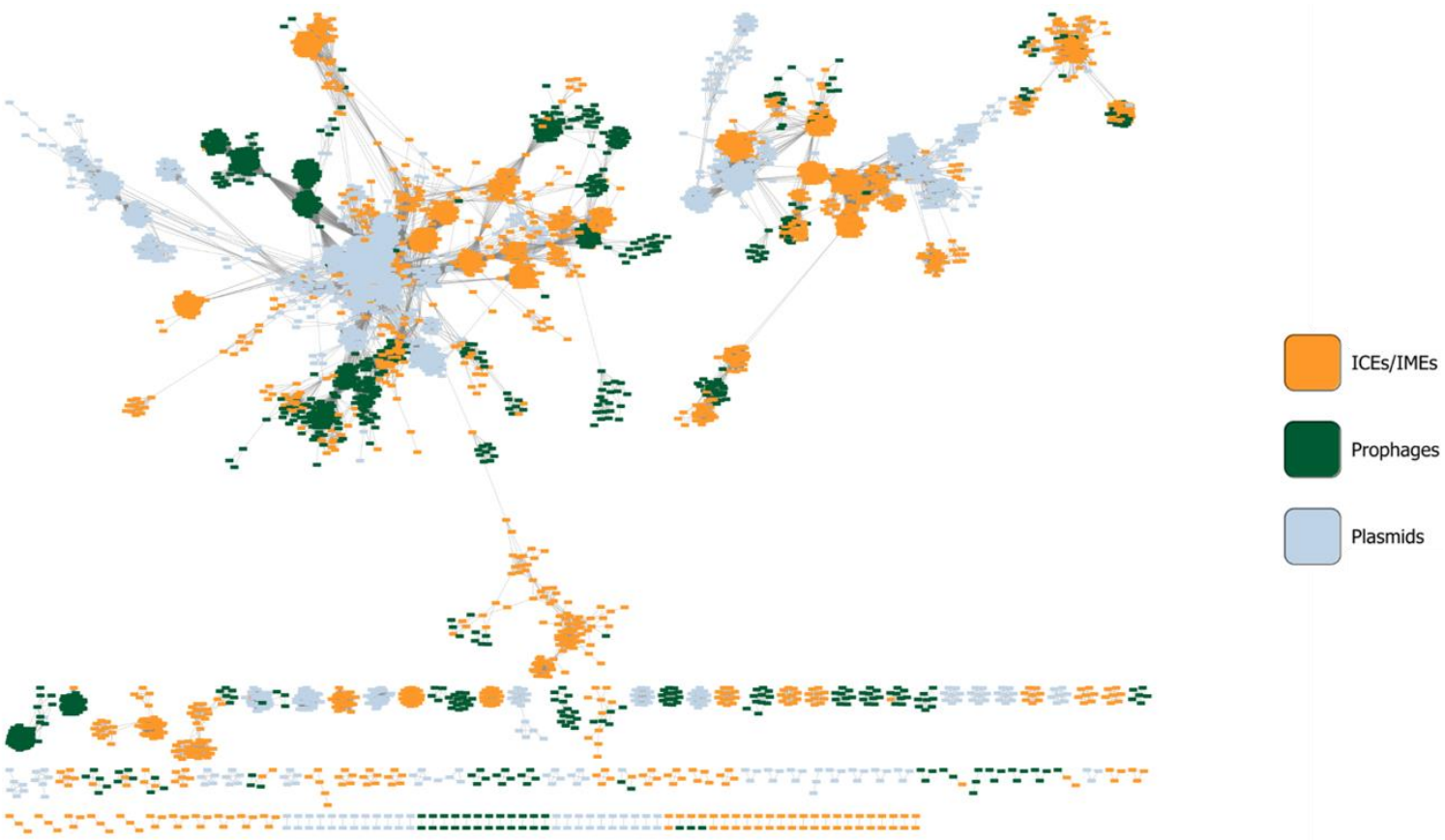


Fig 1. Distribution of MGEs across the ESKAPE pathogens. Proportion of genomes carrying at least one **A**) plasmid, **B)** ICE/IME, or **C)** prophage. **D**) Total number of MGEs and of considered genomes per ESKAPE pathogen. Size of the circles is proportional to the number of identified elements. **E)** Total number of RGPs and ICEs/IMEs per ESKAPE. The size of the green bars is proportional to the total number of ICEs/IMEs identified per ESKAPE pathogen, and the relative number of ICEs/IMEs per RGPs is shown in percentage next to the green bars. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

A)



B)

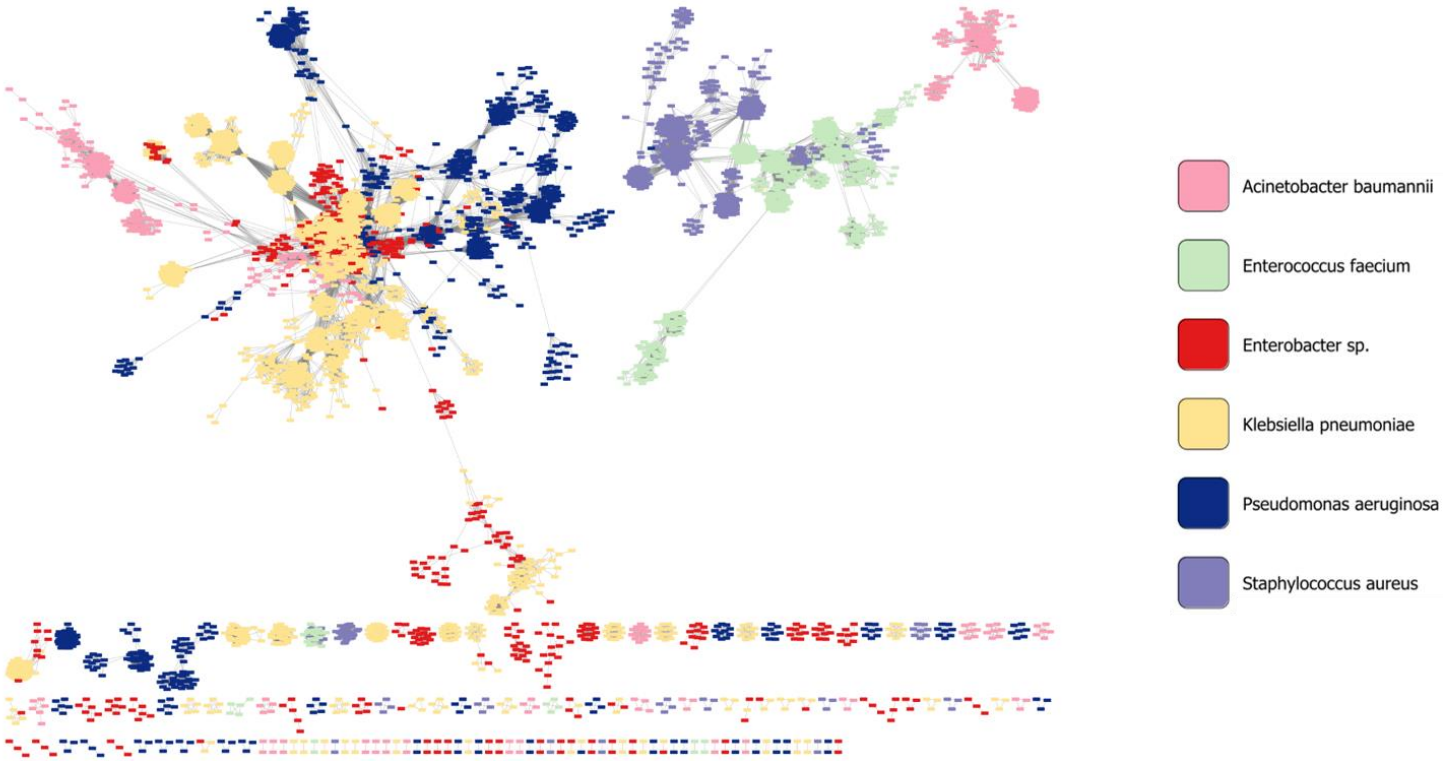


Figure 2. Network of clustered MGEs, using the mean JI as a threshold. Network grouped by **A)** MGE type; **B)** species. Each MGE is represented by a node, connected by edges according to the pairwise distances between all MGE pairs. The network has a clustering coefficient of 0.836, a density of 0.028, a centralization of 0.102, and a heterogeneity of 0.942.

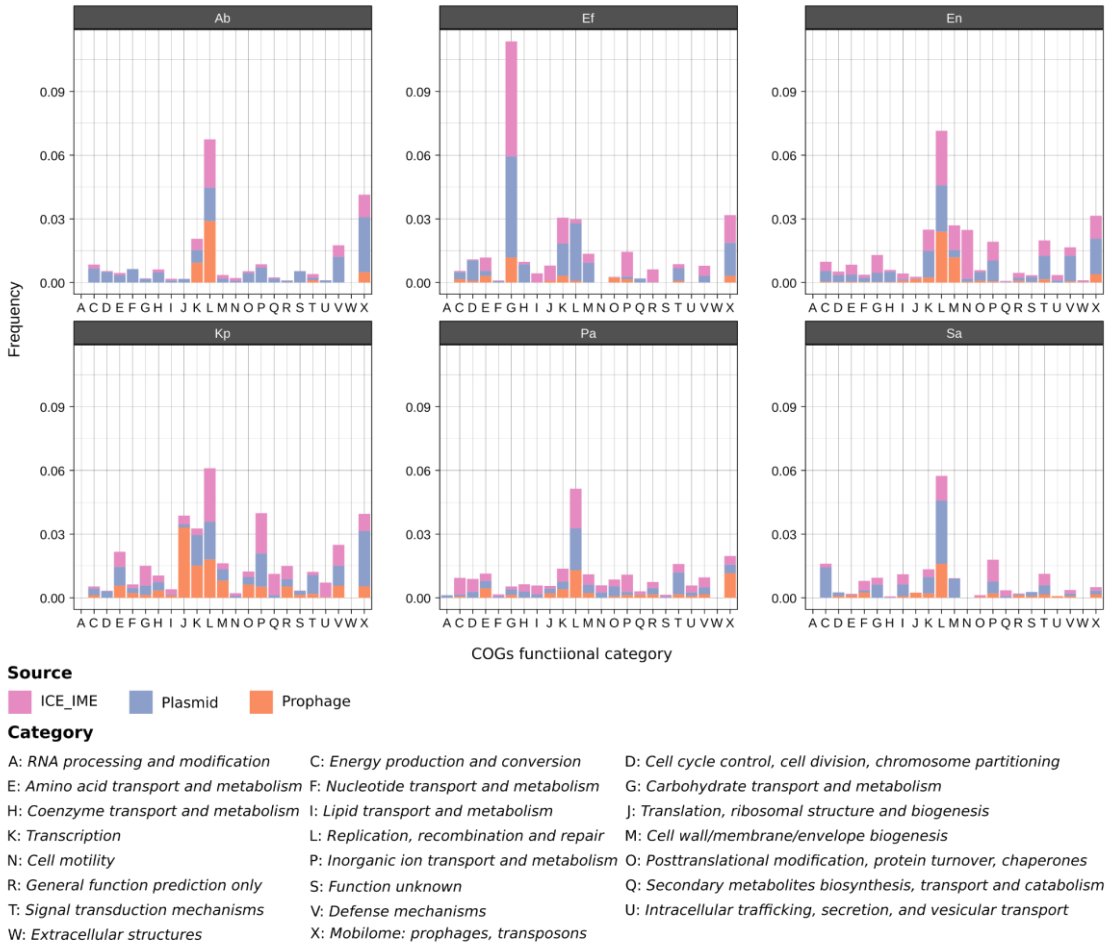


Figure 3. Relative contribution of the ESKAPE MGEs proteome to COG functional categories. Each ESKAPE pathogen is shown in a separate panel. The relative frequency of proteins in the different COG functional categories was calculated separately for each MGE type by dividing the number of proteins belonging to a given category by the total number of proteins observed in the corresponding MGE/ESKAPE pair. Hence, the bars illustrate the incidence of proteins of a given functional category in the proteomes of the different MGE types per ESKAPE taxon. Different COG functional categories are shown along the X axis, including: A - RNA processing and modification; C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and metabolism; I - Lipid transport and metabolism; J - Translation, ribosomal structure and biogenesis; K - Transcription; L - Replication, recombination and repair; M - Cell wall/membrane/envelope biogenesis; N - Cell motility; O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Function unknown; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion,

and vesicular transport; V - Defense mechanisms; W - Extracellular structures; X - Mobilome: prophages, transposons. COG functional categories B, Y, and Z (standing for chromatin structure and dynamics, nuclear structure, and cytoskeleton, respectively) were not identified in the dataset. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

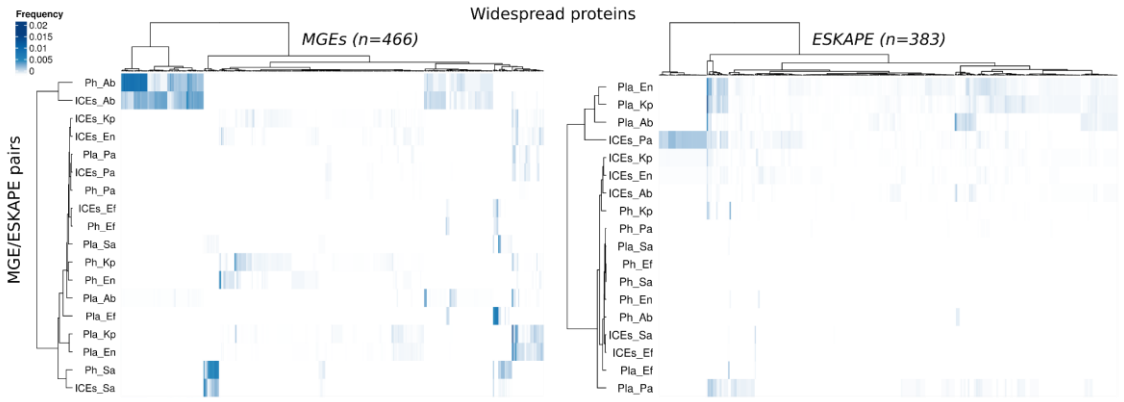


Fig 4. Distribution of widespread protein clusters across MGE/ESKAPE pairs. The heatmaps show the distribution and relative frequency of protein clusters (columns) identified as widespread in either ESKAPE pathogens (i.e. present in at least three different ESKAPE; left side) or MGEs (i.e. present at least once in the three MGE types; right side). MGE/ESKAPE pairs are indicated on the left side of the heatmaps. The relative frequency of the protein clusters in the different MGE/ESKAPE pairs was calculated by dividing the number of occurrences identified by the total number of proteins observed in a given pair. The trees displayed at the top and left side of the heatmaps illustrate the hierarchical clustering of protein clusters and MGE/ESKAPE pairs using the ward.D method with euclidean distance. A list of the protein clusters, their products and relative frequency values is provided in the (**Table S3**). Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

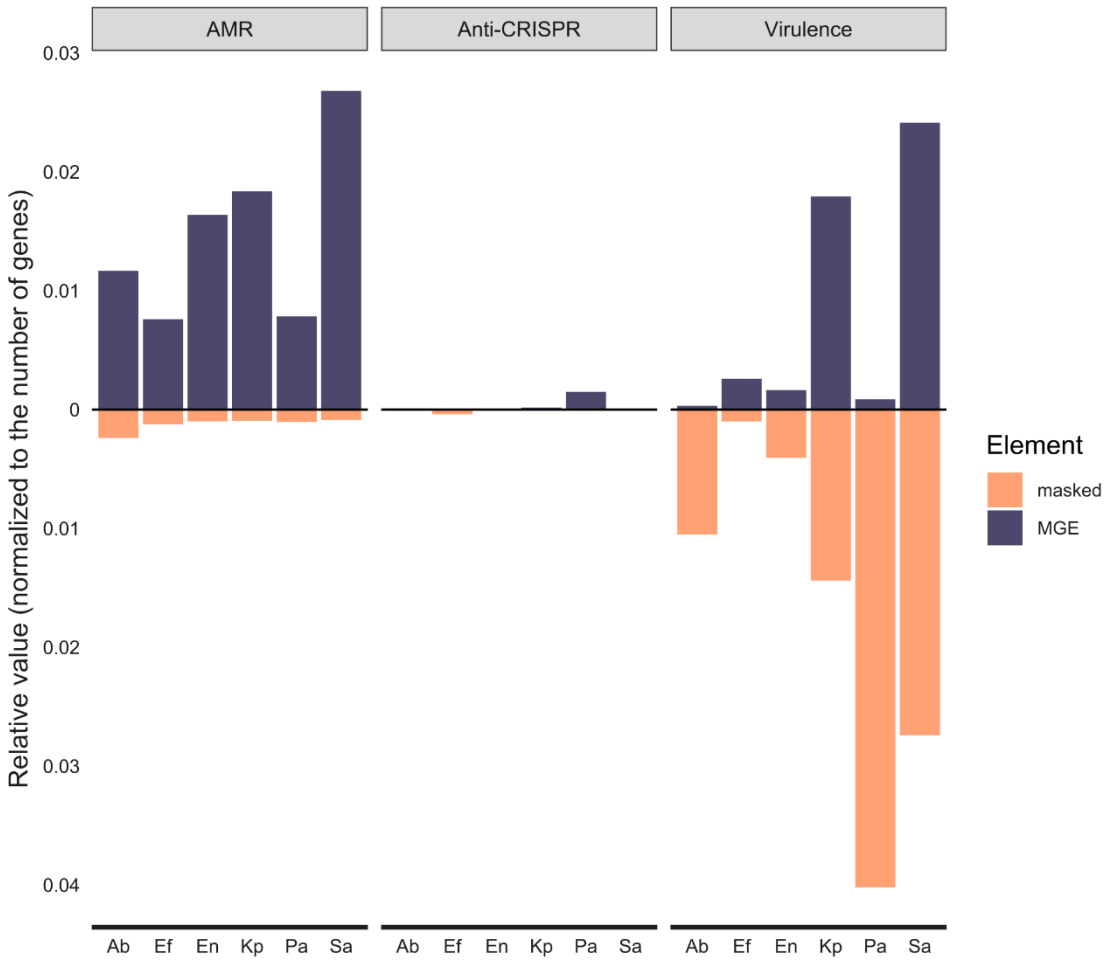


Figure 5. Relative proportion of AMR, virulence, and Anti-CRISPRs between MGEs and masked genomes across the ESKAPE pathogens. The number of AMR, virulence, or Anti-CRISPRs proteins found for MGEs or masked genomes per ESKAPE was normalized to the total number of proteins found for each element per ESKAPE pathogen.

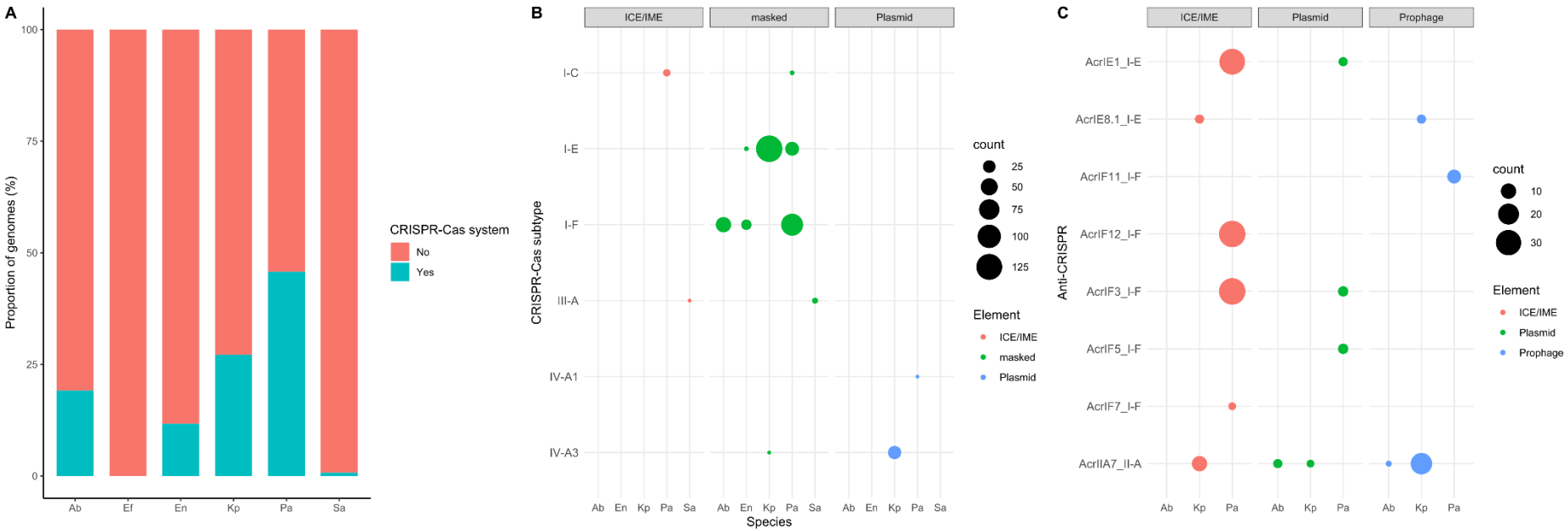


Figure 6. Distribution of CRISPR-Cas and Anti-CRISPR across the ESKAPE mobilome. **A)** Proportion of CRISPR-Cas positive genomes across the ESKAPE pathogens. **B)** Distribution of CRISPR-Cas systems within ICEs/IMEs, plasmids, and masked genomes. Prophages are not shown because we found no CRISPR-Cas system in these elements. **C)** Distribution of Anti-CRISPR proteins across ICEs/IMEs, plasmids, and prophages. Even though we found Anti-CRISPRs on masked genomes, we only plotted those found on MGEs. A complete list of Anti-CRISPRs is given in **Table S6**. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

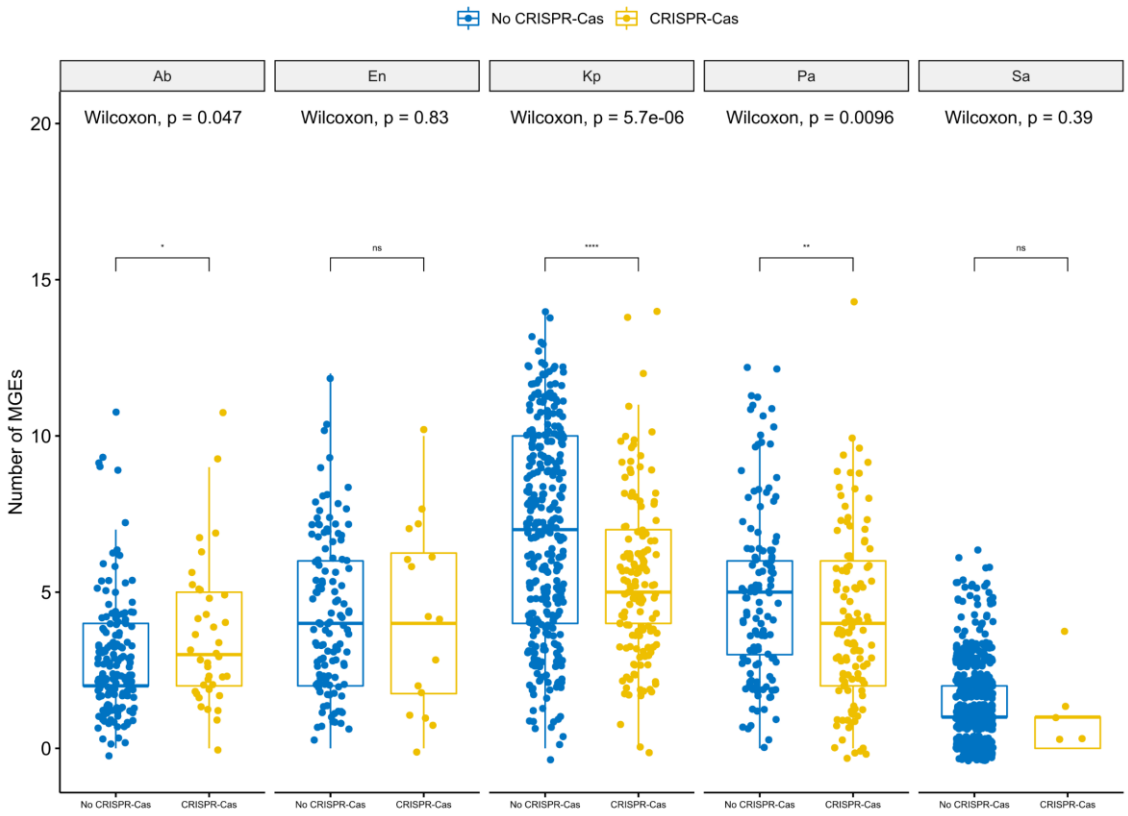


Figure 7. Boxplots comparing the number of MGEs present in pairs of conspecific ESKAPE pathogens, with and without CRISPR-Cas systems. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001. Ab, *A. baumannii*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

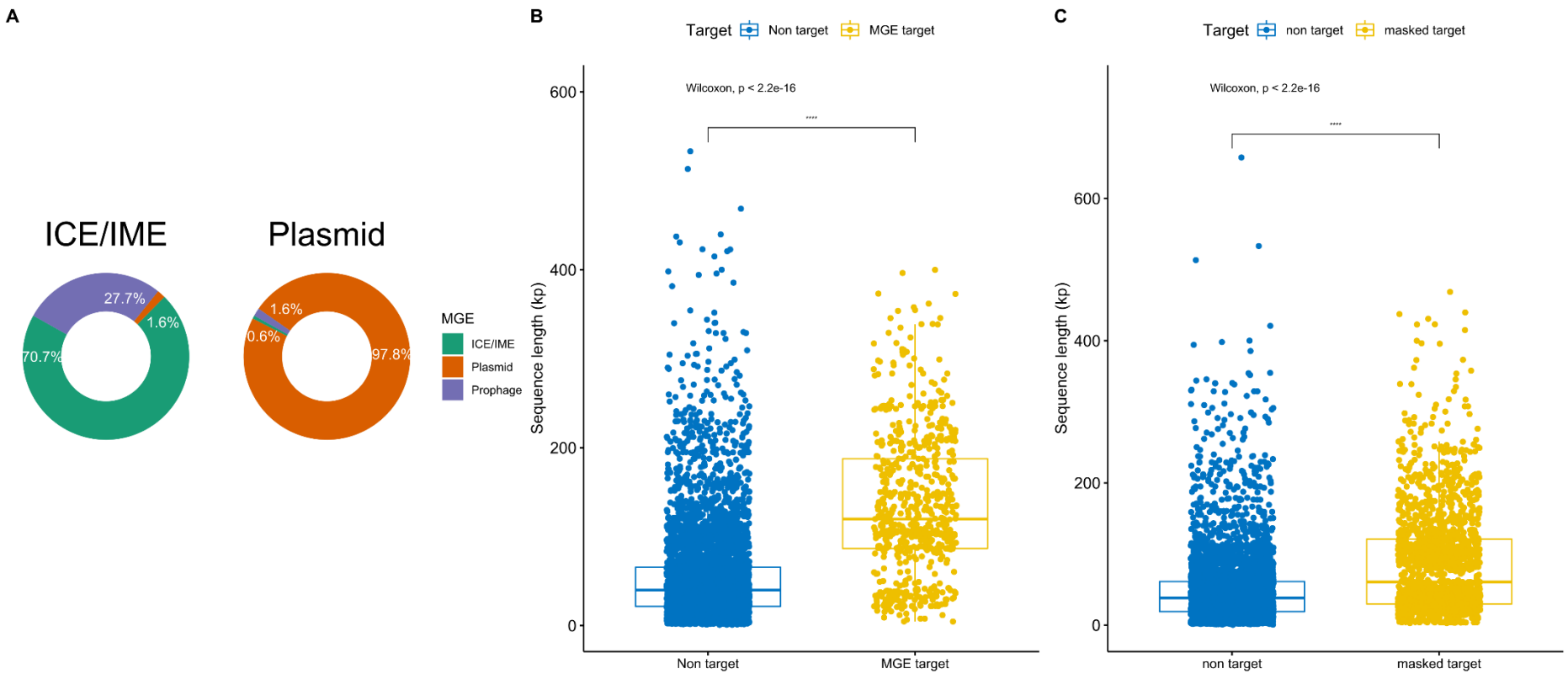


Figure 8. CRISPR spacers are involved in MGE-MGE conflict. **A)** CRISPR spacers found on ICEs/IMEs (left side) and plasmids (right side), and their interactions with protospacers identified in ESKAPE MGEs. **B)** Boxplots comparing the sequence length of MGEs targeted by CRISPR spacers found in MGEs. **C)** Boxplots comparing the sequence length of MGEs targeted by CRISPR spacers found in masked genomes. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001. Ab, *A. baumannii*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

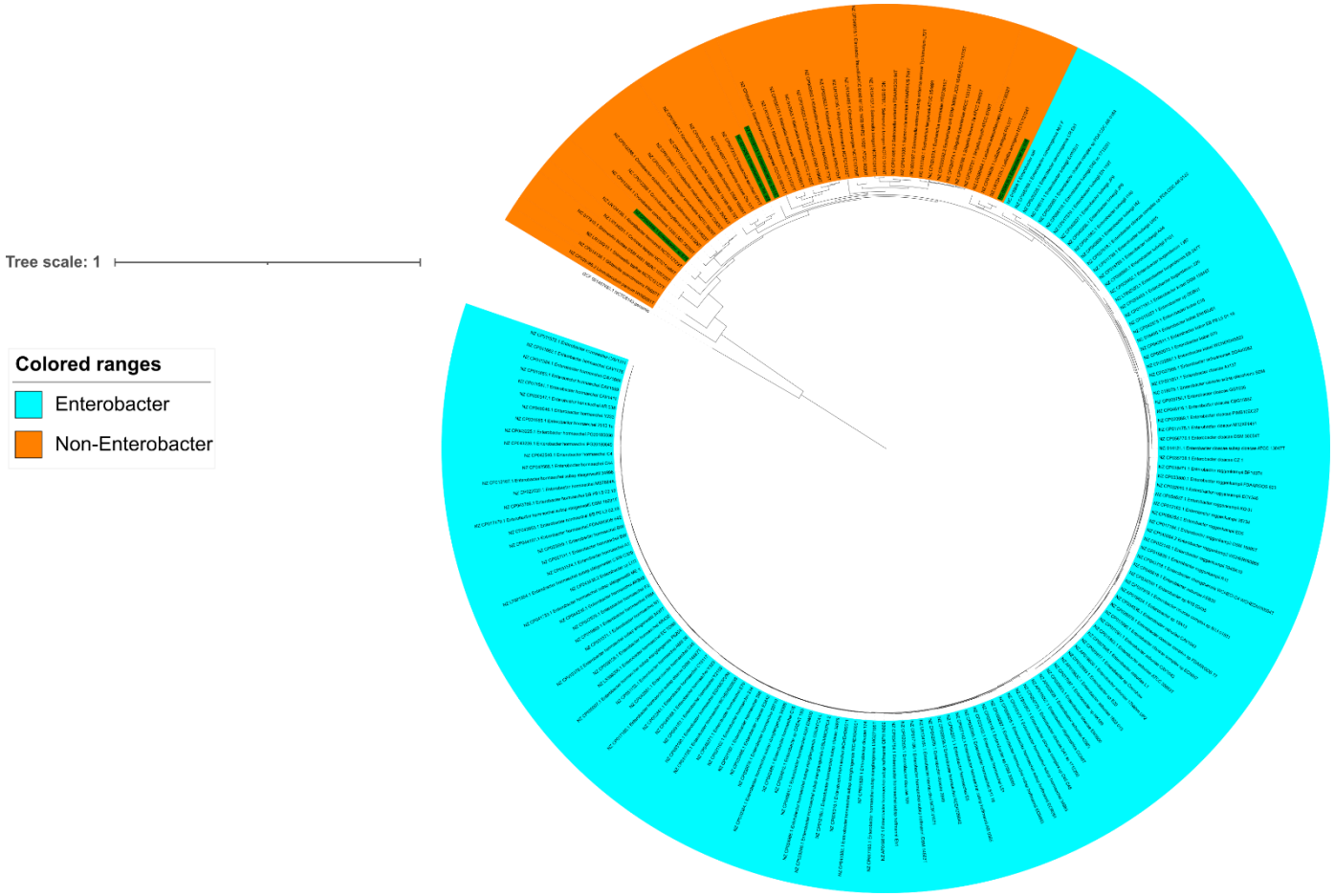


Figure S1. Phylogenomic tree, representing type strains belonging to the *Enterobacteriaceae* family and all complete genomes classified as *Enterobacter*. Genomes that are positioned in the non-*Enterobacter* cluster, and poorly classified as belonging to the *Enterobacter* genus, were highlighted in green.

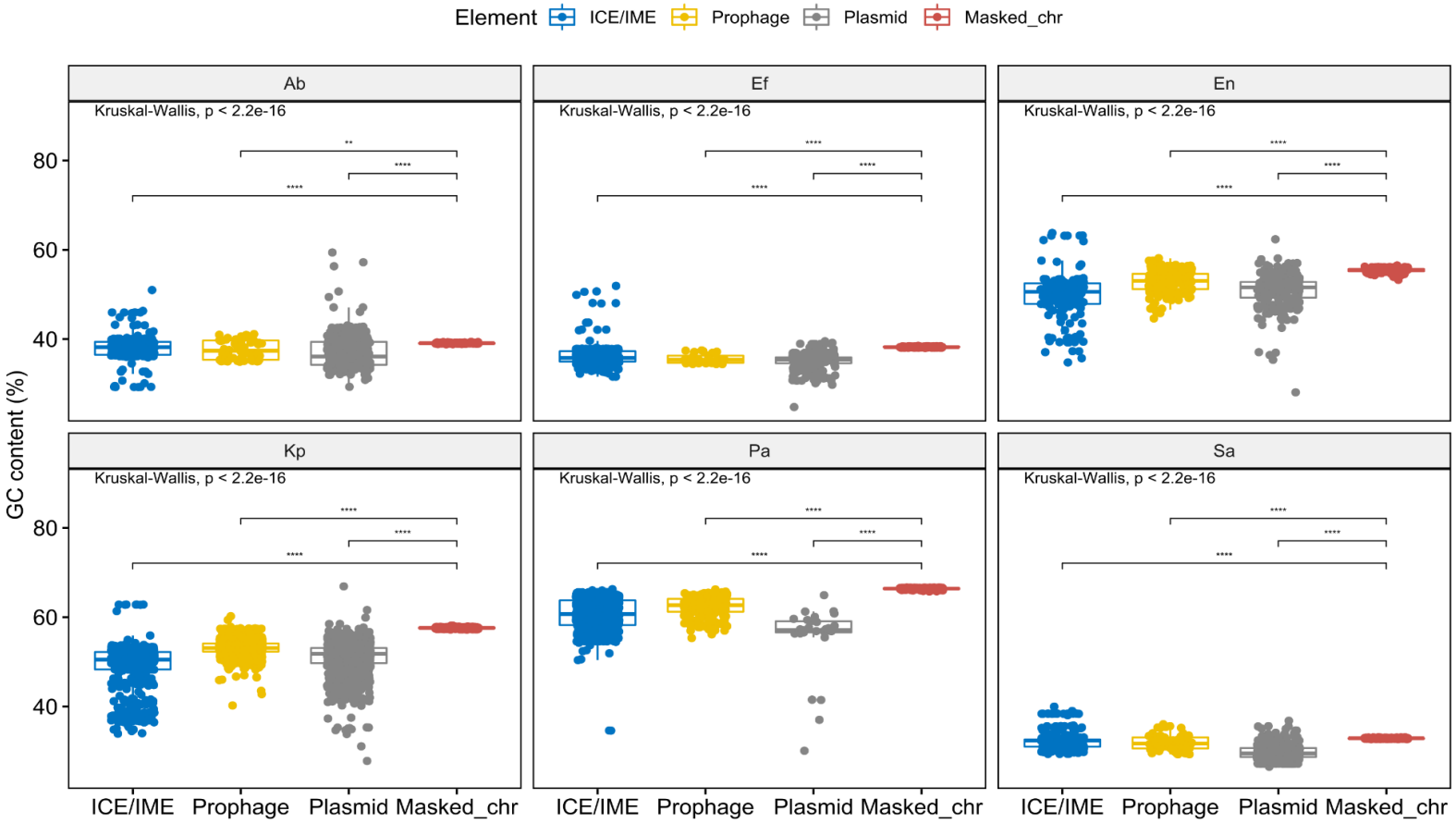


Figure S2. Variation in GC content between MGEs and masked genomes. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001.

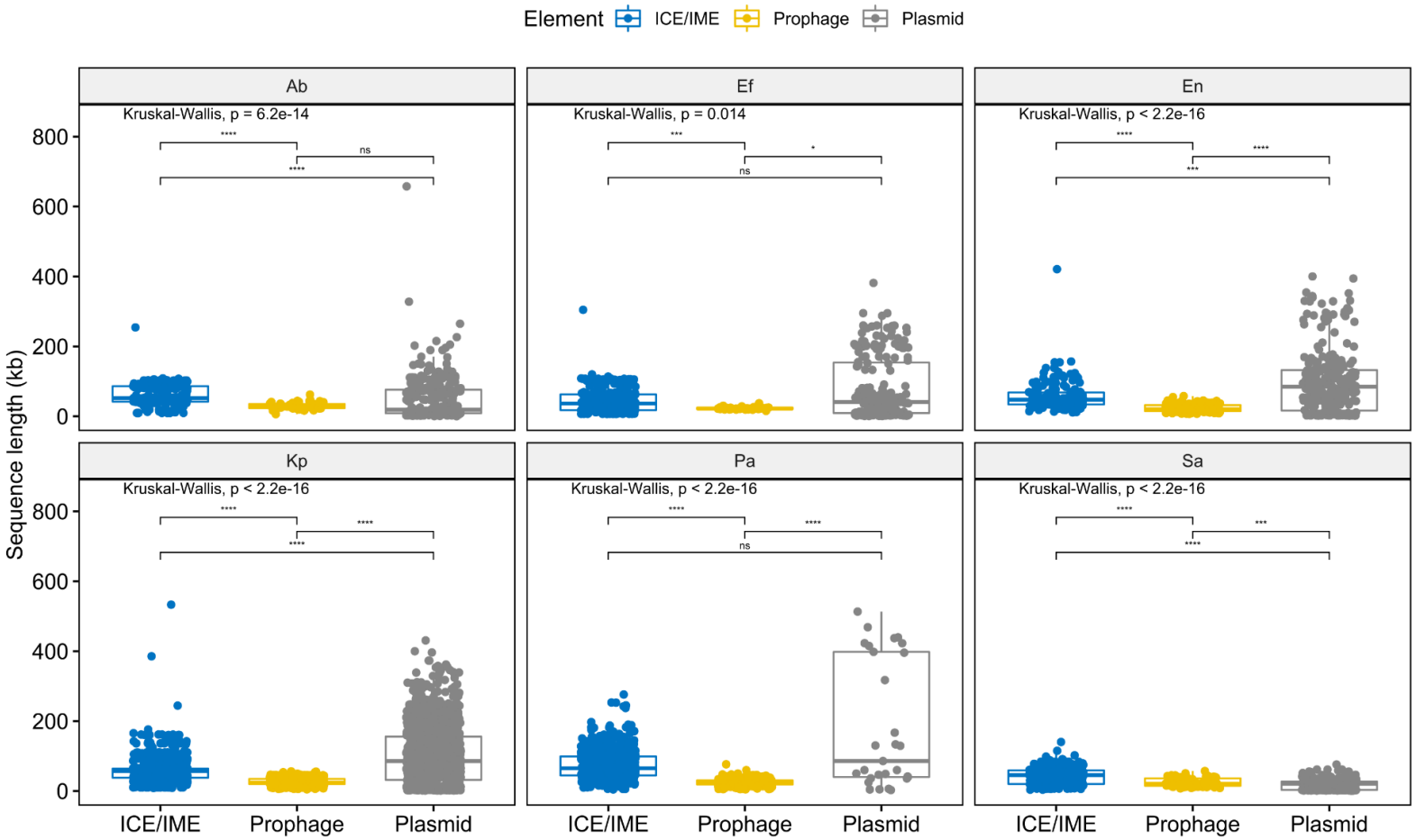


Figure S3. Variation in sequence length between MGEs. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001.

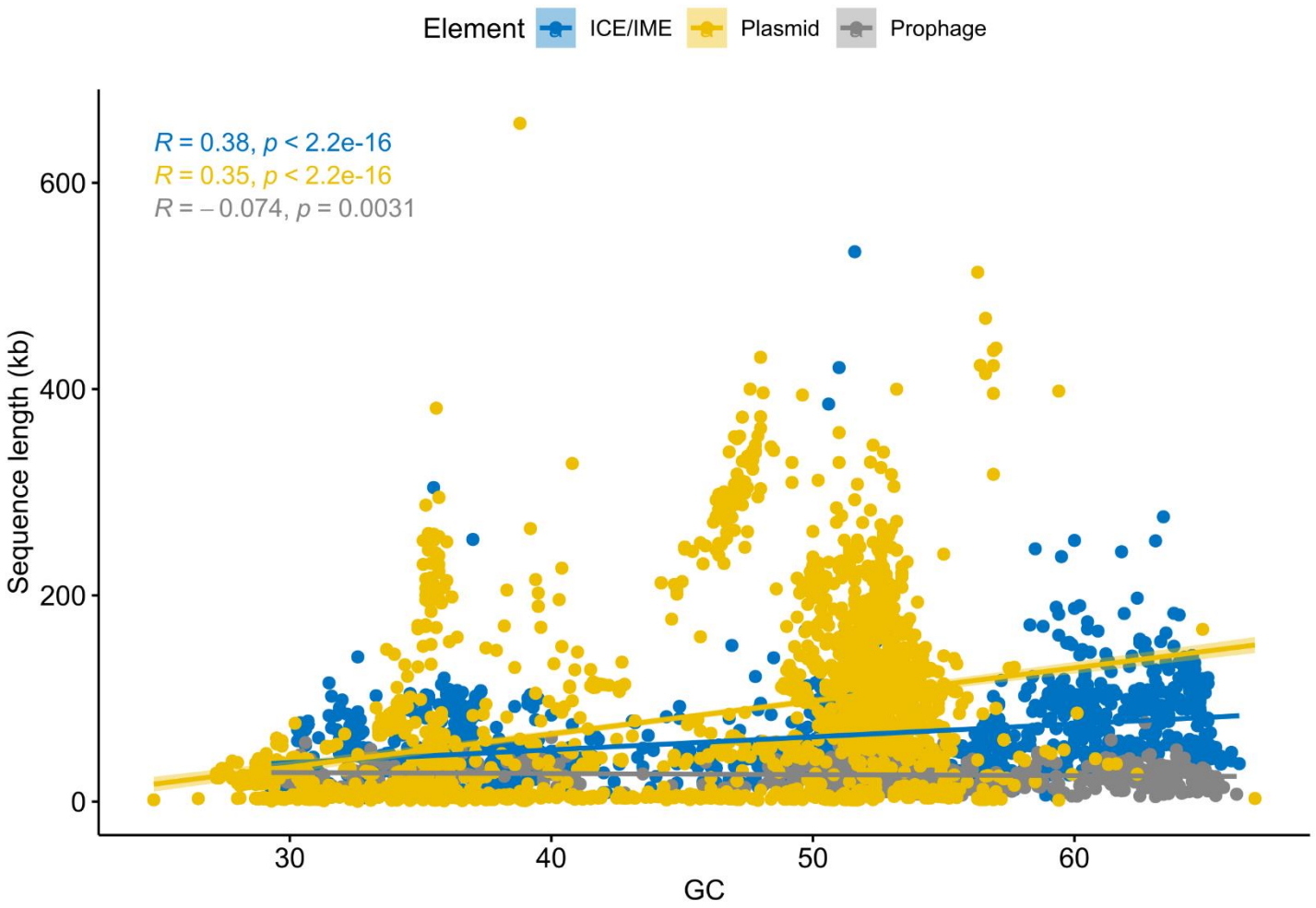
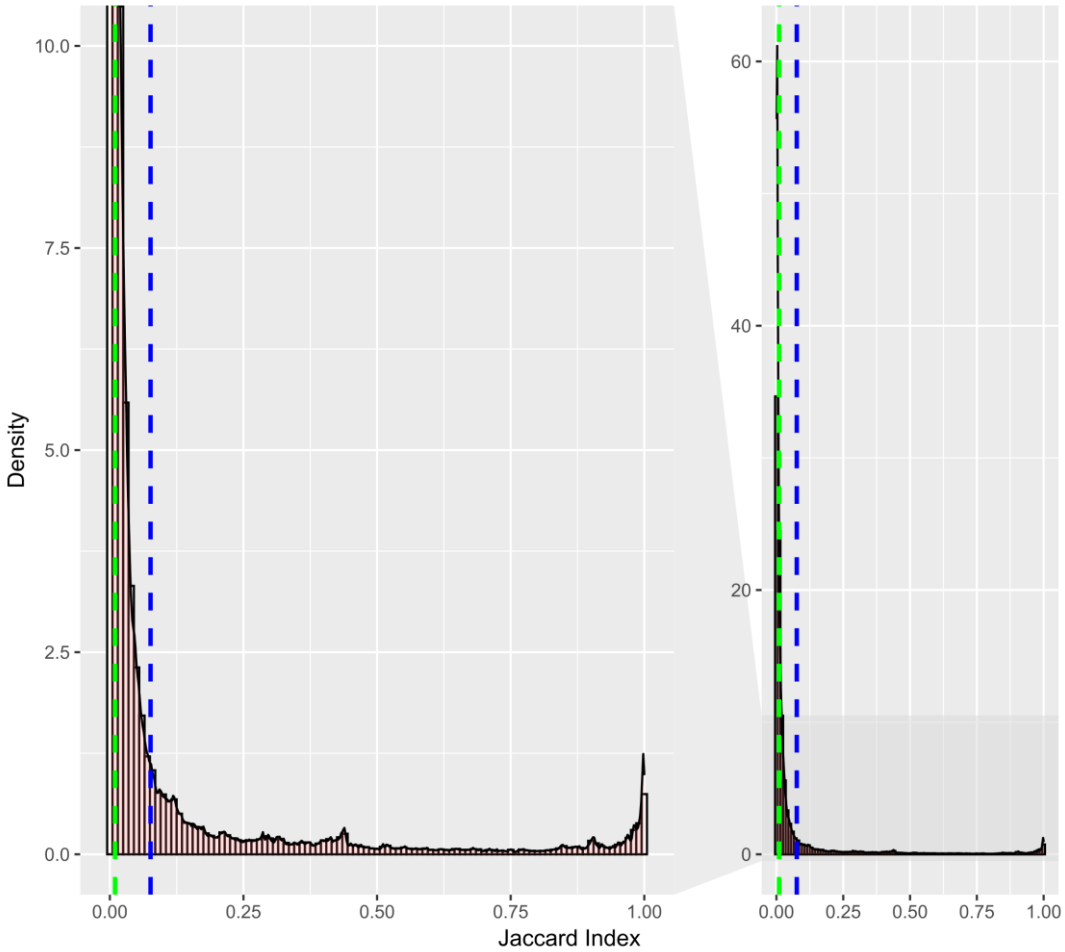


Figure S4. Scatter plot between the different MGEs' sequence length and GC content. Regression lines for each MGE type are shown in different colours. Confidence interval is displayed around smooth. Level of confidence interval is 0.95.

A


Density
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00
Jaccard Index
60
40
20
0
0.00
0.25
0.50
0.75
1.00

B

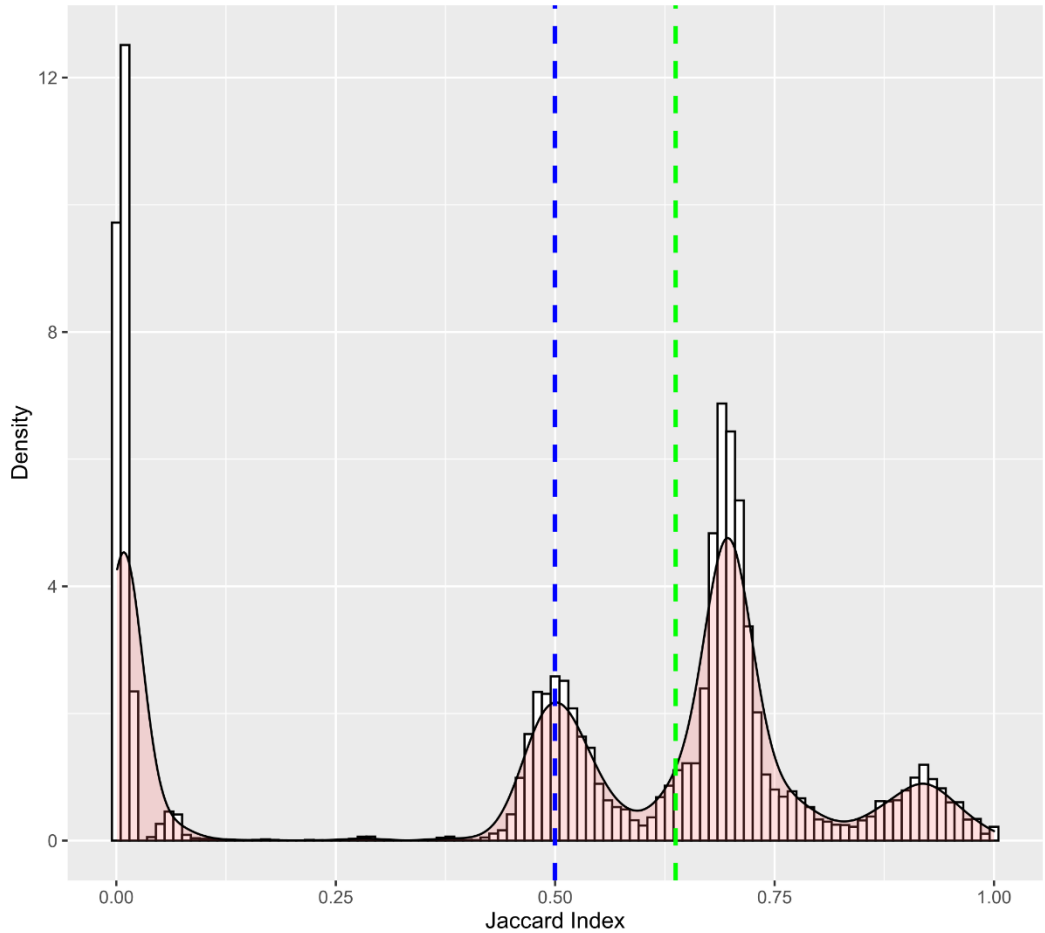


Figure S5. Density plot showing the distribution of Jaccard index (JI) values for the **A)** MGE and **B)** masked genomes network. The mean distribution for the JI is shown in blue, while the median is shown in green. In A), density values from 0-10 were zoomed in and shown in the left.

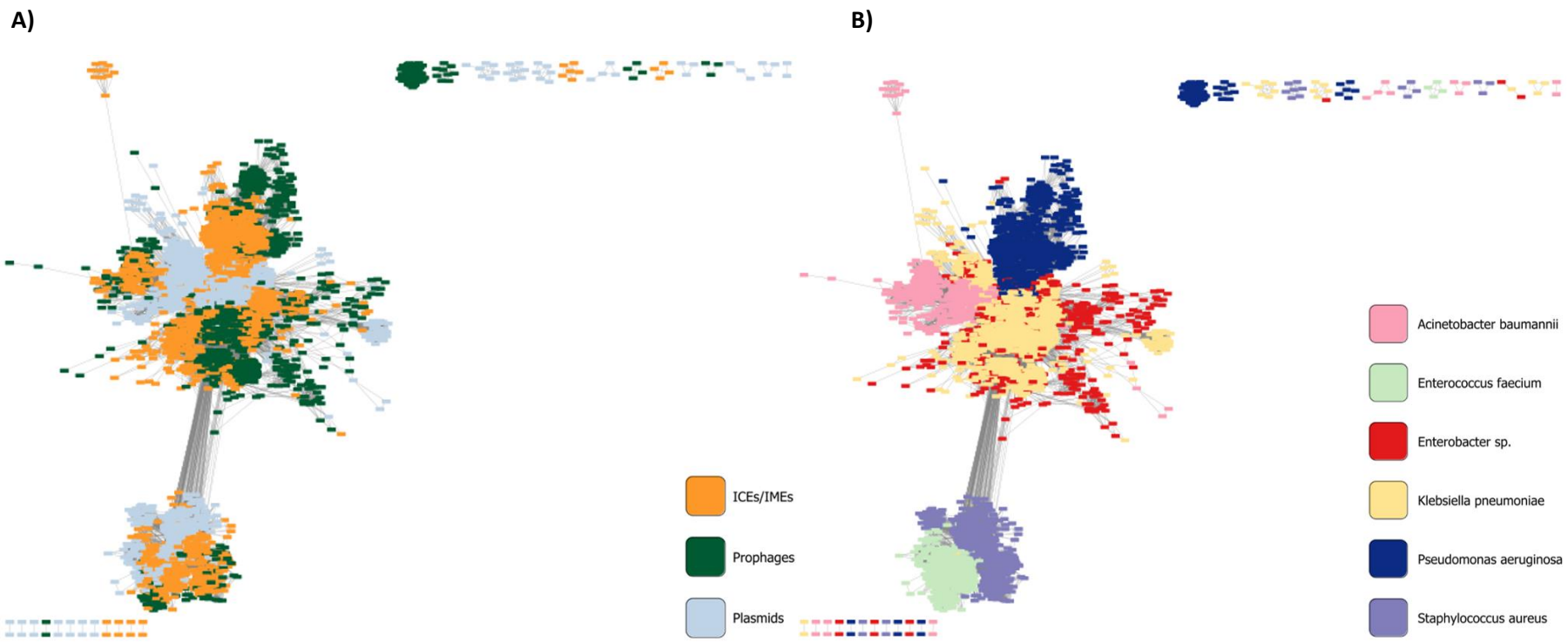


Figure S6. Network of clustered MGEs, using the median JI as a threshold. Network grouped **A)** MGE type; **B)** species. Each MGE is represented by a node, connected by edges according to the pairwise distances between all MGE pairs. The network has a clustering coefficient of 0.783, a density of 0.042, a centralization of 0.168, and a heterogeneity of 1.088.

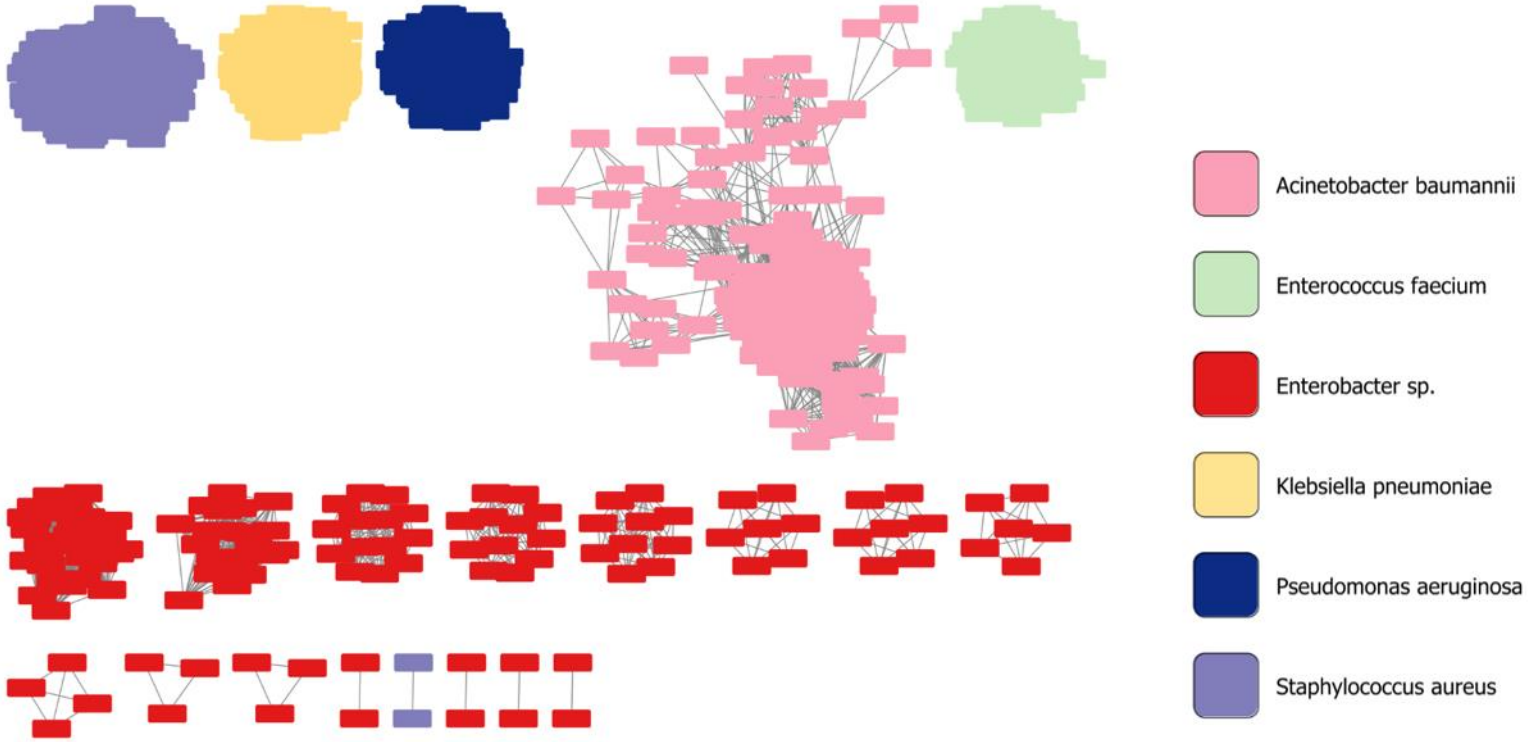


Figure S7. Network of clustered masked genomes grouped by species, using the mean JI as a threshold. Each masked genome is represented by a node, connected by edges according to the pairwise distances between all masked genome pairs.

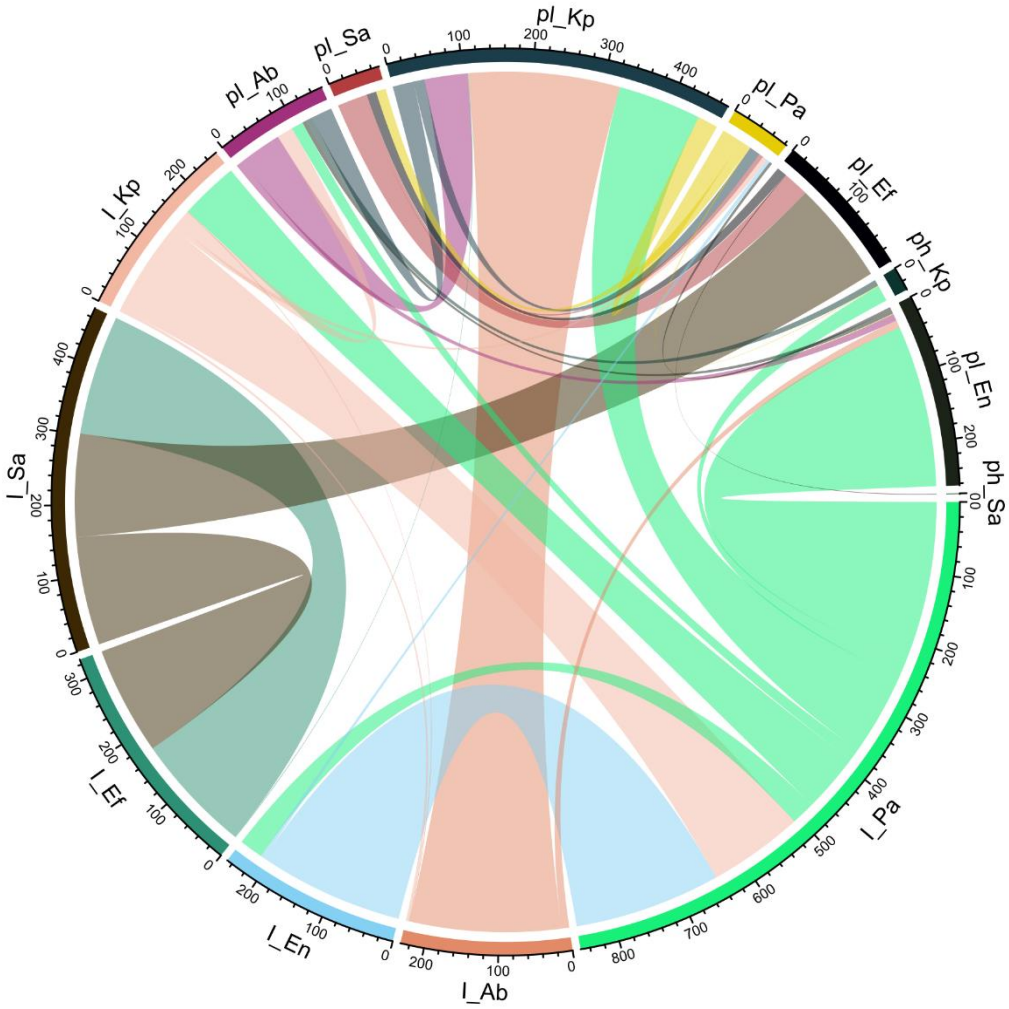


Figure S8. Chord diagram showing DNA sharing events between different ESKAPE pathogens. The width of sectors represents total number of inferred interactions between two different MGE/ESKAPE pairs. The width of links are proportional to the number of inferred DNA sharing events. Interactions between the same ESKAPE/MGE pairs were excluded. Interactions between the same ESKAPE, but different MGEs, were also removed. Interactions between *K. pneumoniae* and *Enterobacter* sp. were also removed, since interactions between these species is very common, as both belong to the *Enterobacteriaceae* family. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

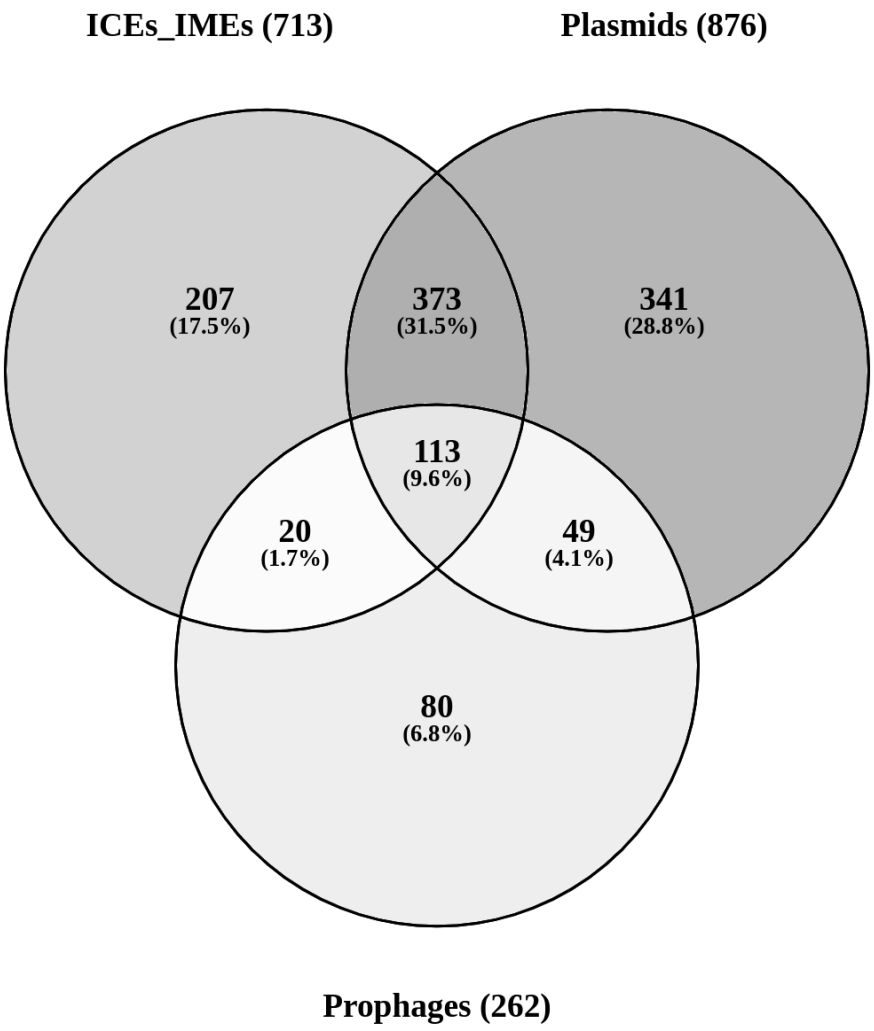


Figure S9. COGs identified in ESKAPE MGEs. Venn diagram showing the distribution of 1183 different COGs identified in ESKAPE ICEs/IMEs, plasmids and phages. The number of different COGs found for each MGE type is indicated in parenthesis. Note that the numbers displayed in the figure do not correspond to the number of proteins with an identifiable COG in MGE sequences but to the distinct COGs detected in the MGEs proteome.

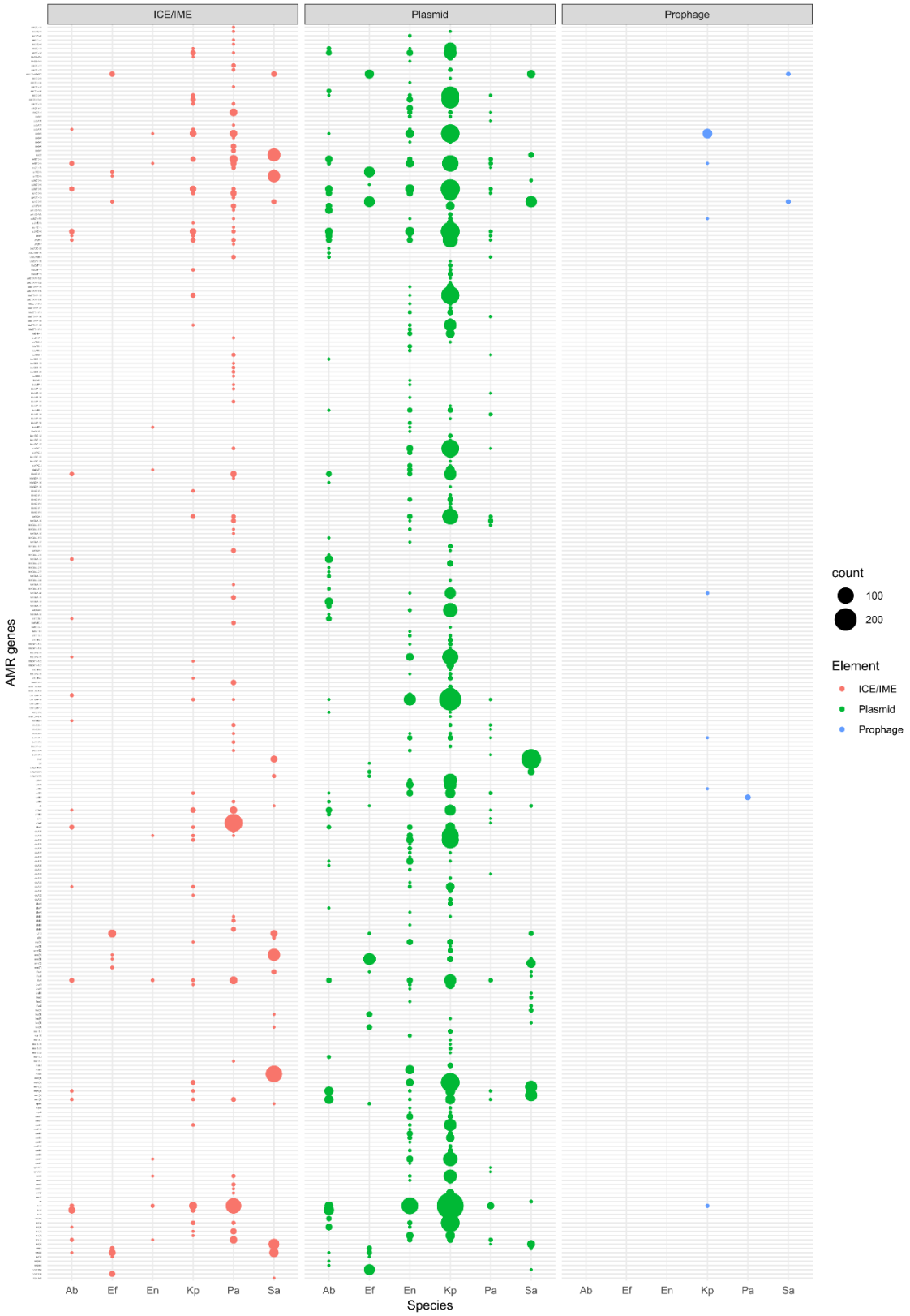


Figure S10. Distribution of antimicrobial resistance (AMR) genes across different ESKAPE pathogens and mobile elements. The size of the circles is proportional to total counts. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

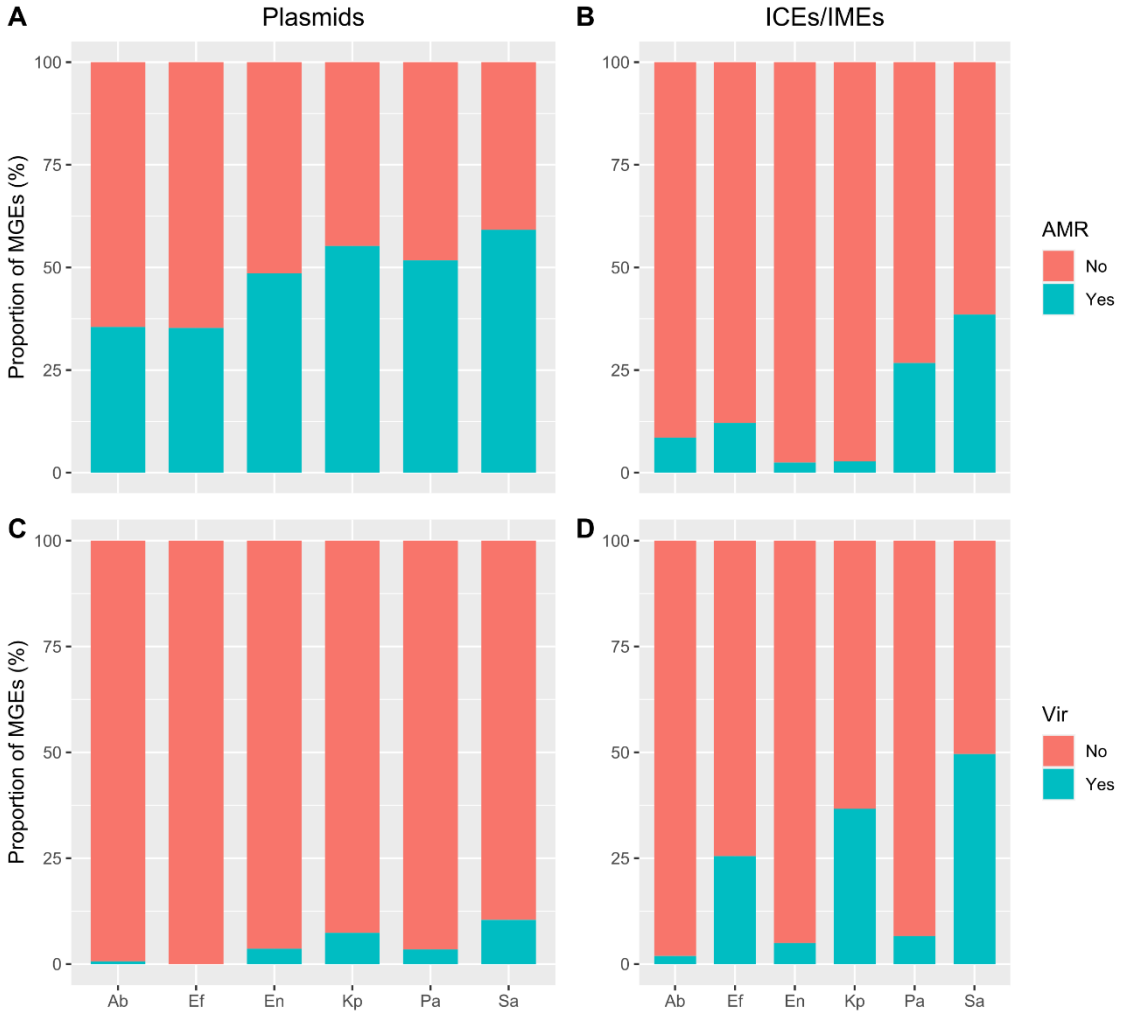


Figure S11. Proportion of mobile genetic elements (MGEs) carrying antimicrobial resistance (AMR) genes, identified in **A)** plasmids and **B)** ICEs/IMEs, and virulence (Vir) genes identified in **C)** plasmids and **D)** ICEs/IMEs. A separate proportion is shown for the different ESKAPE pathogens. Each MGE was considered as positive for AMR or Vir if at least one gene was identified. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

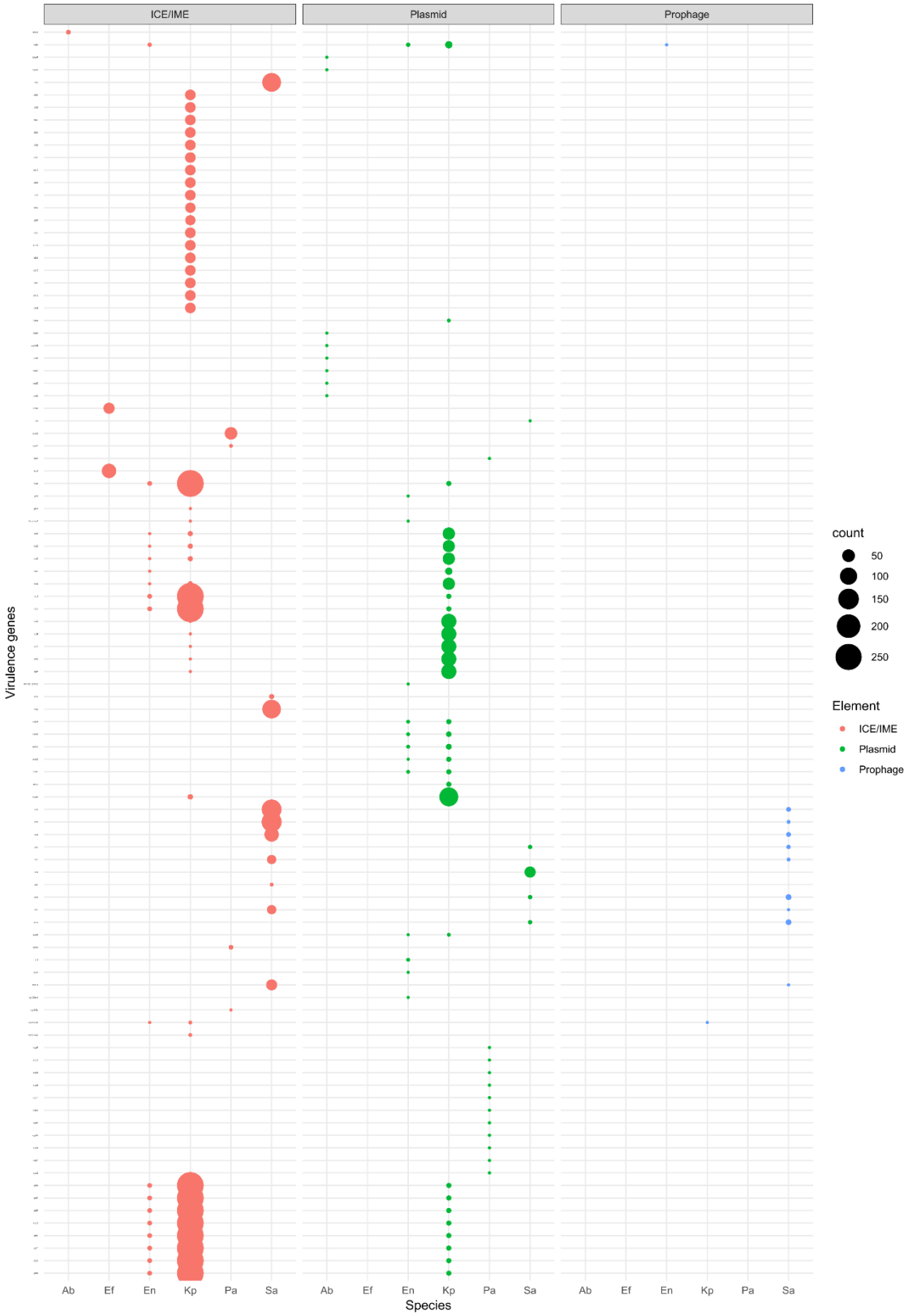


Figure S12. Distribution of virulence genes across different ESKAPE pathogens and mobile elements. The size of the circles is proportional to total counts. Ab, *A. baumannii*; Ef, *E. faecium*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

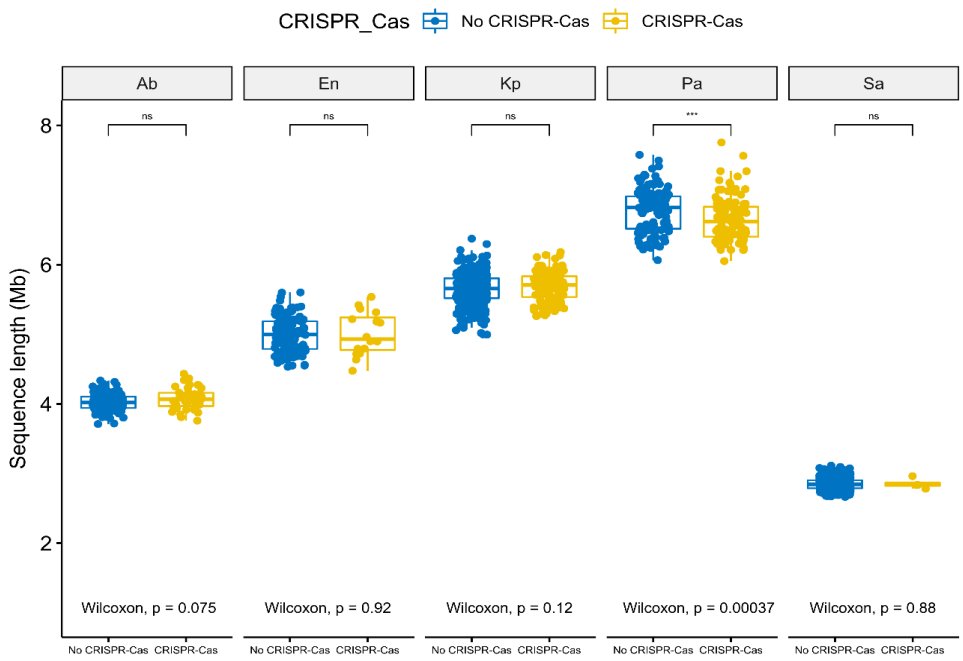


Figure S13. Boxplots comparing the sequence length in pairs of conspecific ESKAPE pathogens, with and without CRISPR-Cas systems. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001. Ab, *A. baumannii*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

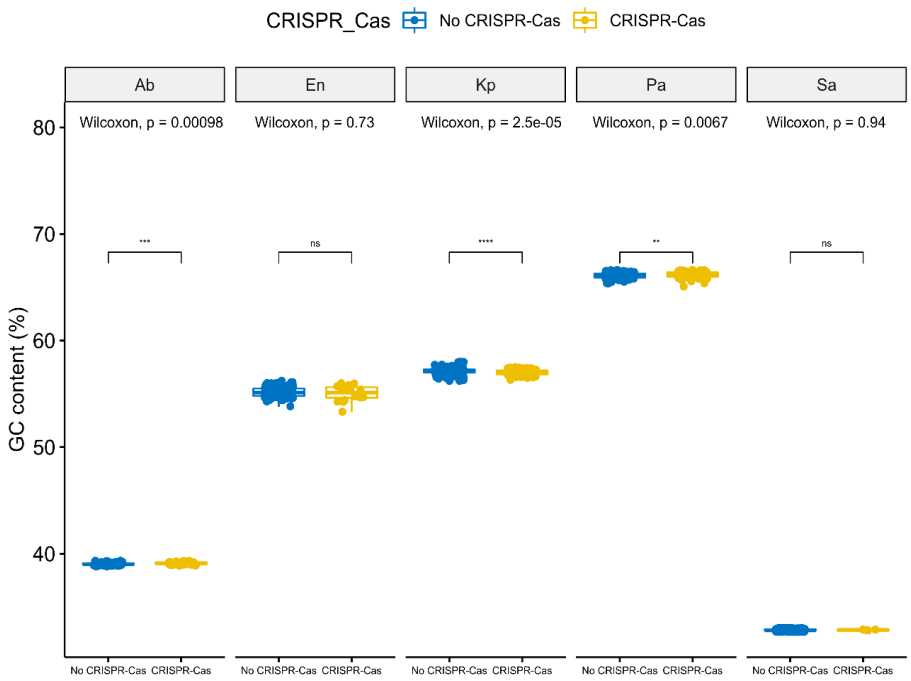


Figure S14. Boxplots comparing the GC content in pairs of conspecific ESKAPE pathogens, with and without CRISPR-Cas systems. Values above 0.05 were considered as non-significant (ns). We

used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001. Ab, *A. baumannii*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

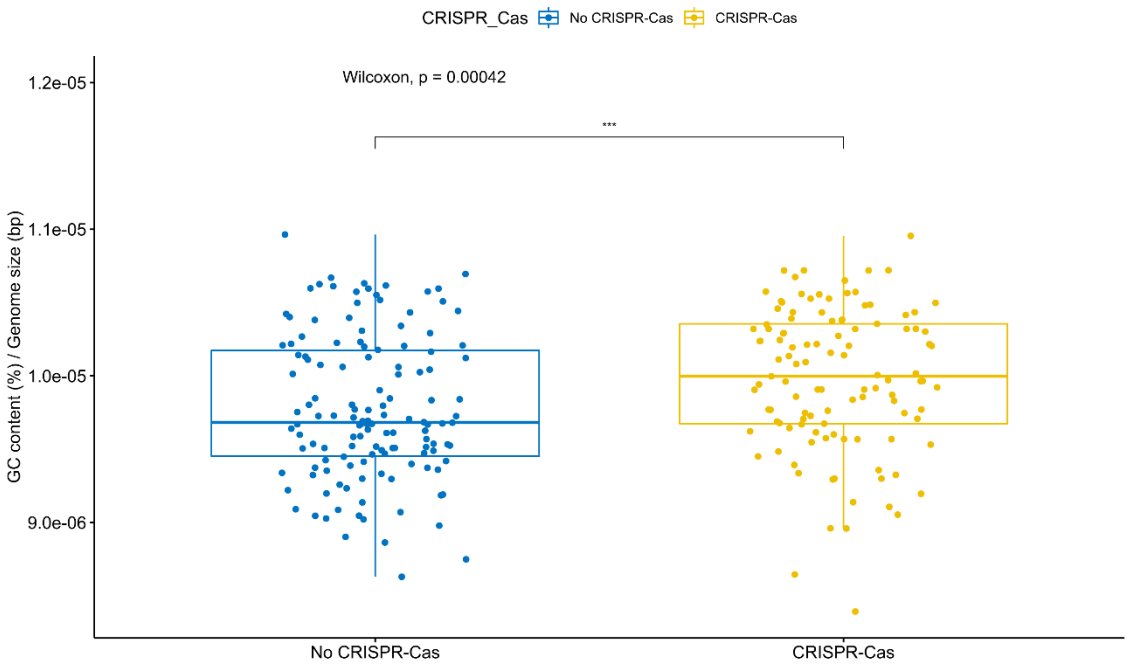


Figure S15. Boxplots comparing CRISPR-Cas positive and negative *P. aeruginosa* genomes, with GC content normalized to sequence length of each genome. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001.

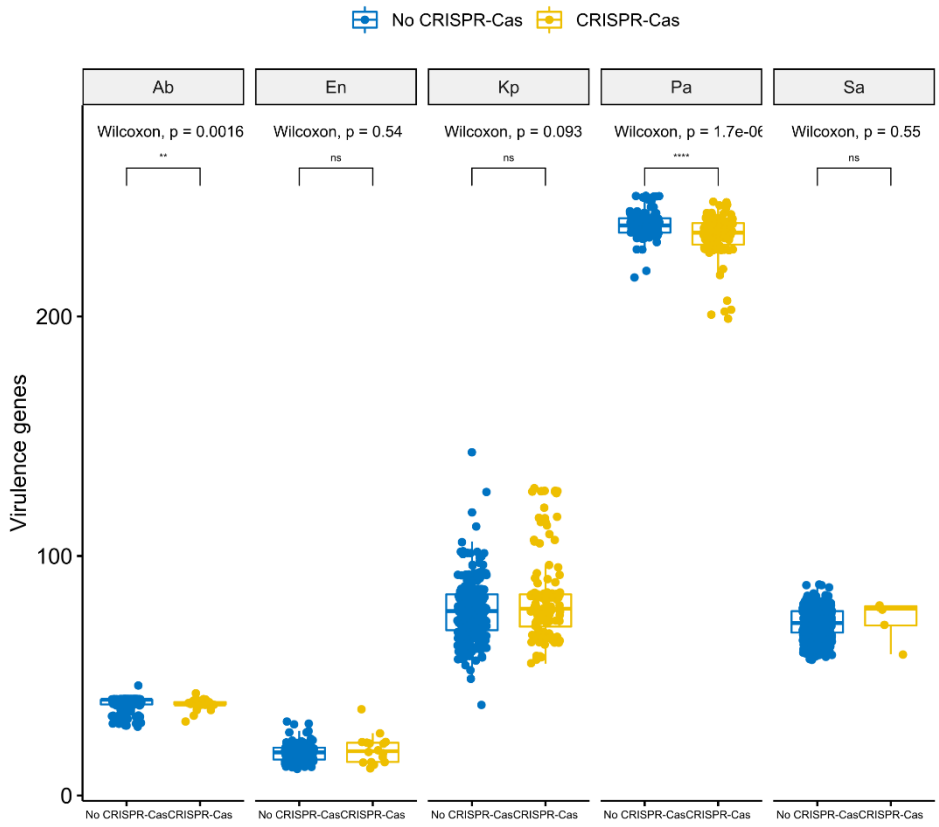


Figure S16. Boxplots comparing the number of virulence genes present in pairs of conspecific ESKAPE pathogens, with and without CRISPR-Cas systems. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001. Ab, *A. baumannii*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

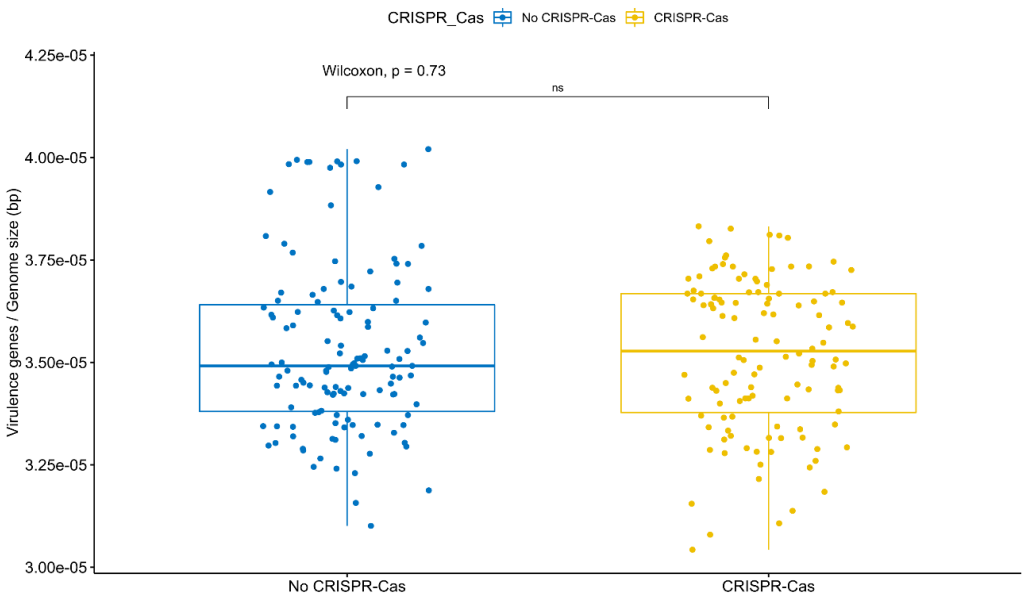


Figure S17. Boxplots comparing CRISPR-Cas positive and negative *P. aeruginosa* genomes, with virulence genes normalized to sequence length of each genome. Values above 0.05 were considered as non-significant (ns).

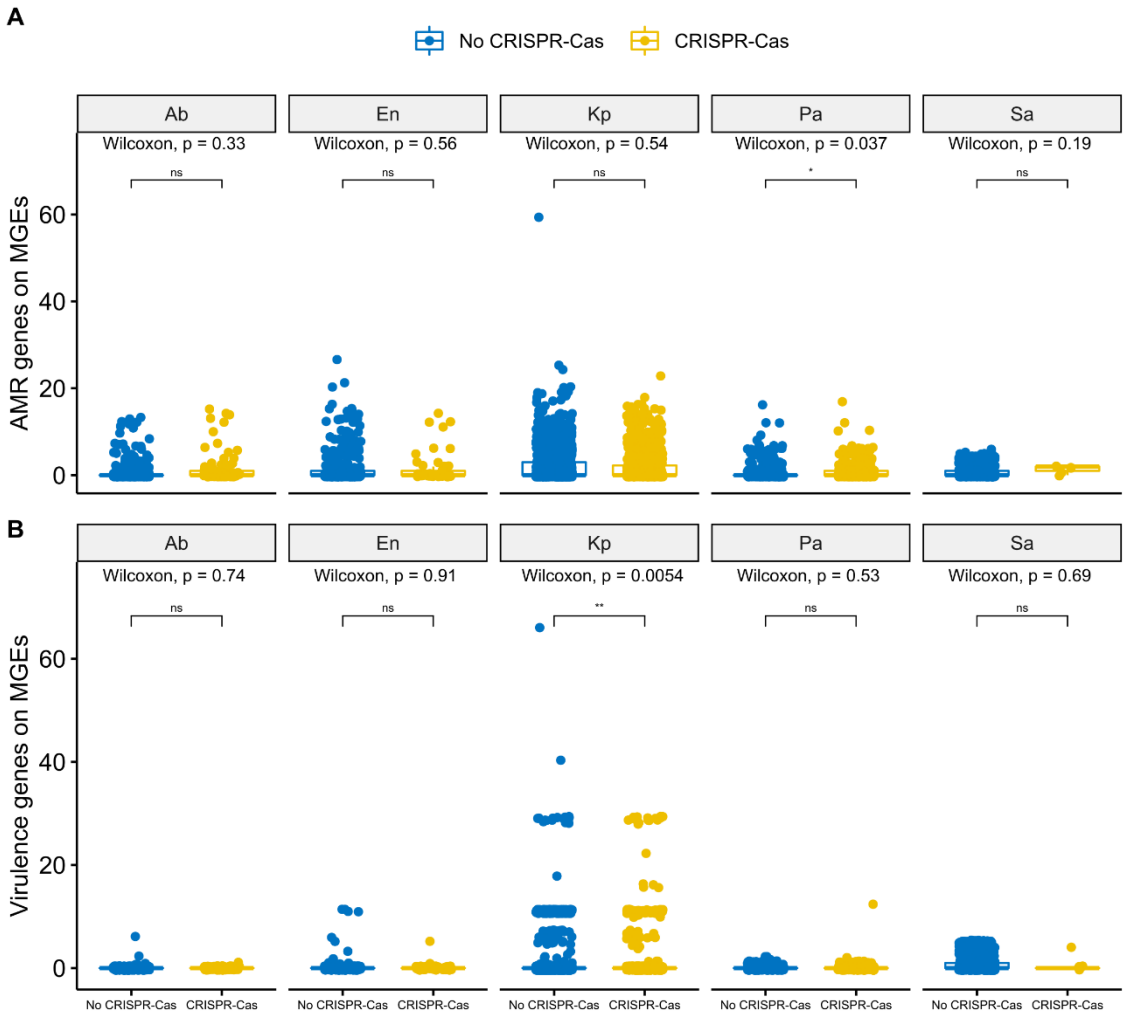


Figure S18. Boxplots comparing pairs of conspecific ESKAPE pathogens, with and without CRISPR-Cas systems, regarding MGE-encoded **A**) antimicrobial resistance (AMR) and **B**) virulence genes. Values above 0.05 were considered as non-significant (ns). We used the following convention for symbols indicating statistical significance: * for p <= 0.05, ** for p <= 0.01, *** for p <= 0.001, and **** for p <= 0.0001. Ab, *A. baumannii*; En, *Enterobacter* sp.; Kp, *K. pneumoniae*; Pa, *P. aeruginosa*; Sa, *S. aureus*.

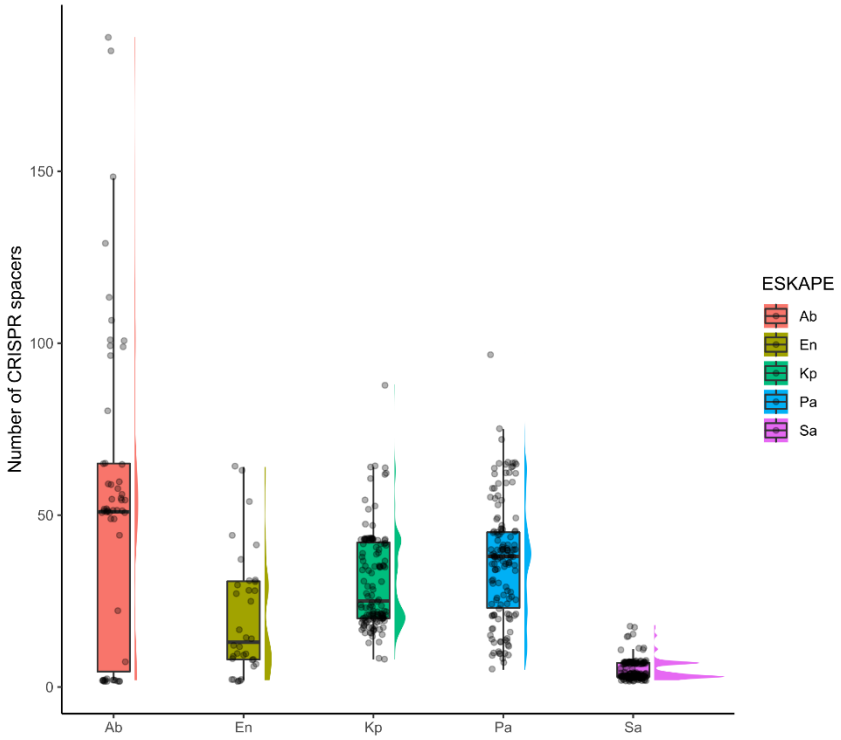


Figure S19. Total number of CRISPR spacers per masked genome identified across the ESKAPE pathogens.