Title

Defence systems drive accessory genome interactions in *Pseudomonas aeruginosa*.

Authors

Charlotte E. Chong[1,12], Aaron Weimann[2,3,4], Aleksei Agapov[5,12], Joanne L. Fothergill[6,12], Michael A. Brockhurst[7,12], Julian Parkhill[4], R. Andres Floto[2,8,9,10], Mark D. Szczelkun[11,12], Edze R. Westra[5,12], Multi-Defence Consortium[12], Kate S. Baker*[1,12]

1. Department of Genetics, University of Cambridge, Cambridge, UK
2. VPD Heart Lung Research Institute, University of Cambridge, Cambridge, UK
3. Cambridge Centre for AI in Medicine, University of Cambridge, UK
4. Department of Veterinary Medicine, University of Cambridge; Cambridge, UK
5. Environment and Sustainability Institute, Biosciences, University of Exeter, Penryn, UK
6. Department of Clinical Infection, Microbiology and immunology, University of Liverpool, Liverpool, UK
7. Division of Evolution, Infection and Genomics, School of Biological Sciences, University of Manchester, Manchester, UK
8. Cambridge Centre for Lung Infection, Royal Papworth Hospital, Cambridge, UK
9. Department of Medicine, University of Cambridge, Cambridge, UK
10. Cambridge University Hospitals Trust, Cambridge, UK
11. DNA-Protein Interactions Unit, School of Biochemistry, University of Bristol, Bristol, UK
12. Multi-Defence Consortium, https://sites.exeter.ac.uk/multidefence/, Environment and Sustainability Institute, Biosciences, University of Exeter, Penryn, UK

*Corresponding Author: kb827@cam.ac.uk

Abstract

As well as undergoing mutational selection, bacterial genomes are shaped by a complex evolutionary interplay among diverse accessory genome elements (AGEs). In this study we define AGEs as encompassing both autonomous elements including

phage, plasmids, integrative conjugative elements (ICEs) and carried elements such as defence systems (DSs) and antimicrobial resistance genes (ARGs). Genomic studies can facilitate our better understanding of the relationships among these AGEs within microbial genomes, offering deeper insights into their roles in bacterial adaptation and evolution. Emerging evidence suggests that bacterial DSs can interact both synergistically and antagonistically with each other, and ecological studies describe non-random patterns of co-occurrence and avoidance of DSs in bacterial populations. Here, we analysed the distributions of DSs and other AGEs in a recently curated dataset of 4,288 *Pseudomonas aeruginosa* genomes. Genomic DS content varied by ecological niche, with higher numbers per genome in non-CF isolates (average n=7.9) compared to CF isolates (average n=6.9). We observed multiple phylogenetically independent associations (n=426) and dissociations (n=50) among DSs, and among other AGEs, many of which had a plausible biological basis. We further explored the relative importance of interactions among different AGEs, revealing that DSs and anti-defence systems engage in the most significant interactions with other AGEs and, most notably, with each other. Ultimately, these patterns of DS interactions and their variation across ecological niches reveals the evolutionary conflicts shaping bacterial accessory genomes and provides a significant public resource for downstream studies on DS and other AGE interactions.

Keywords

Bacterial genomics, bacterial immunity, defence systems interactions, co-occurrence, bacterial defence

Introduction

Bacterial evolution is partially driven by the acquisition of beneficial traits like antimicrobial resistance (AMR), which is often facilitated by horizontal gene transfer (HGT) of accessory genome elements (AGEs) such as plasmids, integrative conjugative elements (ICEs) and phages (1, 2). However, the uptake of this genetic material can also affect fitness and in some cases cause cell death, so in response to this evolutionary pressure, bacteria have acquired immunity against these types of AGEs. Bacteria can contain many complex defence systems (DSs) to defend against harmful invasion ensuring a balance between adaptability and genomic integrity (3). DSs such as

60 restriction-modification (RM) and CRISPR-Cas are powerful contributors to the control
61 of HGT but in some case can stabilise AGEs like phage and plasmids (4, 5).

62 Multiple and diverse sets of DSs are often found within bacterial genomes, with
63 an average of five DSs found in over 20,000 sequenced prokaryotic genomes (6, 7).
64 Furthermore, recent studies demonstrate that DSs may be found clustered in defence
65 islands (3, 7-11). These genomic regions are also commonly associated with
66 autonomous AGEs that are thought to be responsible for their mobilisation (12, 13). The
67 proximity to autonomous AGEs and resultant genome plasticity offers a fitness benefit to
68 bacteria living and responding to a phage-rich environment (3). It has been hypothesised
69 that defence islands form within bacterial genomes due to mechanistic synergism
70 between clustered DSs promoting co-localisation and mobilisation between genomes
71 (11, 14). Additionally, it has also been suggested that defence islands are promoted by
72 the same evolutionary forces-such as HGT, selective advantage, and genetic linkage that
73 drive the aggregation of AMR and pathogenicity genes on autonomous AGEs, such as
74 plasmids, ICEs and phage (11, 15, 16). However, little is known about which DS
75 interactions in naturally evolving populations might represent putative mechanistic
76 interactions that shape HGT, and which are merely co-localisations of convenience.

77 To address these deficiencies in understanding, here we comprehensively
78 analyse phylogenetically independent interactions among DSs, and other AGEs,
79 including plasmids, integrative conjugative elements (ICEs), antimicrobial resistance
80 genes (ARGs), phages, and the newly described anti-defence systems in a recently
81 curated global dataset of 4,288 *P. aeruginosa* genomes (17). *P. aeruginosa* is an ideal
82 organism to study DS interaction as previous studies have found it contains varied DSs
83 (6, 9, 18). It is a widely distributed, Gram-negative bacterium and is found in a diverse
84 range of ecological niches. As an opportunistic pathogen, it is a major concern in hospital
85 settings, causing a broad range of infections, particularly in individuals with cystic
86 fibrosis (CF). Its adaptability has contributed to its status as a World Health Organisation
87 high-priority AMR pathogen (19). Furthermore, DSs were recently revealed to play a key
88 role in its adaptation to different transmission models, making it a valuable model for
89 studying bacterial immunity (9, 17, 20). These interaction patterns among DS and other

90 AGEs provide valuable insights into microbial ecology, evolution, and potential
91 mechanistic interactions.

92 **Methods**

93 Genomic dataset

94 *P. aeruginosa* genomes used in this study were previously described (17). We
95 utilised 4,288 genomes from studies of AMR in healthcare settings (21-24), individuals
96 with cystic fibrosis (17, 25, 26), bacteraemia (27) and non-CF bronchiectasis (28), from
97 studies into high-risk *P. aeruginosa* (29-31) and from the International *Pseudomonas*
98 Consortium (32) (Supplementary Table 1). Isolate genomes from 9 deeply sampled CF
99 patients from the TeleCF study were (17) down sampled, and only genomes
100 representative of diversity were included in the analysis. Genomes were assembled from
101 short-read sequence data, using the Velvet (33) or SPAdes (34) assembler using the
102 Assembly Improvement pipeline (35), as described by Weimann, et al. (2024). Bacterial
103 genomes were annotated with PROKKA v1.14.6 (36).
104

105 Detection of accessory genome elements in bacterial genomes

106 DSs and anti-defence systems were detected in *P. aeruginosa* draft assembly
107 FASTA files using DefenseFinder-DB v1.2.2 (accessed April 2023) and Anti-
108 DefenseFinder-DB v1.3.0 (accessed December 2024) respectively (6). These databases
109 contained 152 DSs and 41 anti-defence systems. Integrative conjugative elements (ICEs)
110 and plasmids were detected in bacterial genomes using ICEberg v.2.0 (1032 database
111 entries) (37) and PlasmidFinder v.2.1.6 (116 database entries) (38) databases via
112 ABRicate v.1.0.1(39), using default parameters. Phage genomes present within the *P.*
113 *aeruginosa* isolates were detected using geNomad v. 1.7.0 (40) and further characterised
114 with PHASTEST webserver (~410,000 viral sequences) (41). Additionally, ARGs were
115 identified using NCBI AMRFinderPlus v. v3.11.11, which contains 6,428 ARG entries (42).
116

117 Identifying co-occurrence

118 Coinfinder v.1.10 (43) was used for detecting significant interactions (i.e.
119 associations and dissociations) among AGEs in the *P. aeruginosa* dataset, using the
120 presence and absence data of defence systems, phage, plasmids, ICEs and ARGs, along

121 with a maximum likelihood tree phylogenetic tree of 3012 patient representative isolates
122 generated by Weimann *et al*, as input (Supplementary Table 1, see Figure 1C in Weimann,
123 et al 2024) (17). A smaller dataset was used in the phylogeny (n= 2941) to avoid
124 overrepresentation of clones from deeply sampled patients whilst ensuring all the
125 genetic diversity was kept (17). Gephi (44) was used to further interrogate interactions
126 and visualise the effect size (observed/expected occurrences) of significant interactions
127 (defined as Bonferroni-corrected binomial exact test $p<0.01$). In addition to effect size,
128 significant interactions were also triaged by the frequency (n= 400 for DS only
129 interactions and n= 300 for AGE interactions) of the AGE and their given D-value (D-value
130 >0), a measure of phylogenetic independence that quantifies the phylogenetic signal of
131 a binary trait. A lower D-value indicates a strong phylogenetic conservation (traits
132 following Brownian motion evolution), while a higher D-value suggests more random
133 distribution independent of phylogeny.
134
135 Statistical analyses
136 A Wilcoxon rank sum test was used to compare the frequency of AGEs in genomes
137 between niches. The association of DSs with CF and non-CF genomes was tested using
138 Fisher's exact test with Bonferroni-adjusted p-values. T-tests were used to evaluate the
139 differences in number of DSs, anti-defence systems and other AGEs per genome length
140 (kb) between genomes isolates between niches, and to assess differences in DS
141 frequency between isolates containing CRISPR-Cas Type I-F compared to those without.
142 To compare the length of contiguous sequences (kb) containing Pycsar and Shedu DSs
143 (to those containing any other DS), we performed a Kruksall-Wallis rank sum test
144 followed by a Dunn's post hoc test for pairwise comparisons. All analyses were
145 performed using R v 4.4.0.
146 To assess the relative importance of interactions among different AGE features,
147 we calculated the proportion of significant interactions as a fraction of total theoretically
148 possible pairs (among detected elements in the AGE types). All significantly interacting
149 pairs were calculated based on the total number of each genomic element found (i.e.
150 ARG = 449, anti-defence system = 30, DS = 131, ICE = 19, phage = 30, plasmid = 13 see
151 Supplementary Table 1). To determine the total number of theoretically possible pairs
152 between different AGE databases (i.e. types of accessory genome feature e.g. ARGs,

153 DSs), we multiplied the size of the respective database (i.e. N_{AMR} x N_{DS}, where *N* is the
154 number of elements detected in the dataset). For pairs within the same category (e.g.,
155 AMR-AMR, we calculated the possible pairs as N_{AMR} x (N_{AMR}-1)/2 to adjust for duplicates.
156 The proportion of the observed number of significant interactions for each pair type was
157 then divided by the total theoretically possible pairs for that category to determine the
158 proportion of (possible) pairs that were significantly interacting.
159
160 **Results**
161 The accessory genome features of *P. aeruginosa* differ between niches
162 To explore the variation in the DSs of *P. aeruginosa* we analysed the DS repertoire
163 of 4,288 *P. aeruginosa* genomes which were isolated from several sources (Figure 1A)
164 and included globally distributed human, animal and environmental isolates. We
165 grouped isolates into two categories: CF and non-CF and compared their frequency of
166 AGE carriage (i.e. presence and absence). This categorisation was informed by findings
167 from Weimann, et al (2024), which reported enrichment of genes involved in
168 transcriptional regulation, inorganic ion transport, lipid metabolism and significant
169 depletion of genes involved in bacterial defence in epidemic clones of *P. aeruginosa*.
170 Therefore, we sought to examine whether similar patterns were evident in the distribution
171 across a representative dataset of CF and non-CF isolates. This revealed that *P.*
172 *aeruginosa* were rich in AGEs (i.e. DS (CF: 99.4% (n=2011/2023), non-CF: 99.8%
173 (n=2262/2265)), anti-defence system (CF: 57.5% (n=1164/2023), non-CF: 66.9%
174 (n=1516/2265)), plasmid (CF: 8.2% (n=166/2023), non-CF: 13.2% (n=301/2265)), ARG
175 (CF: 100% (n=2023/2023), non-CF: 100% (n=2265/2265)), ICE (CF: 97.1% (n=1965/2265),
176 non-CF: 98.1% (n=2222/2265)) and phage (CF: 30.1% (n=609/2023), non-CF: 48.1%
177 (n=1089/2265)) in both niches (Figure 1A, 1B, 1C & Table 1).
178 Overall, *P. aeruginosa* contained an average of 7 DSs, with 99.7 % (n=4274/4288)
179 of isolates containing at least 1 DS, consistent with previous studies (6, 18)(Figure 1B,
180 1C, Supplementary Table 1 & 2). Numerous DSs (n=131) were found, with RM Type I being
181 the frequently found system, present in 72 % (n=3081/4288) of genomes. This is
182 consistent with published observations stating that RM systems are present in >74% of
183 prokaryotic genomes, with an average of ≥2 RM systems per genome (4). To explore
184 whether any DS was associated with either the CF or non-CF niche, we compared the

185 proportion of genomes containing each of the DSs per niche, which revealed that 31%
186 (n=41/131) of DSs detected in this dataset were significantly associated with niche; 37%
187 (n=15/41) with the CF niche and the remainder with the non-CF niche (Bonferroni-
188 corrected Fisher's Exact Test, $p < 0.05$). The DSs associated most significantly with CF
189 and non-CF niches were DarTG and Druantia II respectively (Figure 1C). CRISPR-Cas
190 Type I-F was found in the most genomes within the CF niche and was significantly
191 associated with CF isolates, while RM Type I was the most common and significantly
192 associated with non-CF isolates.

193 To ensure changes in accessory genome content did not result from differences
194 in total genome content between the two niches, we then ascertained AGE content while
195 adjusting for genome size. Specifically, we found that *P. aeruginosa* genomes isolated
196 from CF patients had significantly smaller genomes (t-test, $p<0.05$) and overall
197 contained fewer DSs, anti-defence systems, plasmids, ARGs and phage (Wilcoxon rank-
198 sum test, $p<0.05$) (Figure 1B, Supplementary Figure 1). These CF-associated depletions
199 in AGEs were true, even when frequencies were adjusted to frequencies per kilobase of
200 genome (t-test, $p<0.05$, Supplementary Figure 1). These findings are congruent with
201 previous studies that show isolates from CF patients exhibit genome reduction
202 compared to non-CF isolates as host-specialisation is more advanced in bacteria
203 adapting to chronically infected lungs compared to strains causing more acute
204 infections (17), and is consistent with a broader trend of genomic streamlining in host-
205 adapting pathogens (45-50).
206

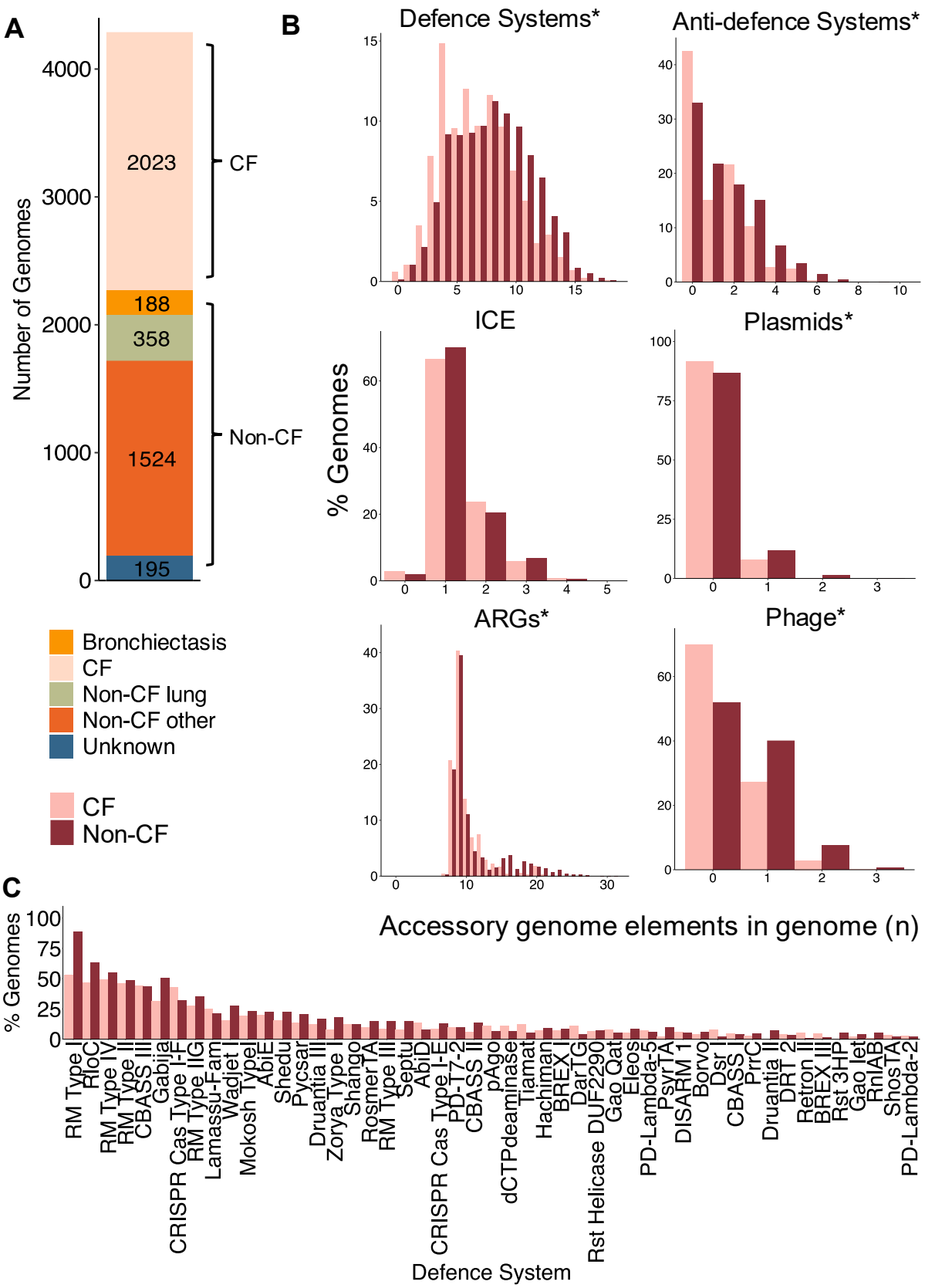


Figure 1. The distribution of accessory genome elements in *P. aeruginosa*. (A) Isolation niches for genomes in dataset (*n* = 4289). **(B)** The number of accessory genome elements (DSs, anti-DSs, ICEs, plasmids, ARGs and phages) per genome is shown as a

211 proportion of genomes containing that number of element, stratified by CF/non-CF
212 niche. Significant differences between the proportion of isolates containing any number
213 of accessory genome elements from CF and non-CF niche's is denoted with a *
214 (Wilcoxon rank sum test $p<0.05$). **(C)** The proportion of genomes containing the 50 most
215 prevalent DSs by niche (see Supplementary Table 1 for full dataset). ICE = Integrated
216 Conjugative Element ARGs = Antimicrobial Resistance Genes CF = cystic fibrosis DS =
217 Defence System**.**
218
219 **Table 1. Summary of differences between genomes from CF and non-CF patients**

Feature	Niche	Mean	Range (Min-Max)	IQR (Q1-Q3)
Isolates	CF	2023	-	-
	non-CF	2265	-	-
Genome Size (Mb)	CF	6.6	5.9-7.5	6.4-6.5
	non-CF	6.7	5.8-7.7	6.5-6.9
DSs	CF	6.9	0-18	4-9
	non-CF	7.9	0-18	5-10
Anti-defence Systems	CF	1.4	0-7	0-2
	non-CF	1.6	0-9	0-3
ARGs	CF	10	7-21	9-10
	non-CF	11	7-31	9-12
Plasmids	CF	0.08	0-2	0-0
	non-CF	0.2	0-3	0-0
ICEs	CF	1.4	0-5	1-2
	non-CF	1.3	0-4	1-2
Phages	CF	0.3	0-3	0-1
	non-CF	0.6	0-3	0-1

220
221 <u>Determining defence system interactions</u>
222 To identify putative synergistic or antagonistic DS interactions across *P.*
223 *aeruginosa*, we explored which DSs were more frequently associated (putative
224 synergism) or dissociated (putative antagonism) with each other than expected by
225 chance and independent of evolutionary relationships. For this we employed CoinFinder
226 (43), which identifies phylogenetically independent associations and dissociations
227 between traits (in our case, accessory genome categories of elements) and also
228 estimates the phylogenetic independence (as measured by D-value) of traits.

229 This association analysis revealed 182 pairwise interactions among DSs (27.9%,
230 n=182 of 653 possible pairs) that were significantly associating or dissociating with one
231 another (Bonferroni-corrected binomial exact test $p<0.01$, Figure 2, Supplementary
232 Table 3 & 4). Association interactions made up 85.2% (n=155/182) of the total significant
233 DS pairs and the remaining interactions were dissociations. Overall, the dominance of
234 associations over dissociations supports current hypotheses that DS complement each
235 other (14, 18, 51, 52). Each interacting DS (n=36) had a variable number of associations
236 and dissociations, with an average of 10 interactions (range 2-19), and RM Type I having
237 the most interactions (n=19). The high frequency of interactions between RM Type I and
238 other systems may reflect that RM shapes the DS complement of genomes. Other DSs
239 also exhibited high numbers of associations, such as RM Type IV, Pycsar and AbiE, which
240 all associated with 17 other DSs (Figure 1C, 2 & Supplementary Table 3 & 4).
241 To focus our attention on those associations that may have significant roles in
242 shaping *P. aeruginosa* populations, we triaged DSs to focus on those that were
243 comparatively mobile and more common. As a proxy for mobility, we used the
244 phylogenetic independence (D-value) of each DS. As a lower D-value indicates a strong
245 phylogenetic conservation we considered DSs with a D-value of over 0 (Figure 2A) in our
246 downstream analysis. Additionally, we filtered the interactions to only include DSs that
247 were present in over 400 (9.3%, n= 400/4288 of the total dataset, full data in
248 Supplementary Tables 3 & 4) genomes (Figure 2A). Of the 15 DSs that passed these
249 thresholds, 13 were found to both associate and dissociate with other systems, with the
250 remaining 2 having only associations (Figure 2A). Among these 15 DSs, RM Type II, Gabija
251 Lamassu and AbiE were the most frequent and exhibited the most phylogenetic
252 independence (Figure 2A).
253 Having triaged on frequency and phylogenetic independence, we then sought to
254 interrogate DS interactions by visualising the effect size (inferred from the ratio of
255 observed vs expected occurrences, methods) of each association alongside the number
256 of interactions of each DS (Figure 2 B & C, Supplementary Table 3 & 4). This revealed that
257 among the 54 (n= 54/155 total associations) interactions between the 15 triaged DSs,
258 Druantia III and Mokosh Type I association exhibited the strongest interaction (Figure 2B).
259 This indicates a possible mechanistic synergism and is consistent with experimental

260 evidence suggesting Druantia III and Mokosh Type I both protect against similar phages
261 such as T4, through nuclease-based mechanisms (7, 53, 54).
262 We then examined DS pairs that were found to dissociate from each other. As
263 these only made up only 14.8% (n=27/182) of the total significant interactions (Figure 2C
264 & Supplementary Table 4) we applied no triage thresholds. Here we found that CRISPR-
265 Cas Type I-F had the most dissociations with other DSs, such as other RMs Type I & II and
266 Pycsar (Figure 2C & Supplementary Figure 2). Isolates containing CRISPR-Cas Type I-F
267 (n=159/4288) contained fewer DSs, compared to isolates not containing CRISPR-Cas
268 Type I-F (t-test, $p<0.05$) (Supplementary Figure 2). These strong indicators for
269 dissociation of CRISPR-Cas Type I-F are consistent with it being commonly found in *P.*
270 *aeruginosa* and blocking HGT of foreign DNA and even other DS (5, 55, 56).
271 As recent studies have shown that DSs are frequently encoded within horizontally
272 transferrable defence islands or genomic hotspots (3, 9-11, 53) we sought to explore
273 whether the DS associations resulted from co-localisation. Therefore, we assessed
274 whether the significant associating and dissociating DS pairs (n= 77, Figure 2B & C) were
275 present on the same contiguous sequence (contig) in our draft genome assemblies (as
276 an imperfect but practicable proxy of co-localisation, Figure 3). This revealed that of all
277 DS containing contigs, only 29% (n=7086/24211) carried two or more DSs. Within the
278 7086 contigs containing ≥2 DS, associating DS pairs were co-located in only 22% of cases
279 i.e. were on the same contig 4,367 times out of a total of 19,472 pairs occurring across
280 genomes (Figure 3). As a counterfactual for understanding the rates of co-localisation
281 among associating DS pairs, we examined co-localisation rates among dissociating DS
282 pairs. As anticipated, the rates of co-localisation were far lower i.e. 0.5%, n=175/37,098
283 pairs across all genomes, though this varied among interacting pairs (Figure 3). We found
284 the rate of co-localisation for 22 (of 54) associating pairs was lower than 0.034% (with
285 0.034% being the highest co-localisation proportion for a dissociating pairs) (Figure 3).
286 Overall, this indicated that associations detected here are phylogenetically independent
287 and are unlikely to have resulted from co-localisation. We observed that some
288 associating DS, such as Shedu and Pycsar, were less-frequently co-located than
289 systems that were dissociated in the population (Figure 3 & Supplementary Figure 3).
290 In summary, our analyses of DS interactions in *P. aeruginosa* revealed a complex
291 network of both association and dissociation interactions. Most significant interactions

292 were associations, with DSs such as RM Type II, Gabija, Lamassu and AbiE frequently
293 associating with other systems. Strong interactions between DSs may point towards
294 functional synergy. Conversely, frequent dissociations with CRISPR-Cas Type I-F
295 indicates its role in restricting HGT and conferring a selective disadvantage to co-
296 occurring DSs. Although co-localisation of DSs on contigs was generally low, some
297 associations such as Mokosh Type I, and RM Type IV were consistently (n=415/415
298 genomes containing the pair) co-located. While such co-localisation can suggest
299 potential functional synergy (52, 57), in this case, the phylogenetically independent co-
300 occurrence patterns could indicate that their co-inheritance is likely driven by genetic
301 linkage rather than necessarily reflecting synergistic interactions between the systems
302 (Figure 2A). These findings highlight the dynamic interplay between DSs, with
303 implications for understanding how they may shape the adaptive landscape of *P.*
304 *aeruginosa* populations.
305

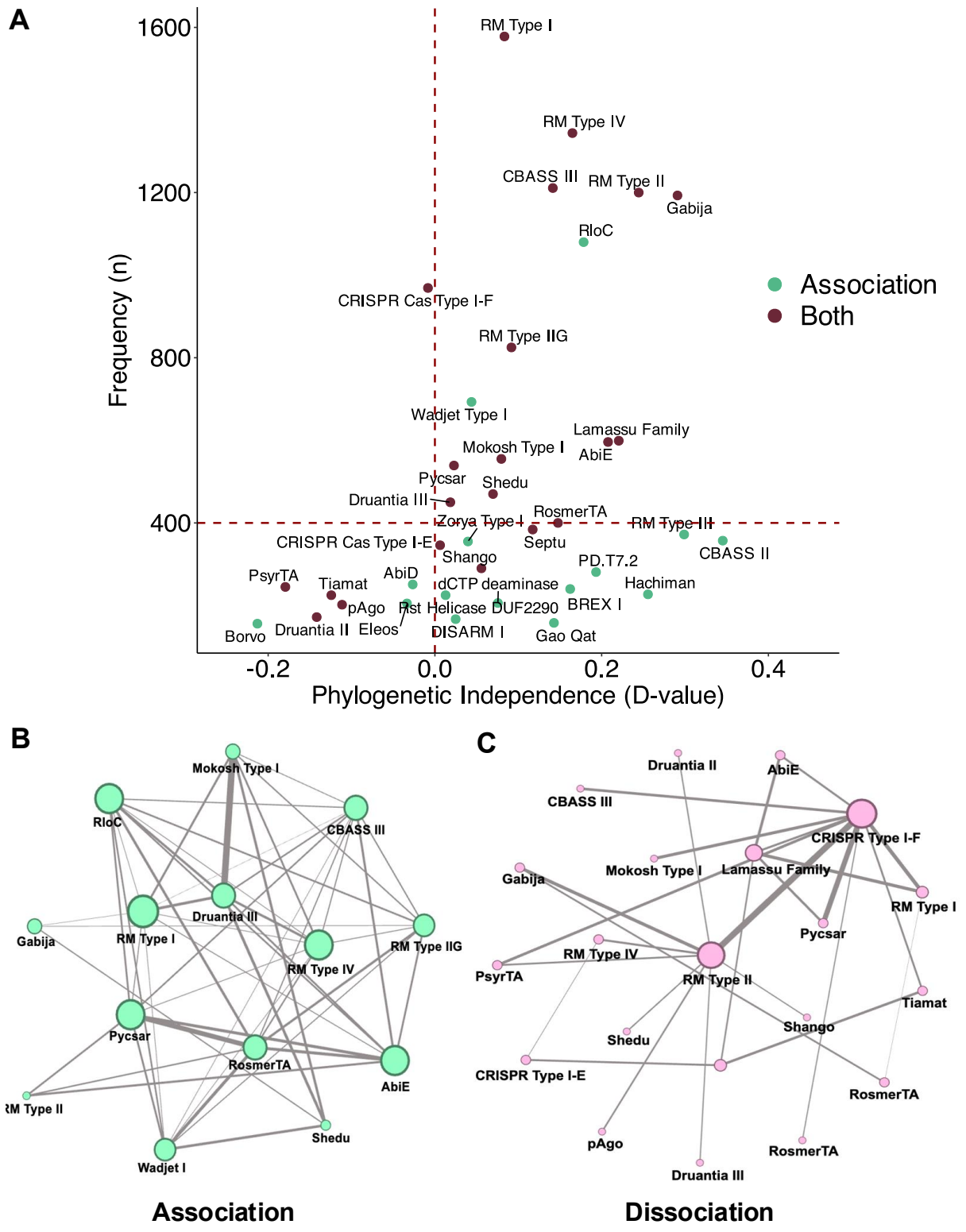


Figure 2. The distributions, associations and dissociations of defence systems in the global *P. aeruginosa* population. (A) The frequency and phylogenetic independence of DSs with significant interactions (i.e. associations or dissociations). Green dots highlight DSs that were found to have significant associations while purple dots indicate DSs that had both significant associations and dissociations (no systems were found to only have

significant dissociations, full data in Supplementary Table 3 & 4). Red dashed lines show triage cut-off values of 0 and 400 for Phylogenetic independence and Frequency respectively. **(B)** Interaction network of associations among DSs. Each node depicts a DS scaled in size by the number of significant associations it had. Edges represent individual associations between DSs with the thickness corresponding to the effect size (i.e. strength of association). **(C)** Interaction network of DS dissociations. Each node similarly depicts a DS scaled in size by the number of significant dissociations it had. Edges represent individual dissociations between DSs with thickness corresponding to effect size.

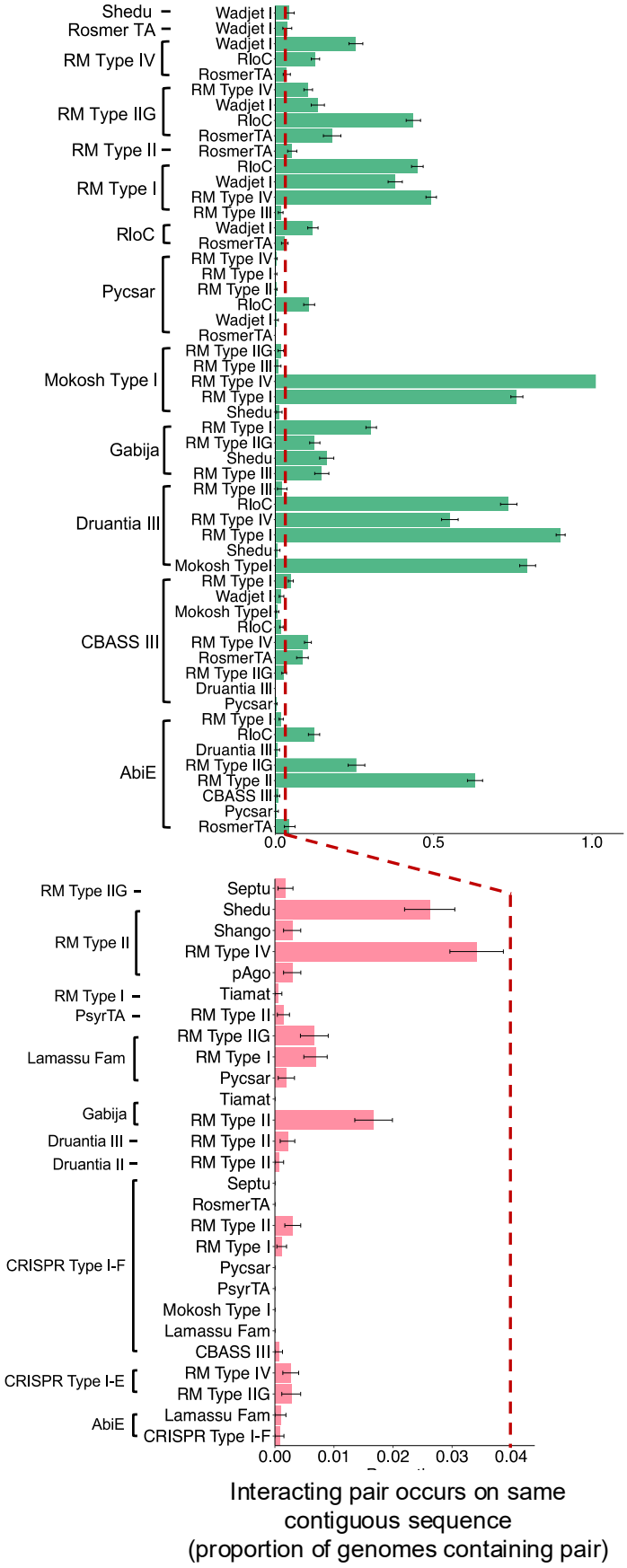


Figure 3. The proportion of pairs of DSs showing interactions of association (upper) and dissociation (lower) that occur on the same contiguous sequences in draft

325 **assemblies.** Axes labels are ordered alphabetically. Error bars represent standard error.
326 The red dashed line intersects the x-axis at 0.04 on both plots to highlight the difference
327 in scale.

328 Interactions of DS with other accessory genome elements

329 To explore the complexity of integral links among DSs and other AGEs, including
330 the recently described anti-defence systems, we used similar methods to characterise
331 the interactions among DSs and all other AGEs. This analysis revealed a total of 591
332 significant interactions between specific AGEs distributed across accessory genome
333 feature categories (Figure 4 & Supplementary Table 3 & 5). As with DS comparisons, we
334 triaged the interactions by frequency (using those present in 7%, n=300/4288) of each
335 element across the dataset and phylogenetic independence (D-value > 0) (Figure 4A).
336 Analysis of the 36 (of 68) AGEs meeting these triage criteria revealed a network with 201
337 associations of various effect sizes and differences in the number of associations
338 (ranging between 2 and 31 per element) (Figure 4A & Supplementary Table 3 & 5). Of the
339 201 pairs passing the threshold, anti-CRISPR AcrIF3 and anti-CRISPR AcrIF5 show the
340 strongest association (Figure 4A & Supplementary Table 5). Other comparatively strong
341 associations included those between ARGs *qacE*Δ and *sul1*, and DS AbiE and ICE bhp.sal
342 (Figure 4A & Supplementary Table 5).
343 Regarding the dissociation interactions among these individual AGEs, CRISPR-
344 Cas Type I-F was again found to have the most negative interactions (29.8%, n=17/57
345 total dissociations with no triage criteria applied). The significant dissociation between
346 CRISPR-Cas Type I-F and RM Type II was also exhibited in this analysis, further supporting
347 the observation that they may not act synergistically in *P. aeruginosa* (Figure 4B &
348 Supplementary Table 5). The opposite is seen for the Gram-positive bacteria *Listeria*,
349 *Streptococcus* and *Staphylococcus*, where CRISPR-Cas Type IV and VI systems have
350 been observed to provide synergistic protection against phage infection with RM systems
351 suggesting synergy maybe system type specific and species specific (58-60). We also
352 observed that the ARG blaPDC3 had 2 strong dissociations with both RM Type IIG and
353 blaOXA-488 (Figure 4B & Supplementary Table 5). Notably, we also observed distinct
354 patterns of interaction between DSs and their cognate inhibitors, where anti-CRISPR
355 systems associate with CRISPR-Cas Type I-E but dissociate from CRISPR-Cas Type I-F

356 and anti-CBASS systems associate with CBASS systems. Hence, the phylogenetically
357 independent associations and dissociations we have found among individual AGEs in
358 this analysis represent putative interactions for which there is likely a strong mechanistic
359 or ecological basis.

360 To draw out broader trends in interactions among this plethora of AGEs, we then
361 sought to explore the broader dynamics and relative importance of interactions among
362 different feature categories of the AGEs. To do this, AGEs were collapsed into feature
363 categories (i.e. ARGs, DS, phages, etc) and the associations examined. Overall, this
364 analysis revealed more associations than dissociations, with 90.4% (n=534) of the 591
365 interactions being associations. Among accessory genome features, we found the
366 highest numbers of associations between ARGs and DSs (31.1%, n=166/534 interacting
367 pairs) followed by DSs and DSs (27% n=144/534) and anti-defence systems and DSs
368 (n=11% 60/534, Figure 4C). Conversely, ARGs and DSs (35.1% n= 20/57) along with DSs
369 and DSs (45.6% n=26/57) had the most dissociations (Figure 4C). These analyses suggest
370 that the accessory genome landscape of *P. aeruginosa* has many ARG-DS and DS-DS
371 interactions.

372 To move these results closer to meaningful insights on the relative importance of
373 accessory genome feature interactions within *P. aeruginosa*, we contextualised these
374 results by adjusting for the total number of detected AGEs in each feature category (to
375 account for possible bias arising from database size differences). Specifically, we
376 assessed the number of significantly interacting pairs between accessory genome
377 feature categories as a proportion of total theoretically possible pairs enumerated from
378 the two sets of detected AGEs (e.g. for ARG to DS interactions, there were a total of 166
379 significant associations among a theoretical total of 58,819 possible interactions,
380 enumerated from the detection of 449 unique ARGs and 131 unique DS, giving a
381 proportion of 0.003 significant pairs (Methods, Figure 4D). Overall, this revealed that DS
382 association interactions were the most often significant, and associated with anti-
383 defence systems and ICEs were also high (Figure 4D). Self-interactions between anti-
384 defence systems as well with DSs were also high. Dissociation interactions, although
385 lower in abundance, were more frequently significant between DS-DS, anti-defence-DS,
386 and plasmid-DS pairs (Figure 4D). Hence, our results suggest that the interactions of DSs
387 and anti-defence systems are important interactions among the interplay of the

accessory genome, likely shaping the composition of bacterial genomes, impacting genetic diversity, AMR, and evolution. These findings are consistent with an apparent continuous arms race between both DSs and anti-defence systems, taking place alongside external selective pressures which affect retention and diversification of the accessory genome (3, 61, 62). Furthermore, the predominance of ICE and plasmid interactions highlights their known role in facilitating HGT of DSs, anti-defence systems, and ARGs; additionally shaping genome composition (3, 63).

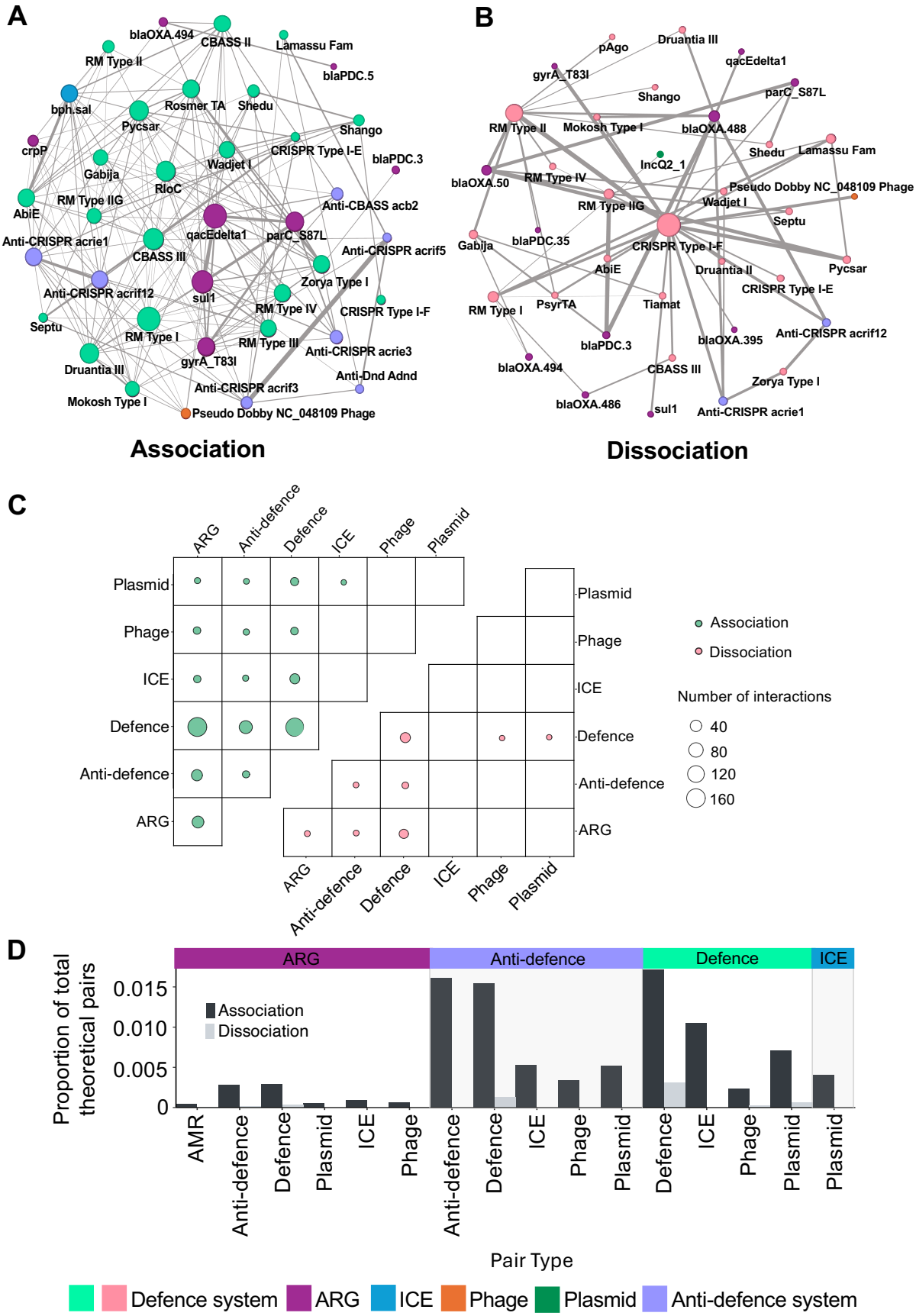


Figure 4. Interactions among accessory genome elements in *P. aeruginosa*. (A) Significant associations network among individual accessory genome elements. Each

399 node depicts a genomic element coloured by type according to the inlaid key. Nodes are
400 scaled by the number of significant associations the element has. Edges represent
401 significant associations between elements with thickness scaled by effect size.
402 Elements have been triaged using a Phylogenetic independence value of ≥0 and being
403 found in >300 genomes. **(B)** Significant dissociations network among individual
404 accessory genome elements. Each node depicts a genomic element coloured by type
405 according to the inlaid key. Nodes are scaled by the number of significant dissociations
406 the element has. Edges represent significant dissociations between elements with
407 thickness scaled by effect size. **(C)** Bubble plots representing significant associations
408 (upper) or dissociations (lower) interactions between pairs of accessory genome feature
409 categories (axes). Bubbles are scaled by the number of significant interactions between
410 those categories. **(D)** The relative importance of AGE interactions is shown as bar plots
411 showing the proportion of feature category pair interactions that were found to be
412 significant associations (dark grey) or dissociations (light grey) with the denominator of
413 the total number of theoretical pairs (to account for database size). Feature category
414 pairs are indicated by groups above the bar plots (coloured according to the inlaid key)
415 and x-axis labels.

416 **Discussion**

417 In this study we investigated AGEs across ecological niches of global *P.*
418 *aeruginosa* population to find biologically meaningful relationships among the densely
419 packed accessory genome.

420 The comprehensive analysis of 4,288 *P. aeruginosa* genomes revealed 31% of
421 detected DSs were significantly associated with bacterial niche, highlighting the role of
422 environmental pressures in shaping genome defence strategies. Notably, the CF niche
423 isolates were enriched for DarTG and CRISPR-Cas Type I-F, while non-CF niche isolates
424 exhibited a higher prevalence of Druantia II and RM-Type I systems. These findings
425 suggest that CF-adapted *P. aeruginosa* may rely on distinct defence mechanisms,
426 potentially reflecting selection for systems that counteract host-adapted phages or
427 environmental stress. Different bacteria DSs may be favoured under distinct ecological
428 conditions, yet identifying the environmental drivers of their distribution in complex
429 environments remains challenging. Although, there are studies that have previously

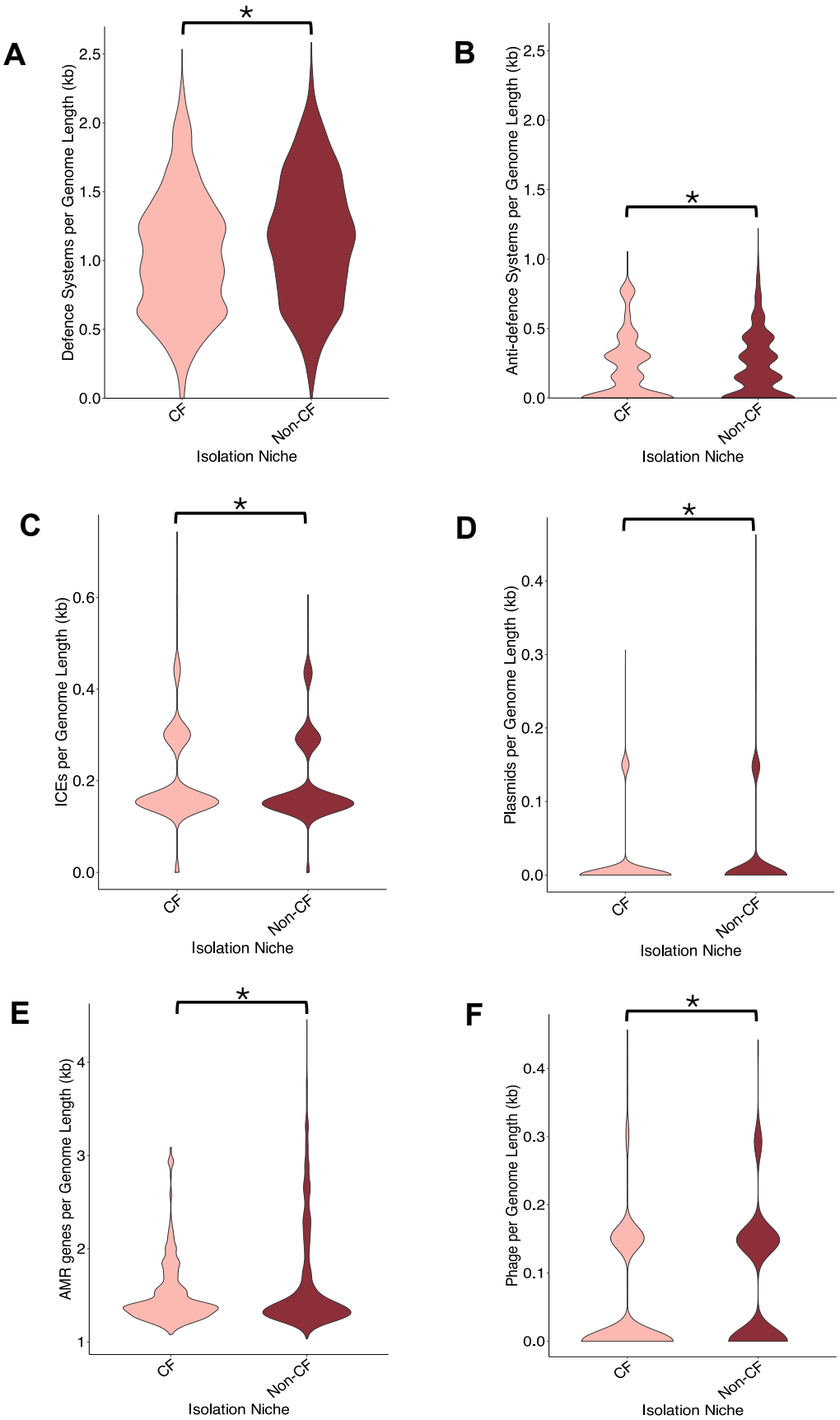
430 linked some ecological factors such as force of infection to the evolution of immune
431 strategies, where bacteria in lower infection risk environments can favour CRISPR
432 immunity compared to surface modification immunity (64, 65). This is supported by
433 increased abundance of CRISPR systems in thermophiles that reside in low infection risk
434 environments (66, 67) and may go some way to explain the enrichment of CRISPR-Cas
435 Type I-F in *P. aeruginosa* isolates in this study. Furthermore, pan-genome differences
436 between different phylogroups of *P. aeruginosa* suggest that variation in defence
437 systems likely modulates the extent and nature of homologous recombination, with
438 more diverse of abundant systems potentially restricting the uptake of exogenous DNA
439 (68). Additionally, CF isolates exhibited significant genome reduction and contained
440 fewer DSs, anti -defence systems, plasmids, ARGs and phages per kilobase of genome
441 length, supporting the hypothesis that long-term adaptation to the host-lung
442 environment drives genomic streamlining, through loss of genes linked to motility,
443 virulence and environmental adaptability due to focused selection pressures in chronic
444 infection.

445 We found a high density of DSs in *P. aeruginosa* genomes, consistent with
446 previous studies, and extended this work by illuminating the many significant
447 interactions among DSs. Associating interactions were far more common than
448 dissociating ones, supporting the hypotheses that DSs function in a complementary
449 manner, working alongside each other to provide a multi-layered approach against
450 invading AGEs, providing protection if other defence mechanisms fail (6, 14, 18, 51-53,
451 69, 70). Distinctly however, we reveal that associations are unlikely due to co-
452 localisations in defence islands.

453 Conversely, CRISPR-Cas Type I-F had the highest number of dissociations,
454 particularly with RM systems and Pycsar. Given its well documented role in blocking
455 HGT, including the acquisition of other DSs, this may explain why genomes containing
456 CRISPR-Cas Type I-F harboured significantly fewer DSs than those without it (55, 56,
457 71).Together, these findings highlight the complex interactions shaping DS organisation
458 and suggests that while many DSs complement one another, CRISPR-Cas Type I-F may
459 act as a gatekeeper, limiting genomic plasticity in *P. aeruginosa* living in potentially
460 nutrient deficient, highly competitive or stressful environments. Hence, the interactions
461 we show among defence systems here represent plausible meaningful biological

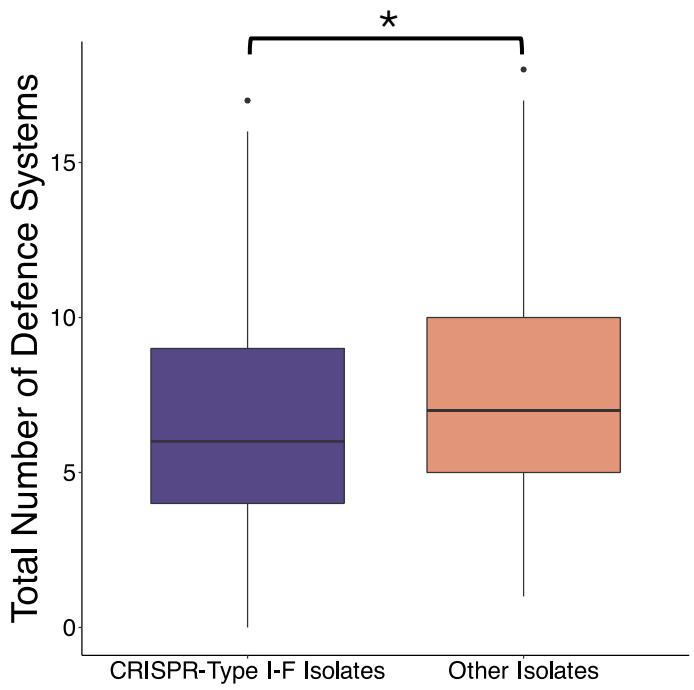
462 mechanistic (or ecological) interactions that occur among DSs that are not attributable
463 to co-inheritance on defence islands (7, 53, 72).
464 Overall, our study of the accessory genome interactions in *P. aeruginosa* revealed
465 a complex network of interactions among accessory genome feature categories. The
466 most frequent interactions were between ARGs and DSs, followed by DS-DS interactions
467 and DS-anti-defence system interactions, suggesting that DSs play a central role in
468 structuring the *P. aeruginosa* accessory genome. Many defence genes are found in the
469 accessory genome and are a key contributor to intraspecific variation of bacteria (73, 74).
470 However, their high turnover rate, co-acquisition with AGEs and persistence in the
471 accessory genome all affect the amount of protection that they offer against HGT (65, 75,
472 76), as will shifting environments, as highlighted above.
473 At a finer scale, out data revealed strong associations between specific AGEs,
474 such as anti-CRISPR AcrIF3 and AcrIF5, as well as the ARG pair *qacE*Δ and *sul1* and the
475 DS AbiE with ICE bhp.sal. These findings reinforce the notion that AGEs interact to
476 promote dissemination and survival in the environment, where AcrIF5 works closely with
477 AcrIF3 to block Cas 2/3 recruitment to ensure that CRISPR-Cas Type I-F cannot degrade
478 foreign DNA (69). The spread of ARGs is facilitated somewhat by DSs, where systems
479 such as Abi, Gabija, CBASS and RM have been found to be more frequently located within
480 or next to AGEs like ICEs, to affect AGE propagation, stability and HGT to confer
481 resistance in bacteria (3, 53, 75, 77, 78). Our study reveals key associations and
482 dissociations that are phylogenetically independent and likely driven by strong
483 mechanistic or ecological interactions.
484 There are several limitations to this study. Firstly, although our dataset includes
485 isolates from diverse ecological niches, many isolates originate from studies on AMR,
486 which may bias genome composition. Our findings are also reliant on the content and
487 curation of various AGE databases, so may have failed to detect significant interactions
488 with yet undescribed AGEs. Furthermore, our association interactions appeared to be
489 robust against incidental co-localisation but ascertaining the full impact of co-
490 localisation would be better evaluated with complete genomes assembled from long-
491 read sequencing data (e.g. as in Wu et al., 2024). Another limitation is that we relied
492 solely on DefenseFinder (6) to identify DSs, which may not fully capture the diversity of
493 all DSs. As a result, we may have missed relevant interactions due to incomplete or

494 imperfect annotations. Our *in-silico* investigation into putative synergism or antagonism
495 among AGEs in *P. aeruginosa* does not offer explanations of the mechanistic or
496 ecological basis for those interactions. However, the approach is robust to many
497 confounding factors (e.g. phylogenetic dependence and the impact of genome size and
498 co-localisation), and many had a plausible mechanistic basis, so we put forward these
499 interactions as strong candidates for future research on biologically meaningful
500 interactions among the accessory genome milieu in *P. aeruginosa*.
501 The results of this work highlight the importance and complexity of accessory
502 genome interactions in bacteria, particularly the expanding array of DSs and anti-
503 defence systems in shaping accessory genome interactions. Multiple association and
504 dissociation interactions between DSs, and other accessory genome features were
505 found and suggest adaptation of *P. aeruginosa* is driven by the selective forces in their
506 environment. Future research should aim to explore and validate these putative
507 interactions both phenotypically and mechanistically, and to look for other possible
508 drivers of synergy, such as co-regulation.
509
510 **Supplementary Material**
511 **Supplementary Table 1**. *Pseudomonas aeruginosa* isolates, associated metadata and
512 accessory genome feature content used in this study.
513 **Supplementary Table 2.** DefenseFinder and Anti-defenseFinder output from *P.*
514 *aeruginosa* dataset.
515 **Supplementary Table 3.** Accessory genome elements found to significantly interact by
516 Coinfinder, their associated D-value, frequencies and interactions.
517 **Supplementary Table 4**. List of DSs pairs found to associate and dissociate by
518 Coinfinder.
519 **Supplementary Table 5**. List of accessory genome feature pairs found to associate and
520 dissociate by Coinfinder.



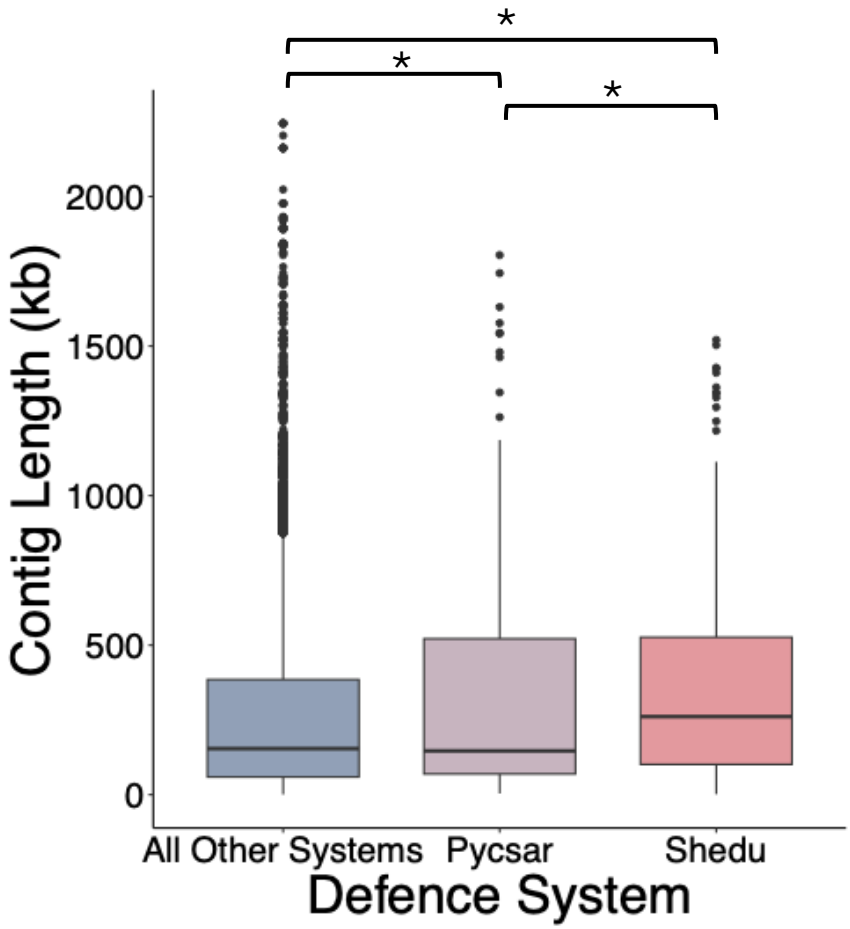
Supplementary Figure 1. Number of defence systems **(A)**, anti-defence systems **(B)**, ICEs **(C)**, plasmids **(D)**, ARGs **(E)** and phage **(F)** present in *P. aeruginosa* per genome

524 length (kb) of genomes isolated from CF patients and non-CF sources. Significant
525 differences in number of accessory genome elements per genome length (kb) between
526 genomes isolates from CF and non-CF niches is denoted with a * (*t*-test $p<0.05$). ICE =
527 Integrated Conjugative Element ARGs = Antimicrobial Resistance Genes CF = cystic
528 fibrosis DS = Defence System.



To evaluate the possible dominant negative association of CRISPR-Cas Type I-F, we determined that those genomes in this study containing CRISPR-Cas Type I-F (n=159/4288) contained fewer DSs 6.7 (range 0-17), compared with 7.2 (range 1-18) in non-CRISPR-Cas Type I-F containing genomes (t-test, $p<0.05$).
529

530 **Supplementary Figure 2**. Comparison of the number of DSs found in *P. aeruginosa*
531 isolates containing the DS CRISPR-Type I-F (n=1591) and isolates that did not (n=2683).
532 The plot shows the median (line), interquartile ranges (box), and range (whiskers) of the
533 number for each group. Outliers are shown as individual markers. The asterisk
534 indicates a statistically significant difference ($p<0.05$) as measured by a *t*-test.



Shedu and Pycsar, which were on average found to co-localise in lower proportions, compared to other non-Pycsar/Shedu containing pairs (average proportion of co-localisation in DS pairs containing, Pycsar = 0.01, Shedu = 0.06 and all other system pairs = 0.2. To confirm that these apparently low-rates of non-co-localisation of Pycsar and Shedu were not the result of disproportionately shorter contigs, we assessed the average contig length (kb) of sequences containing these DSs relative to other DS-containing contigs. This revealed that Shedu or Pycsar containing contigs (average of 347 kb and 316 kb respectively) were actually slightly longer than contigs containing other DSs (average of 262 kb) (Kruksall-Wallis rank sum test and Dunn's post hoc test = $p<0.05$) suggesting that this is a genuine interaction between Shedu and Pycsar, and other systems and might represent either functional redundancy or mechanistic antagonism with other systems.

536 **Supplementary Figure 3**. Comparison of length in kilobases of contiguous sequences
537 containing either Pycsar (n=739), Shedu (n=830) or all other DS (n=33650). The plot
538 shows the median (line), interquartile ranges (box), and range (whiskers) of the number
539 for each group. Outliers are shown as individual markers. The asterisk indicates a
540 statistically significant difference ($p<0.05$) as measured by a Kruksall-Wallis rank test
541 followed by a Dunn's post hoc test for pairwise comparisons.

542 **Funding Information**

543 This work was supported by a grant from the Biotechnology and Biological
544 Sciences Research Council sLoLa BB/X003051/1, awarded to K.S.B, M.A.B, M.D.S and
545 E.R.W and a Wellcome Discovery Award (226602/Z/22/Z) to R.A.F and A.W.

546 **Author contributions**

547 C.E.C and K.S.B contributed to the writing of the original draft. All authors (C.E.C,
548 A.W, A.A, J.L.F, M.A.B, J.P, R.A.F, M.D.S, E.R.W, M.C and K.S.B) contributed to reviewing
549 and editing of subsequent drafts. C.E.C was responsible for formal analysis,
550 investigation, visualisation and methodology. A.W, J.P, R.A.F, and C.E.C. contributed to
551 data curation. K.S.B, M.A.B, M.D.S, C.E.C, and E.R.W contributed to funding acquisition
552 and conceptualisation. K.S.B and E.R.W were responsible for project administration.
553 K.S.B was responsible for supervision.

554 **Acknowledgements**

555 We thank all current and past members of the Multi-Defence Consortium for
556 providing scientific support.

557

558 **Competing Interests Statement**

559 E.R.W. is inventor on patent GB2303034.9.

560 **Data availability**

561 Short-read DNA was downloaded from the European Nucleotide Archive (ENA) by
562 Weimann et al. (2024)(17). ENA run accessions can be found in Supplementary Table 1.

563

564 **References**

1. Vos M, Buckling A, Kuijper B, Eyre-Walker A, Bontemps C, Leblond P, et al. Why do mobile genetic elements transfer DNA of their hosts? Trends in Genetics. 2024;40(11):927-38.
2. Arnold BJ, Huang IT, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. Nature Reviews Microbiology. 2022;20(4):206-18.
3. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol. 2020;18(2):113-9.
4. Oliveira PH, Touchon M, Rocha EPC. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Research. 2014;42(16):10618-31.
5. Watson BNJ, Staals RHJ, Fineran PC. CRISPR-Cas-Mediated Phage Resistance Enhances Horizontal Gene Transfer by Transduction. mBio. 2018;9(1):10.1128/mbio.02406-17.
6. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nature Communications. 2022;13(1):2561.
7. Millman A, Melamed S, Leavitt A, Doron S, Bernheim A, Hör J, et al. An expanded arsenal of immune systems that protect bacteria from phages. Cell Host Microbe. 2022;30(11):1556-69.e5.
8. Maestri A, Pons BJ, Pursey E, Chong CE, Gandon S, Custodio R, et al. The bacterial defense system MADS interacts with CRISPR-Cas to limit phage infection and escape. Cell Host & Microbe. 2024.
9. Johnson MC, Laderman E, Huiting E, Zhang C, Davidson A, Bondy-Denomy J. Core defense hotspots within *Pseudomonas aeruginosa* are a consistent and rich source of anti-phage defense systems. Nucleic Acids Res. 2023;51(10):4995-5005.
10. Rousset F, Depardieu F, Miele S, Dowding J, Laval A-L, Lieberman E, et al. Phages and their satellites encode hotspots of antiviral systems. Cell host & microbe. 2022;30(5):740-53. e5.
11. Hochhauser D, Millman A, Sorek R. The defense island repertoire of the *Escherichia coli* pan-genome. PLOS Genetics. 2023;19(4):e1010694.
12. Novick RP, Ram G. The Floating (Pathogenicity) Island: A Genomic Dessert. Trends in Genetics. 2016;32(2):114-26.
13. Koonin EV, Makarova KS, Wolf YI, Krupovic M. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. Nature Reviews Genetics. 2020;21(2):119-31.
14. Wu Y, Garushyants SK, van den Hurk A, Aparicio-Maldonado C, Kushwaha SK, King CM, et al. Bacterial defense systems exhibit synergistic anti-phage activity. Cell Host & Microbe. 2024;32(4):557-72.e6.
15. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev. 2018;31(4).
16. Rocha EPC, Bikard D. Microbial defenses against mobile genetic elements and viruses: Who defends whom from what? PLOS Biology. 2022;20(1):e3001514.
17. Weimann A, Dinan AM, Ruis C, Bernut A, Pont S, Brown K, et al. Evolution and host-specific adaptation of *Pseudomonas aeruginosa*. Science. 2024;385(6704):eadi0908.

18. Costa AR, van den Berg DF, Esser JQ, Muralidharan A, van den Bossche H, Bonilla BE, et al. Accumulation of defense systems in phage-resistant strains of *Pseudomonas aeruginosa*. Sci Adv. 2024;10(8):eadj0341.
19. World Health Organisation (WHO). WHO Bacterial Priority Pathogens List, 2024 2024 [Available from: https://iris.who.int/bitstream/handle/10665/376776/9789240093461-eng.pdf?sequence=1.
20. Botelho J, Tüffers L, Fuss J, Buchholz F, Utpatel C, Klockgether J, et al. Phylogroup-specific variation shapes the clustering of antimicrobial resistance genes and defence systems across regions of genome plasticity in *Pseudomonas aeruginosa*. eBioMedicine. 2023;90:104532.
21. Kos VN, Déraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, et al. The resistome of *Pseudomonas aeruginosa* in relationship to phenotypic susceptibility. Antimicrobial agents and chemotherapy. 2015;59(1):427-36.
22. Chilam J, Argimón S, Limas MT, Masim ML, Gayeta JM, Lagrada ML, et al. Genomic surveillance of *Pseudomonas aeruginosa* in the Philippines, 2013–2014. Western Pacific surveillance and response journal: WPSAR. 2021;12(2):4.
23. del Barrio-Tofiño E, López-Causapé C, Cabot G, Rivera A, Benito N, Segura C, et al. Genomics and susceptibility profiles of extensively drug-resistant *Pseudomonas aeruginosa* isolates from Spain. Antimicrobial agents and chemotherapy. 2017;61(11):10.1128/aac. 01589-17.
24. Khaledi A, Weimann A, Schniederjans M, Asgari E, Kuo TH, Oliver A, et al. Predicting antimicrobial resistance in *Pseudomonas aeruginosa* with machine learning-enabled molecular diagnostics. EMBO molecular medicine. 2020;12(3):e10264.
25. Stapleton PJ, Izydorcyzk C, Clark S, Blanchard A, Wang PW, Yau Y, et al. *Pseudomonas aeruginosa* strain-sharing in early infection among children with cystic fibrosis. Clinical Infectious Diseases. 2021;73(9):e2521-e8.
26. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. Nature genetics. 2015;47(1):57-64.
27. Reynolds R, Hope R, Williams L. Survey, laboratory and statistical methods for the BSAC Resistance Surveillance Programmes. Journal of antimicrobial chemotherapy. 2008;62(suppl_2):ii15-ii28.
28. Hilliam Y, Moore MP, Lamont IL, Bilton D, Haworth CS, Foweraker J, et al. *Pseudomonas aeruginosa* adaptation and diversification in the non-cystic fibrosis bronchiectasis lung. European Respiratory Journal. 2017;49(4).
29. Moore MP, Lamont IL, Williams D, Paterson S, Kukavica-Ibrulj I, Tucker NP, et al. Transmission, adaptation and geographical spread of the *Pseudomonas aeruginosa* Liverpool epidemic strain. Microbial genomics. 2021;7(3):000511.
30. López-Causapé C, Sommer LM, Cabot G, Rubio R, Ocampo-Sosa AA, Johansen HK, et al. Evolution of the *Pseudomonas aeruginosa* mutational resistome in an international cystic fibrosis clone. Scientific reports. 2017;7(1):5555.
31. Turton JF, Wright L, Underwood A, Witney AA, Chan Y-T, Al-Shahib A, et al. High-resolution analysis by whole-genome sequencing of an international lineage (sequence type 111) of *Pseudomonas aeruginosa* associated with metallo-carbapenemases in the United Kingdom. Journal of clinical microbiology. 2015;53(8):2622-31.

32. Freschi L, Vincent AT, Jeukens J, Emond-Rheault J-G, Kukavica-Ibrulj I, Dupont M-J, et al. The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. Genome biology and evolution. 2019;11(1):109-20.
33. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research. 2008;18(5):821-9.
34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology. 2012;19(5):455-77.
35. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, et al. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microbial genomics. 2016;2(8):e000083.
36. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
37. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. Nucleic Acids Research. 2018;47(D1):D660-D5.
38. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903.
39. Seemann T. Abricate. Github.
40. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nature Biotechnology. 2024;42(8):1303-12.
41. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res. 2023;51(W1):W443-w50.
42. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021;11(1):12728.
43. Whelan FJ, Rusilowicz M, McInerney JO. Coinfinder: detecting significant associations and dissociations in pangenomes. Microb Genom. 2020;6(3).
44. Bastian M, Heymann S, Jacomy M, editors. Gephi: an open source software for exploring and manipulating networks. Proceedings of the international AAAI conference on web and social media; 2009.
45. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, et al. Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. Proceedings of the National Academy of Sciences. 2015;112(3):863-8.
46. Bengtsson RJ, Dallman TJ, Allen H, Silva PMD, Stenhouse G, Pulford CV, et al. Accessory Genome Dynamics and Structural Variation of *Shigella* from Persistent Infections. mBio. 2021;12(2):10.1128/mbio.00254-21.
47. Grote A, Earl AM. Within-host evolution of bacterial pathogens during persistent infection of humans. Current Opinion in Microbiology. 2022;70:102197.
48. Miles SL, Santillo D, Torraca V, Jiménez ATL, Jenkins C, Baker S, et al. Enhanced virulence and stress tolerance are signatures of epidemiologically successful *Shigella sonnei*. bioRxiv. 2025:2025.02.05.636615.

49. Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, et al. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nature genetics. 2003;35(1):32-40.
50. Oré N, Tekaia FR, Ti AT, WHEELER PR, CHURCHER C, DA V ID HS, et al. The decaying genome of *Mycobacterium leprae*. Lepr Rev. 2001;72(4):387-98.
51. Agapov A, Baker KS, Bedekar P, Bhatia RP, Blower TR, Brockhurst MA, et al. Multi-layered genome defences in bacteria. Current Opinion in Microbiology. 2024;78:102436.
52. Tesson F, Bernheim A. Synergy and regulation of antiphage systems: toward the existence of a bacterial immune system? Current Opinion in Microbiology. 2023;71:102238.
53. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. Science. 2018;359(6379):eaar4120.
54. Gao L, Altae-Tran H, Böhning F, Makarova KS, Segel M, Schmid-Burgk JL, et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. Science. 2020;369(6507):1077-84.
55. Wheatley RM, MacLean RC. CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. The ISME Journal. 2021;15(5):1420-33.
56. Elliott JFK, McLeod DV, Taylor TB, Westra ER, Gandon S, Watson BNJ. Conditions for the spread of CRISPR-Cas immune systems into bacterial populations. The ISME Journal. 2024;18(1).
57. Payne LJ, Hughes TCD, Fineran PC, Jackson SA. New antiviral defences are genetically embedded within prokaryotic immune systems. bioRxiv. 2024:2024.01.29.577857.
58. Williams MC, Reker AE, Margolis SR, Liao J, Wiedmann M, Rojas ER, et al. Restriction endonuclease cleavage of phage DNA enables resuscitation from Cas13-induced bacterial dormancy. Nat Microbiol. 2023;8(3):400-9.
59. Dupuis M-È, Villion M, Magadán AH, Moineau S. CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. Nature communications. 2013;4(1):1-7.
60. Kim H, Marraffini LA. Cas9 interaction with the tracrRNA nexus modulates the repression of type II-A CRISPR-cas genes. Nucleic Acids Research. 2024;52(17):10595-606.
61. Niault T, van Houte S, Westra E, Swarts DC. Evolution and ecology of anti-defence systems in phages and plasmids. Current Biology. 2025;35(1):R32-R44.
62. Murtazalieva K, Mu A, Petrovskaya A, Finn RD. The growing repertoire of phage anti-defence systems. Trends in Microbiology. 2024;32(12):1212-28.
63. Samuel B, Mittelman K, Croitoru SY, Ben Haim M, Burstein D. Diverse anti-defence systems are encoded in the leading region of plasmids. Nature. 2024;635(8037):186-92.
64. Westra Edze R, van Houte S, Oyesiku-Blakemore S, Makin B, Broniewski Jenny M, Best A, et al. Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. Current Biology. 2015;25(8):1043-9.
65. van Houte S, Buckling A, Westra Edze R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. Microbiology and Molecular Biology Reviews. 2016;80(3):745-63.

66. Anderson RE, Brazelton WJ, Baross JA. Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. FEMS Microbiol Ecol. 2011;77(1):120-33.
67. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct. 2006;1:7.
68. Choudhury Samara T, Piper Kathryn R, Montoya-Giraldo M, Ikhimiukor Odion O, Dettman Jeremy R, Kassen R, et al. Heterogeneity in recombination rates and accessory gene co-occurrence distinguish *Pseudomonas aeruginosa* phylogroups. mSystems. 2025;0(0):e00301-25.
69. Xie Y, Zhang L, Gao Z, Yin P, Wang H, Li H, et al. AcrIF5 specifically targets DNA-bound CRISPR-Cas surveillance complex for inhibition. Nature Chemical Biology. 2022;18(6):670-7.
70. Beavogui A, Lacroix A, Wiart N, Poulain J, Delmont TO, Paoli L, et al. The defensome of complex bacterial communities. Nature Communications. 2024;15(1):2146.
71. Pursey E, Dimitriu T, Paganelli FL, Westra ER, van Houte S. CRISPR-Cas is associated with fewer antibiotic resistance genes in bacterial pathogens. Philos Trans R Soc Lond B Biol Sci. 2022;377(1842):20200464.
72. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT. A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. Nature Microbiology. 2022;7(10):1568-79.
73. Smith WPJ, Wucher BR, Nadell CD, Foster KR. Bacterial defences: mechanisms, evolution and antimicrobial resistance. Nature Reviews Microbiology. 2023;21(8):519-34.
74. Meaden S, Fineran PC. Bacterial defense islands limit viral attack. Science. 2021;374(6566):399-400.
75. Liu Y, Botelho J, Iranzo J. Timescale and genetic linkage explain the variable impact of defense systems on horizontal gene transfer. Genome Res. 2025;35(2):268-78.
76. Koonin EV, Makarova KS, Wolf YI. Evolutionary Genomics of Defense Systems in Archaea and Bacteria*. Annual Review of Microbiology. 2017;71(Volume 71, 2017):233-61.
77. Gu Q, Zhu X, Yu Y, Jiang T, Pan Z, Ma J, et al. Type II and IV toxin-antitoxin systems coordinately stabilize the integrative and conjugative element of the ICESa2603 family conferring multiple drug resistance in Streptococcus suis. PLOS Pathogens. 2024;20(4):e1012169.
78. Touchon M, Bernheim A, Rocha EPC. Genetic and life-history traits associated with the distribution of prophages in bacteria. The ISME Journal. 2016;10(11):2744-54.