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

Bacterial genomes represent dynamic ecological systems in which mutational selection and dynamic accessory genome element complements drive evolution. Emerging evidence suggests that bacterial defence systems, which protect against

phages and other mobile genetic elements, interact through cooperative, competitive, and antagonistic relationships. These interactions can influence horizontal gene transfer, shape phage susceptibility, and the diversification of genome composition across environments. Recent ecological studies describe non-random co-occurrence and avoidance of defence systems, suggesting that they may emerge from ecological and evolutionary interactions, rather than by chance. This necessitates exploring whether these patterns reflect functional compatibility, shared selection pressures imposed by ecological niche, or if they merely arise from co-localisations of convenience. Understanding these patterns is key to elucidating how the accessory genome evolves and how defence systems constrain or facilitate genome plasticity. To investigate these patterns, in their genomic context, and provide a resource for future investigations, we analysed the distributions of defence systems and other accessory genome elements in a recently curated global dataset of 2,940 *Pseudomonas aeruginosa* genomes. Genomic defence system content varied by ecological niche, with higher numbers per genome in non-cystic fibrosis derived isolates (average n=7.9) compared to cystic fibrosis-derived isolates (average n=6.5). We identified multiple associations (n=426) and dissociations (n=50) among defence systems, and among other accessory genome elements, many of which had a plausible biological explanation. We also quantitated the relative interactions among accessory genome elements which revealed that defence systems and anti-defence systems engage in the most accessory genome interactions, compared with e.g. antimicrobial resistance genes, plasmids, phages, suggesting that systems are a major driving force in the ecological dynamics of the bacterial accessory genome. Together, these patterns provide new insights into the evolutionary forces shaping bacterial genomes, provide a valuable resource of quantitated interactions in a highly curated *Pseudomonas aeruginosa* dataset, and establish a framework for future mechanistic and ecological investigations of defence system interactions.

Keywords

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

Introduction

61 Bacterial evolution is partially driven by the acquisition of beneficial traits like
62 antimicrobial resistance (AMR), often facilitated by horizontal gene transfer (HGT) of
63 mobile genetic elements (MGEs) such as plasmids, integrative conjugative elements
64 (ICEs) and phages (1, 2). In this study, we define these collectively as 'Accessory Genome
65 Elements (AGEs)', encompassing both MGEs (including phage, plasmids and ICEs) and
66 functionally defined elements (such as defence systems (DSs), antimicrobial resistance
67 genes (ARGs) and anti-defence systems (Table 1). These AGEs collectively contribute to
68 bacterial adaptability, genome plasticity, and evolution but their can also affect fitness
69 and, in some cases, cause cell death. In response to this evolutionary pressure, bacteria
70 have developed immune responses against extraneous AGEs. Bacteria can contain
71 many complex defence systems (DSs) to defend against harmful invasion ensuring a
72 balance between adaptability and genome integrity (3). DSs such as restriction-
73 modification (RM) and CRISPR-Cas are powerful contributors to the control of HGT but
74 in some case can stabilise AGEs like phage and plasmids (4, 5).

75 Multiple and diverse DSs are often found within bacterial genomes, with an
76 average of five DSs in over 20,000 sequenced prokaryotic genomes (6, 7). Furthermore,
77 recent studies demonstrate that DSs may be found clustered in defence islands (3, 7-11)
78 which are commonly associated with mobilisable AGEs, and may facilitate their transfer
79 between genomes (12, 13). The proximity to mobile AGEs and resultant genome plasticity
80 offers a fitness benefit to bacteria living and responding to phage-rich environments (3).
81 It has been hypothesised that defence islands form within bacterial genomes due to
82 mechanistic synergism between clustered DSs promoting co-localisation and
83 mobilisation between genomes (11, 14). Such aggregation is consistent with that seen
84 for AMR and pathogenicity genes on mobile AGEs, such as plasmids, ICEs and phage (11,
85 15, 16), suggesting that defence islands may be similarly promoted by evolutionary
86 forces including HGT, shared selective advantage, and genetic linkage. Owing to these
87 multiple influences, it remains challenging to identify those DS interactions which
88 represent putative mechanistic interactions, and those which result from chance co-
89 localisation.

90 To address these deficiencies in understanding, here we comprehensively
91 analyse interactions among DSs, and other AGEs, including plasmids, ICEs, ARGs,

92 phages, and the newly described anti-defence systems in a recently curated global
93 dataset of 2,940 *P. aeruginosa* genomes (17). *P. aeruginosa* is an ideal organism to study
94 DS interaction as previous studies have found it contains varied DSs (6, 9, 18). It is a
95 widely distributed, Gram-negative bacterium and is found in a diverse range of ecological
96 niches. As an opportunistic pathogen, it is a major concern in hospital settings, causing
97 a broad range of infections, particularly in individuals with cystic fibrosis (CF). Its
98 adaptability has contributed to its status as a World Health Organisation high-priority
99 AMR pathogen (19). Furthermore, DSs were recently revealed to play a key role in its
100 adaptation to different transmission models, making it a valuable model for studying
101 bacterial immunity (9, 17, 20). We identify a weak but significant effect of niche (CF vs.
102 non-CF) in driving accessory genome complements and then go on to characterise
103 ecological interactions among DSs and a wide range of AGEs, recovering multiple
104 putative phylogenetically independent interactions for downstream mechanistic
105 investigation. We also demonstrate that DSs and anti-defence systems are the principal
106 drivers of AGEs in *Pseudomonas* genomes providing valuable insights into microbial
107 ecology, evolution, and potential mechanistic interactions.

108 **Table 1. Overview of accessory genome elements (AGEs) considered in this study**

Category	Type	Examples	Functional role
Mobile element	Phage	Temperate and lytic phages- Dobby, fMGyn-Pae01	Mediate HGT, impose selection for DSs
	Plasmid	Conjugative and non-conjugative plasmids- IncQ2 type plasmids	Transfer genes, including ARGs and DSs
	Integrative conjugative elements (ICEs)	Tn4371-like elements, bph-sal	Mobilise and integrate genetic material between bacterial genomes
Functional element	Defence systems (DSs)	Restriction modification (RM), CBASS, AbiE, gabija	Protect against phages and MGEs

	Antimicrobial resistance genes (ARGs)	blaOXA, sul1, ermB	Confer resistance to antimicrobials
	Anti-defence systems	Anti-CRISPRs (Arcs) and Anti-CBASS	Inhibit defence systems

Methods

Genomic dataset

P. aeruginosa genomes used in this study were previously described (17). We utilised 2,940 (from a total of 9,829 genomes) genomes from studies of AMR in healthcare settings (21-24), individuals with cystic fibrosis (17, 25, 26), bacteraemia (27) and non-CF bronchiectasis (28), from studies into high-risk *P. aeruginosa* (29-31) and from the International *Pseudomonas* Consortium (32) (Supplementary Table 1). Isolate genomes from 9 deeply sampled CF patients from the TeleCF study were (17) down sampled, and only genomes representative of diversity were included in the analysis. Specifically, only patient representative strains were retained (i.e. the oldest sample for each sequence within a patient) to avoid overrepresentation of clones from deeply sampled patients. Genomes were assembled from short-read sequence data, using the Velvet (33) or SPAdes (34) assembler using the Assembly Improvement pipeline (35), as described by Weimann, et al. (2024). Bacterial genomes were annotated with PROKKA v1.14.6 (36).

Detection of accessory genome elements in bacterial genomes

DSs and anti-defence systems were detected in *P. aeruginosa* draft assembly FASTA files using DefenseFinder-DB v1.2.2 (accessed April 2023) and Anti-DefenseFinder-DB v1.3.0 (accessed December 2024) respectively (6). These databases contained 152 DSs and 41 anti-defence systems. Integrative conjugative elements (ICEs) and plasmids were detected in bacterial genomes using ICEberg v.2.0 (1032 database entries) (37) and PlasmidFinder v.2.1.6 (116 database entries) (38) databases via ABRicate v.1.0.1(39), using default parameters. Phage genomes present within the *P. aeruginosa* isolates were detected using geNomad v. 1.7.0 (40) and further characterised

135 with PHASTEST webserver (~410,000 viral sequences) (41). Additionally, ARGs were
136 identified using NCBI AMRFinderPlus v. v3.11.11, which contains 6,428 ARG entries (42).
137
138 Identifying co-occurrence
139 Coinfinder v.1.10 (43) was used for detecting significant interactions (i.e.
140 associations and dissociations) among AGEs in the *P. aeruginosa* dataset, using the
141 presence and absence data of defence systems, phage, plasmids, ICEs and ARGs, along
142 with a maximum likelihood tree phylogenetic tree generated by Weimann *et al*, as input
143 (Supplementary Table 1, see Figure 1C in Weimann, et al 2024) (17). This tree was
144 constructed using FastTree v2.1.10 (44) based on a core genome SNP alignment
145 obtained by mapping sequence reads to the *P. aeruginosa* PAO1 reference genome
146 (accession AE004091.2) (17). Gephi (45) was used to further interrogate interactions and
147 visualise the effect size (observed/expected occurrences) of significant interactions
148 (defined as Bonferroni-corrected binomial exact test $p<0.01$). In addition to effect size,
149 significant interactions were also triaged by the frequency (n= 400 for DS only
150 interactions and n= 300 for AGE interactions) of the AGE and their given D-value (D-value
151 >0), a measure of phylogenetic independence that quantifies the phylogenetic signal of
152 a binary trait. A lower D-value indicates a strong phylogenetic conservation (traits
153 following Brownian motion evolution), while a higher D-value suggests more random
154 distribution independent of phylogeny (46). To verify these findings with a more explicitly
155 phylogenetically controlled method, Goldfinder was also used to identify interacting DSs
156 and AGEs and detect co-occurrence and avoidance patterns between these elements
157 (47).
158
159 Statistical analyses
160 A Wilcoxon rank sum test was used to compare the frequency of AGEs in genomes
161 between niches. The association of DSs with CF and non-CF genomes was tested using
162 Fisher's exact test with Bonferroni-adjusted p-values. T-tests were used to evaluate the
163 differences in number of DSs, anti-defence systems and other AGEs per genome length
164 (kb) between genomes isolates between niches, and to assess differences in DS
165 frequency between isolates containing CRISPR-Cas Type I-F compared to those without.
166 Effect sizes for t-tests were calculated using Cohen's d, and for Wilcoxon rank sum tests

167 using rank-biserial correlation. To assess the relationship between DS abundance and
168 ARG content, we performed a simple linear regression analysis separately for both CF
169 and non-CF isolates. To compare the length of contiguous sequences (kb) containing
170 Pycsar and Shedu DSs (to those containing any other DS), we performed a Kruksall-
171 Wallis rank sum test followed by a Dunn's post hoc test for pairwise comparisons. All
172 analyses were performed using R v 4.4.0.

173 To assess the relative importance of interactions among different AGE features,
174 we calculated the proportion of significant interactions as a fraction of total theoretically
175 possible pairs (among detected elements in the AGE types). All significantly interacting
176 pairs were calculated based on the total number of each genomic element found (i.e.
177 ARG = 441, anti-defence system = 30, DS = 130, ICE = 19, phage = 29, plasmid = 13 see
178 Supplementary Table 1). To determine the total number of theoretically possible pairs
179 between different AGE databases (i.e. types of accessory genome feature e.g. ARGs,
180 DSs), we multiplied the size of the respective database (i.e. N_{AMR} x N_{DS}, where *N* is the
181 number of elements detected in the dataset). For pairs within the same category (e.g.,
182 AMR-AMR, we calculated the possible pairs as N_{AMR} x (N_{AMR}-1)/2 to adjust for duplicates.
183 The proportion of the observed number of significant interactions for each pair type was
184 then divided by the total theoretically possible pairs for that category to determine the
185 proportion of (possible) pairs that were significantly interacting.
186

187 **Results**

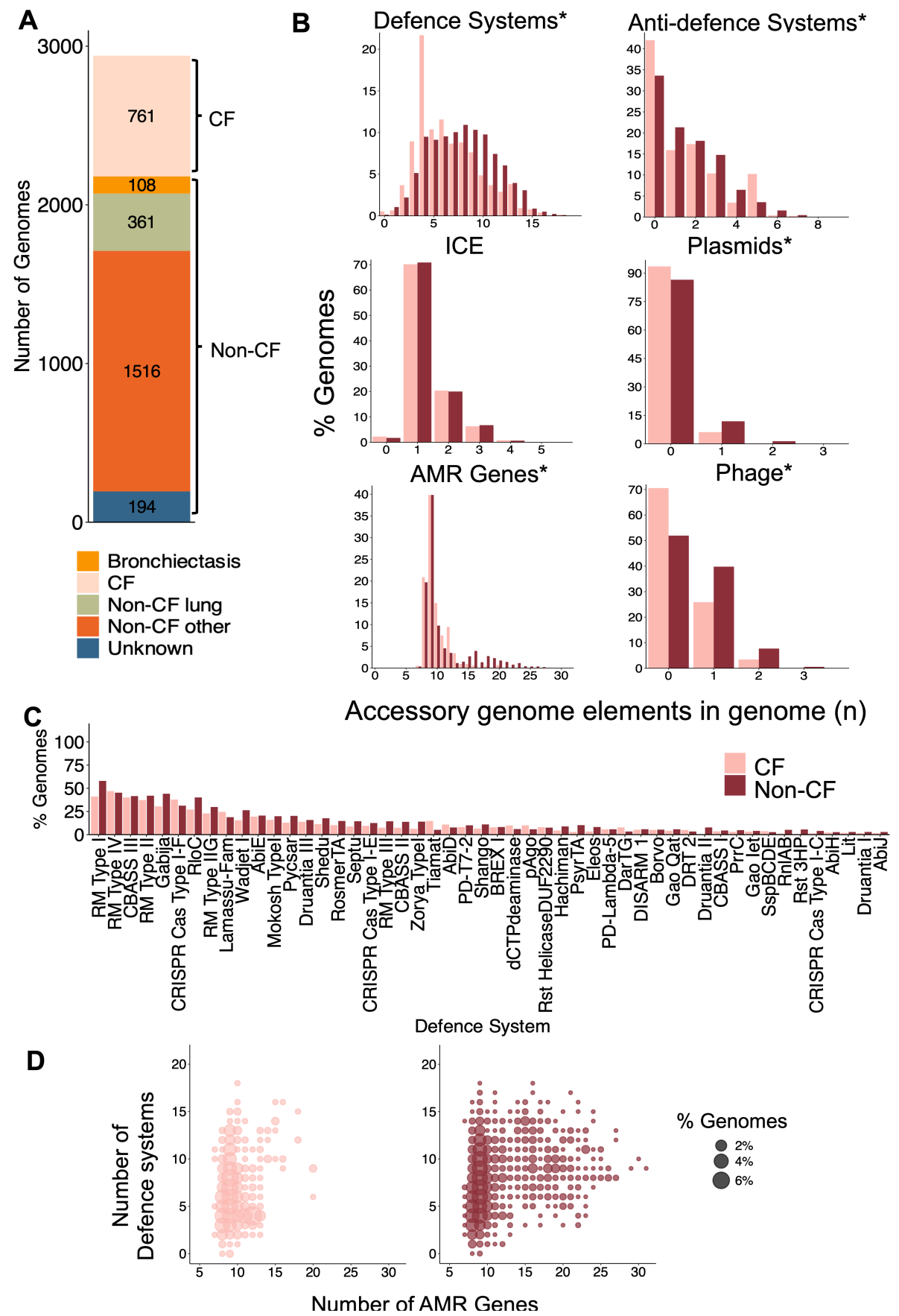
188 The accessory genome features of *P. aeruginosa* differ between niches

189 To explore the variation in the DSs of *P. aeruginosa* we analysed the DS repertoire
190 of 2,940 *P. aeruginosa* genomes which were isolated from several sources (Figure 1A)
191 and included globally distributed human, animal and environmental isolates. We
192 grouped isolates into two categories: CF and non-CF and compared their frequency of
193 AGE carriage (i.e. presence and absence). This division reflects two distinct ecological
194 niches of *P. aeruginosa*, where CF isolates typically originate from chronically infected
195 airways, with altered mucus composition, persistent infection, reduced microbial
196 competition, repeated, but likely purifying antibiotic exposure, and strong host-derived
197 selective pressures. In contrast, non-CF isolates are in this case associated with

198 acute/transient infection, environmental reservoirs and transmission chains where
199 exposure to more diverse competitors and MGEs is greater (48, 49).
200 This categorisation was further motivated by findings from the original study
201 (Weimann, et al (2024)) (17) which reported a significant depletion of genes involved in
202 bacterial defence in epidemic clones of *P. aeruginosa*. Therefore, we sought to examine
203 whether similar patterns were evident in the distribution across a representative dataset
204 of CF and non-CF isolates. Overall, *P. aeruginosa* contained an average of 7.5 DSs, with
205 99.8 % (n=2934/2940) of isolates containing at least 1 DS, consistent with previous
206 studies (6, 18) (Figure 1B, 1C, Supplementary Table 2). Numerous DSs (n=130) were
207 found, with RM Type I being the frequently found system, present in 53.6 %
208 (n=1577/2940) of genomes. This is consistent with published observations stating that
209 RM systems are present in >74% of prokaryotic genomes, with an average of ≥2 RM
210 systems per genome (4). This also revealed that *P. aeruginosa* were rich in AGEs across
211 both niches (i.e. DS (CF: 99.5% (n=757/761), non-CF: 99.9% (n=2176/2179)), anti-
212 defence system (CF: 58.0% (n=441/761), non-CF: 66.4% (n=1446/2179)), plasmid (CF:
213 6.4% (n=49/761), non-CF: 13.4% (n=293/2179)), ARG (CF: 100% (n=761/761), non-CF:
214 100% (n=2179/2179)), ICE (CF: 97.8% (n=744/761), non-CF: 98.3% (n=2141/2179)) and
215 phage (CF: 29.4% (n=224/761), non-CF: 48.0% (n=1047/2179)) (Figure 1A, 1B, 1C). We
216 found that CF genomes (relative to non-CF genomes) were depleted for DSs, anti-
217 defence systems, ARGs, plasmids, and phages (although these were weak effect sizes,
218 Table 2). However, we also found that CF genomes were significantly smaller than non-
219 CF genomes (Table 2) congruent with previous studies showing genomic streamlining in
220 bacteria adapting to chronically infected lungs (compared with more acute infections)
221 (17), and broader trends in host-adapting pathogens (50-55). However, our CF-
222 associated AGE depletions remained statistically significant when frequencies were
223 adjusted for genome size (t-test, $p<0.05$, Supplementary Figure 1).
224 To explore whether any individual DSs were enriched in either niche, we
225 compared the proportion of genomes containing each DS per niche, revealing that 17%
226 (n=22/130) of DSs were significantly associated with niche; 82% (n=18/22) with the non-
227 CF niche and the remainder with the CF niche (Bonferroni-corrected Fisher's Exact Test,
228 $p< 0.05$). The DSs associated most significantly with non-CF and CF niches were
229 Druantia II and Tiamat respectively (Figure 1C). RM Type IV was found in the most

genomes (n=357/761) within the CF niche, while RM Type I was the most common (n=1264/2179) in non-CF isolates.

And finally, to determine whether the ecological differences between CF and non-CF environments influence accessory genome composition, we compared the number of DS and ARGs across isolates from each niche. Consistent with the global comparison results (Table 2), there were a higher number of ARGs in non-CF isolates, though the distributions highlight a long tail of genomes with a high number of ARGs (Figure 1D). This is consistent with isolates in the non-CF environment retaining a higher number of AMR genes for survival in the harsher, more diverse environment and suggests that DS and ARG repertoires are shaped, in part, by niche-specific selection pressures. Despite this influence however, the overall effect sizes of the CF/non-CF niche was weak (Table 2), and as other studies have suggested that the distribution of AGEs is shaped by the ecological (i.e. cooperative vs antagonistic) relationships among DSs (and other AGE classes, Table 1) themselves. We then progressed to conduct a global analysis of the data to explore the potential interactions among the diverse AGEs in *P. aeruginosa* with a view to identifying putative associations and disassociations of DSs (and other AGEs) that might represent putative mechanistic relationships (while controlling for phylogenetic relationships).


A
Number of Genomes
3000
2000
1000
0
761
108
361
1516
194
CF
Non-CF
Bronchiectasis
CF
Non-CF lung
Non-CF other
Unknown
B
Defence Systems*
Anti-defence Systems*
ICE
Plasmids*
AMR Genes*
Phage*
% Genomes
C
% Genomes
Accessory genome elements in genome (n)
CF
Non-CF
RM Type I
RM Type IV
CBASS III
RM Type II
Gabija
CRISPR Cas Type I-F
RloC
RM Type IIG
Lamassu-Fam
Wadjet I
AbiE
Mokosh TypeI
Pycsar
Druantia III
Shedu
RosmerTA
Septu
CRISPR Cas Type I-E
RM Type III
CBASS II
Zorya TypeI
Tiamat
AbiD
PD-T7-2
Shango
BREX I
dCTPdeaminase
pAgo
Rst HelicaseDUF2290
Hachiman
PsyrTA
Eleos
PD-Lambda-5
DarTG
DISARM 1
Borvo
Gao Qat
DRT 2
Druantia II
CBASS I
PrrC
Gao Iet
SspBCDE
RnlAB
Rst 3HP
CRISPR Cas Type I-C
AbiH
Lit
Druantia I
AbiJ
Defence System
D
Number of Defence systems
Number of AMR Genes
% Genomes
2%
4%
6%

249

Figure 1. The distribution of accessory genome elements in *P. aeruginosa*. (A) Isolation niches for genomes in dataset (*n* = 2940. **(B)** The number of accessory genome elements (DSs, anti-DSs, ICEs, plasmids, ARGs and phages) per genome is shown as a percentage of genomes containing that number of element, stratified by CF/non-CF niche. Significant differences between the percentage of isolates containing any number of accessory genome elements from CF and non-CF niche's is denoted with a * ((Wilcoxon rank sum $p<0.05$). **(C)** The percentage of genomes containing the 50 most prevalent DSs by niche (see Supplementary Table 1 for full dataset). ICE = Integrated Conjugative Element ARGs = Antimicrobial Resistance Genes CF = cystic fibrosis DS = Defence System. **(D)** Bubble plots showing the relationship between the number of AMR genes and defence systems per genome, where each point represents each genomic profile and bubble size reflects percentage of genomes with that profile, stratified by CF and non-CF niches.

Table 2. Summary of differences between genomes from CF and non-CF patients

Feature	Niche	Mean	Range (Min-Max)	IQR (Q1-Q3)	*p*-value	Effect size
Genome Size (Mb)	CF non-CF	6.5 6.7	5.9-7.5 5.8-7.7	6.3-6.7 6.5-6.9	9.18×10^{55}	-0.64
DSs	CF non-CF	6.5 7.9	0-18 0-18	4-9 5-10	2.14×10^{25}	-0.25
Anti-defence Systems	CF non-CF	1.5 1.6	0-7 0-9	0-2 0-3	0.017	-0.05
ARGs	CF non-CF	9.7 11	7-20 7-31	9-10 9-12	5.12×10^{4}	-0.08
Plasmids	CF non-CF	0.07 0.2	0-2 0-3	0-0 0-0	1.65×10^{7}	-0.07
ICEs	CF non-CF	1.3 1.3	0-5 0-4	1-2 1-2	0.92	-0.001
Phages	CF non-CF	0.3 0.6	0-3 0-3	0-1 0-1	1.16×10^{19}	-0.19

† Statistical tests performed: *t-test* for differences in genome size (Cohen's *d* effect size reported), *Wilcoxon rank-sum test* for all other features (rank-biserial correlation effect size reported.

Determining defence system interactions

270 To identify putative synergistic or antagonistic DS interactions across *P.*
271 *aeruginosa*, we explored which DSs were more frequently associated (putative
272 synergism) or dissociated (putative antagonism) with each other than expected by
273 chance and independent of evolutionary relationships. For this we first, employed
274 CoinFinder (43), which identifies associations and dissociations between traits (in our
275 case, accessory genome categories of elements) and also estimates the phylogenetic
276 independence (as measured by D-value) of traits.

277 Association analysis with Coinfinder revealed 182 pairwise interactions among
278 DSs (27.9%, n=182 of 653 possible pairs) that were significantly associating or
279 dissociating with one another (Bonferroni-corrected binomial exact test $p<0.01$, Figure
280 2, Supplementary Table 3 & 4). Association interactions made up 85.2% (n=155/182) of
281 the total significant DS pairs and the remaining interactions were dissociations. Overall,
282 the dominance of associations over dissociations supports current hypotheses that DS
283 complement each other (14, 18, 56, 57). Each interacting DS (n=36) had a variable
284 number of associations and dissociations, with an average of 10 interactions (range 2-
285 19), and RM Type I having the most interactions (n=19). The high frequency of interactions
286 between RM Type I and other systems may reflect that RM shapes the DS complement of
287 genomes. Other DSs also exhibited high numbers of associations, such as RM Type IV,
288 Pycsar and AbiE, which all associated with 17 other DSs (Figure 1C, 2 & Supplementary
289 Table 3 & 4).

290 To focus our attention on those associations that may have significant roles in
291 shaping *P. aeruginosa* populations, we triaged DSs to focus on those that were
292 comparatively mobile and more common. As a proxy for mobility, we used the
293 phylogenetic independence (D-value) of each DS. As a lower D-value indicates a strong
294 phylogenetic conservation we considered DSs with a D-value of over 0 (Figure 2A) in our
295 downstream analysis. However, we note that the overall D-values across traits were
296 relatively low (ranging from 0.02-0.4), suggesting that while these DSs were not fully
297 phylogenetically conserved, some dependence on phylogeny likely remains. This
298 limitation is considered in our interpretation of the results. Additionally, we filtered the
299 interactions to only include DSs that were present in over 400 (13.6%, n= 400/2940 of the
300 total dataset, full data in Supplementary Tables 3 & 4) genomes (Figure 2A). Of the 15
301 DSs that passed these thresholds, 13 were found to both associate and dissociate with

302 other systems, with the remaining 2 having only associations (Figure 2A). Among these
303 15 DSs, RM Type II, Gabija Lamassu and AbiE were the most frequent and exhibited the
304 more phylogenetic independence (higher D-value within our range) (Figure 2A).

305 Having triaged on frequency and phylogenetic independence, we then sought to
306 interrogate DS interactions by visualising the effect size (inferred from the ratio of
307 observed vs expected occurrences, methods) of each association alongside the number
308 of interactions of each DS (Figure 2B & C, Supplementary Table 3 & 4). This revealed that
309 among the 54 (n= 54/155 total associations) interactions between the 15 triaged DSs,
310 Druantia III and Mokosh Type I association exhibited the strongest interaction (Figure 2B).
311 This indicates a possible mechanistic synergism and is consistent with experimental
312 evidence suggesting Druantia III and Mokosh Type I both protect against similar phages
313 such as T4, through nuclease-based mechanisms (7, 58, 59).

314 We then examined DS pairs that were found to dissociate from each other. As
315 these only made up only 14.8% (n=27/182) of the total significant interactions (Figure 2C
316 & Supplementary Table 4) we applied no triage thresholds (though notably those with a
317 D-value of <1 should be viewed with caution owing to the potential phylogenetic
318 association of individual DSs, see Figures 1A & C). Here we found that CRISPR-Cas Type
319 I-F had the most dissociations with other DSs, such as other RMs Type I & II and Pycsar
320 (Figure 2C & Supplementary Figure 2). Isolates containing CRISPR-Cas Type I-F
321 (n=969/2940) contained fewer DSs, compared to isolates not containing CRISPR-Cas
322 Type I-F (t-test, $p<0.05$) (Supplementary Figure 2). These strong indicators for
323 dissociation of CRISPR-Cas Type I-F are consistent with it being commonly found in *P.*
324 *aeruginosa* and blocking HGT of foreign DNA and even other DS (5, 60, 61). To better
325 control for phylogenetic dependence, we also ran the program Goldfinder to detect
326 interactions between DSs, complementing the Coinfinder analysis. This additional
327 approach was taken because the D-values for each significantly interreacting DS were
328 relatively low, indicating that many traits showed some phylogenetic association.
329 Encouragingly, Goldfinder identified 9 associations among 11 DSs (with D-values of
330 between 0.02 and 0.4, Supplementary Table 3 & 6), most of which (66.7%, n=6/9 pairs)
331 were also detected by Coinfinder and were within our filtering thresholds.

332 As recent studies have shown that DSs are frequently encoded within horizontally
333 transferrable defence islands or genomic hotspots (3, 9-11, 58) we sought to explore

334 whether our DS associations resulted from co-localisation. Therefore, we assessed
335 whether the significant associating and dissociating DS pairs (n= 77, Figure 2B & C) were
336 present on the same contiguous sequence (contig) in our draft genome assemblies (as
337 an imperfect but practicable proxy of co-localisation, Figure 3). This revealed that of all
338 DS containing contigs, only 29% (n=7086/24211) carried two or more DSs. Within the
339 7086 contigs containing ≥2 DS, associating DS pairs were co-located in only 22% of cases
340 i.e. were on the same contig 4,367 times out of a total of 19,472 pairs occurring across
341 genomes (Figure 3). Notably, the six intersecting associations detected by Goldfinder
342 had some of the highest (>0.5 proportion) rates of co-localisation (Figure 3,
343 Supplementary Table 6). As a counterfactual for understanding the rates of co-
344 localisation among associating DS pairs, we examined co-localisation rates among
345 dissociating DS pairs (Figure 3). As anticipated, the rates of co-localisation were far lower
346 i.e. 0.5%, n=175/37,098 pairs across all genomes, though this varied among interacting
347 pairs (Figure 3). We found the rate of co-localisation for 22 (of 54) associating pairs was
348 lower than 0.034% (with 0.034% being the highest co-localisation proportion for a
349 dissociating pairs) (Figure 3). We observed that some associating DSs were frequently
350 below this 0.034% threshold (e.g. Pycsar) potentially indicating genuine ecological
351 interactions not from co-localisation, and we further verified that this was not
352 attributable to an overall shorter length of contigs containing those DSs (Figure 3 &
353 Supplementary Figure 3). Overall, our results indicate that many associations detected
354 here are unlikely to have entirely resulted from co-localisation.

355 Hence, our analyses of DS interactions in *P. aeruginosa* revealed a complex
356 network of both association and dissociation interactions. Most significant interactions
357 were associations, with DSs such as RM Type II, Gabija, Lamassu and AbiE frequently
358 associating with other systems. Strong interactions between DSs may reflect
359 complementary functions. Conversely, frequent dissociations with CRISPR-Cas Type I-F
360 indicates its role in restricting HGT and conferring a selective disadvantage to co-
361 occurring DSs. Although co-localisation of DSs on contigs was generally low, some
362 associations such as Mokosh Type I and RM Type IV were consistently (n=415/415
363 genomes containing the pair) co-located. While such co-localisation can suggest
364 potential functional synergy (57, 62), in this case, the less-phylogenetically dependent
365 co-occurrence patterns could also indicate that co-inheritance could be driven by

genetic linkage rather than necessarily reflecting synergistic interactions between the systems (Figure 2A). Environmental context may also contribute to these patterns, as DS content can be shaped by local ecological conditions or exposure to specific phage communities. in addition to interactions with plausible functional explanations, our analysis also identified numerous statistically significant DS interactions for which no biological link has yet been described (e.g. CBASS II-Wadjet I & Gabija-Tiamat). These novel patterns may represent unexplored functional compatibility or incompatibility between DSs and therefore warrant further experimental investigation. Overall, our findings highlight the dynamic interplay between DSs, influenced by both genomic context and environment, that may shape the adaptive potential of *P. aeruginosa* populations.

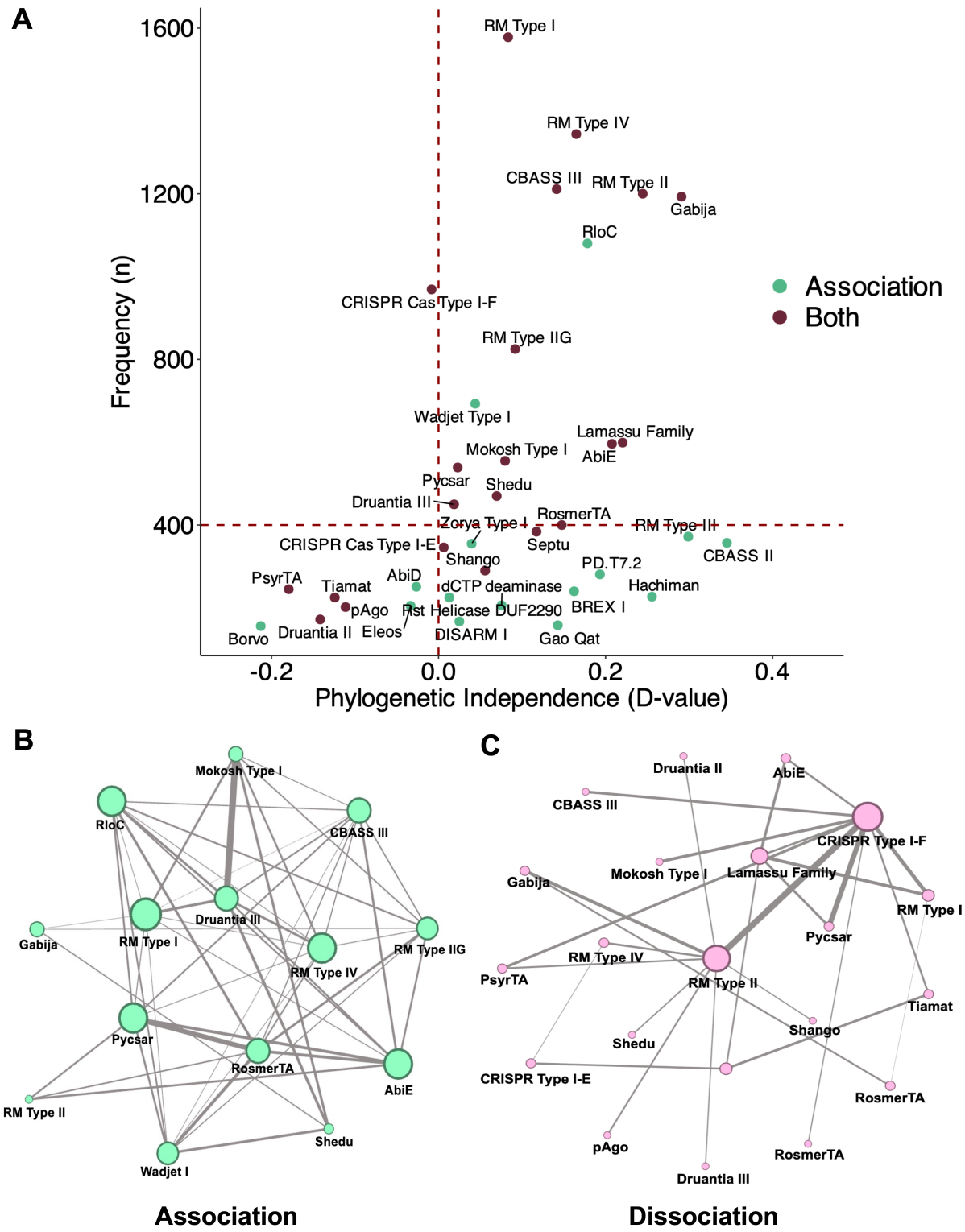
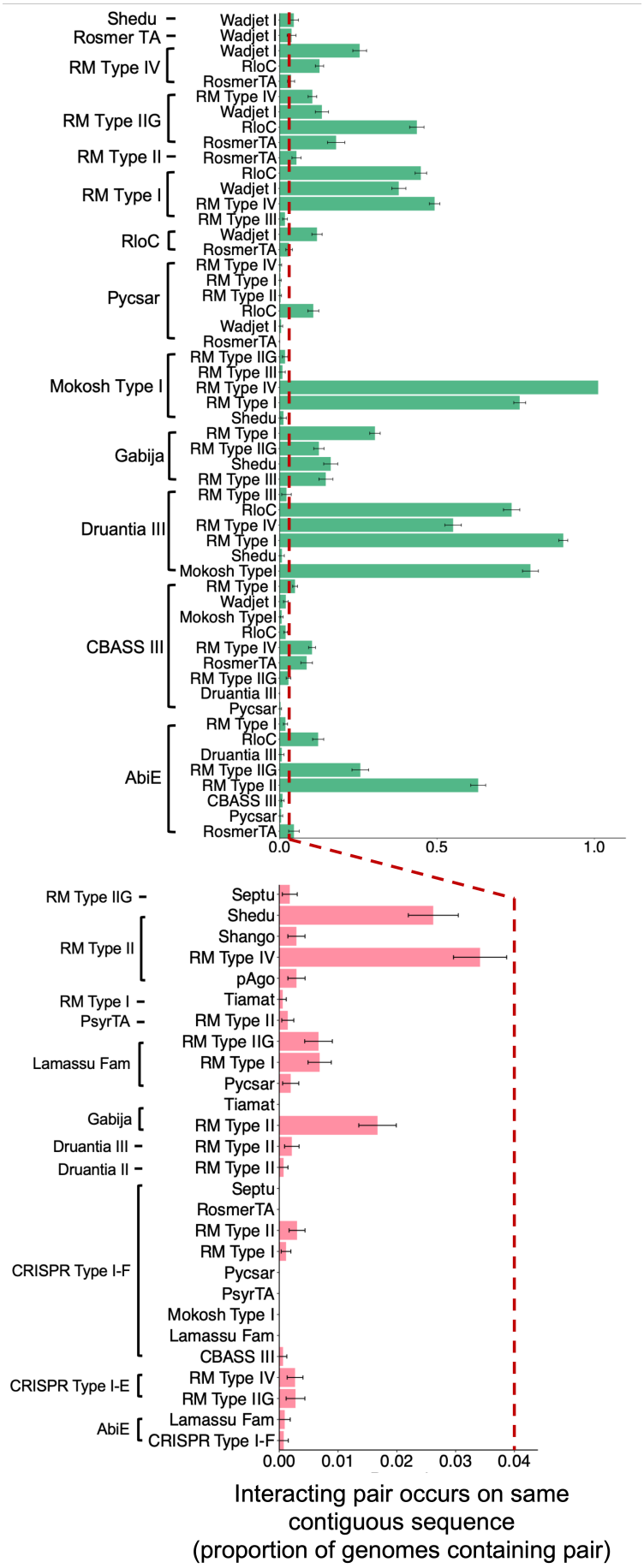


Figure 2. The distributions, associations and dissociations of defence systems in the global *P. aeruginosa* population. (A) The frequency and phylogenetic independence of DSs with significant interactions (i.e. associations or dissociations). Green dots highlight DSs that were found to have significant 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.


Shedu
Rosmer TA
RM Type IV
RM Type IIG
RM Type II
RM Type I
RloC
Pycsar
Mokosh Type I
Gabija
Druantia III
CBASS III
AbiE
Wadjet I
RloC
RosmerTA
RM Type IV
Wadjet I
RloC
RosmerTA
RosmerTA
RloC
Wadjet I
RM Type IV
RM Type III
Wadjet I
RosmerTA
RM Type IV
RM Type I
RM Type II
RloC
Wadjet I
RosmerTA
RM Type IIG
RM Type III
RM Type IV
RM Type I
Shedu
RM Type I
RM Type IIG
Shedu
RM Type III
RM Type III
RloC
RM Type IV
RM Type I
Shedu
Mokosh TypeI
RM Type I
Wadjet I
Mokosh TypeI
RloC
RM Type IV
RosmerTA
RM Type IIG
Druantia III
Pycsar
RM Type I
RloC
Druantia III
RM Type IIG
RM Type II
CBASS III
Pycsar
RosmerTA
0.0
0.5
1.0
RM Type IIG
RM Type II
RM Type I
PsyrTA
Lamassu Fam
Gabija
Druantia III
Druantia II
CRISPR Type I-F
CRISPR Type I-E
AbiE
Septu
Shedu
Shango
RM Type IV
pAgo
Tiamat
RM Type II
RM Type IIG
RM Type I
Pycsar
Tiamat
RM Type II
RM Type II
RM Type II
Septu
RosmerTA
RM Type II
RM Type I
Pycsar
PsyrTA
Mokosh Type I
Lamassu Fam
CBASS III
RM Type IV
RM Type IIG
Lamassu Fam
CRISPR Type I-F
0.00
0.01
0.02
0.03
0.04
Interacting pair occurs on same
contiguous sequence
(proportion of genomes containing pair)

394

395 **Figure 3. The proportion of pairs of DSs showing interactions of association (upper,**
396 **green) and dissociation (lower, pink) that occur on the same contiguous sequences**
397 **in draft assemblies.** Axes labels are ordered alphabetically. Error bars represent
398 standard error. The red dashed line intersects the x-axis at 0.04 on both plots to highlight
399 the difference in scale and show that some associating pairs were less frequently co-
400 located than dissociating pairs.

401

402 Interactions of DS with other accessory genome elements

403 To further explore the complexity of integral links among DSs and other AGEs,
404 including the recently described anti-defence systems, we used similar methods to
405 characterise the interactions among DSs and all other AGEs that maybe influencing
406 accessory genome composition. Coinfinder analysis revealed a total of 591 significant
407 interactions between specific AGEs distributed across accessory genome feature
408 categories (Figure 4 & Supplementary Table 3 & 5). As with DS comparisons, we triaged
409 the interactions by frequency (using those present in 7%, n=300/4288) of each element
410 across the dataset and phylogenetic independence (D-value > 0) (Figure 4A). Analysis of
411 the 36 (of 68) AGEs meeting these triage criteria revealed a network with 201 associations
412 of various effect sizes and differences in the number of associations (ranging between 2
413 and 31 per element) (Figure 4A & Supplementary Table 3 & 5). Of the 201 pairs passing
414 the threshold, anti-CRISPR AcrIF3 and anti-CRISPR AcrIF5 show the strongest
415 association (Figure 4A & Supplementary Table 5). Other comparatively strong
416 associations included those between ARGs *qacE*Δ and *sul1*, and DS AbiE and ICE bhp-
417 sal (Figure 4A & Supplementary Table 5).

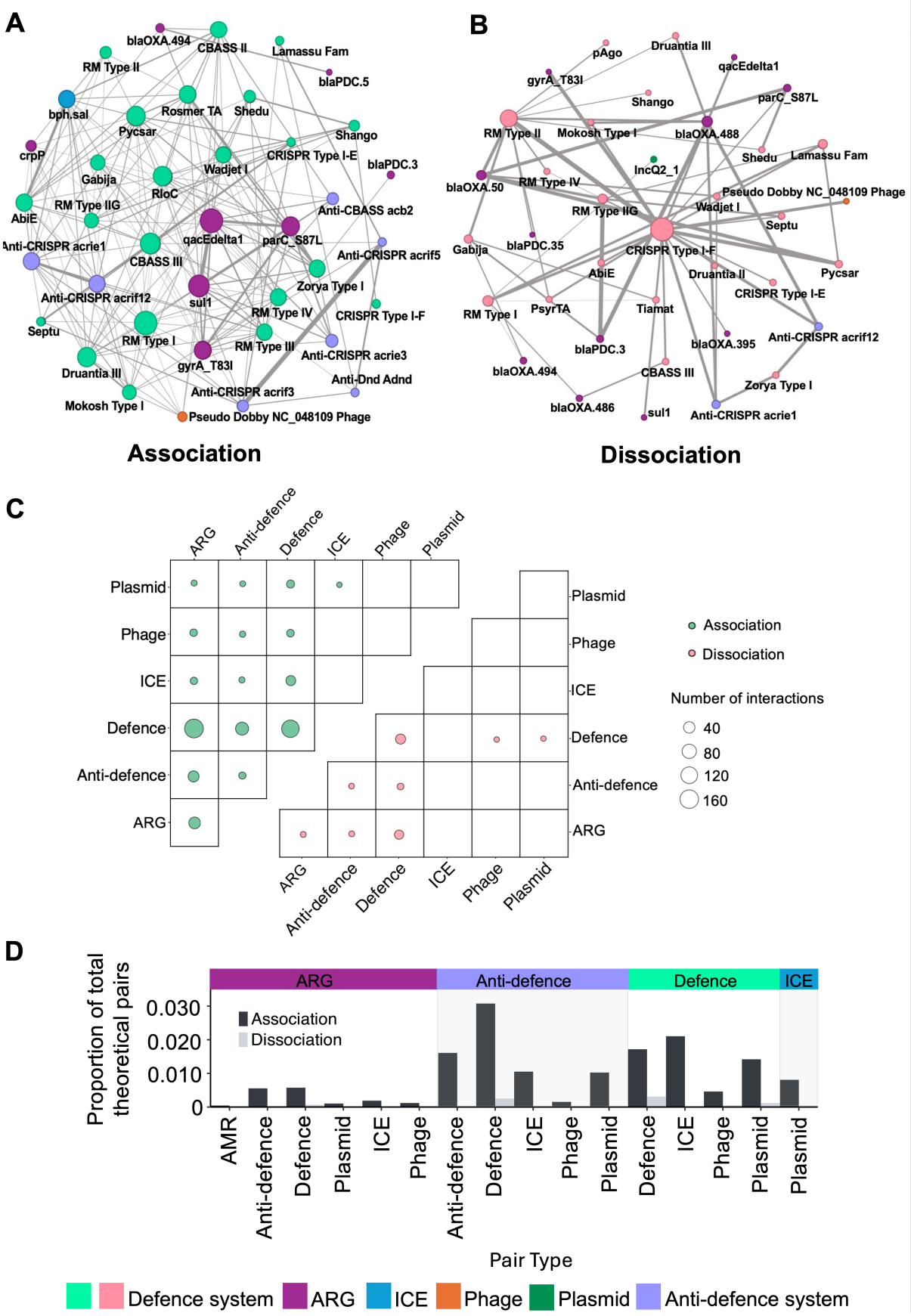
418 Regarding the dissociation interactions among these individual AGEs, CRISPR-
419 Cas Type I-F was again found to have the most negative interactions (29.8%, n=17/57
420 total dissociations with no triage criteria applied). The significant dissociation between
421 CRISPR-Cas Type I-F and RM Type II was also exhibited in this analysis, further supporting
422 the observation that they may not act synergistically in *P. aeruginosa* (Figure 4B &
423 Supplementary Table 5). The opposite is seen for the Gram-positive bacteria *Listeria*,
424 *Streptococcus* and *Staphylococcus*, where CRISPR-Cas Type IV and VI systems have
425 been observed to provide synergistic protection against phage infection with RM systems

426 suggesting synergy maybe system type specific and species specific (63-65). We also
427 observed that the ARG blaPDC3 had two strong dissociations with both RM Type IIG and
428 blaOXA-488 (Figure 4B & Supplementary Table 5). Notably, we also observed distinct
429 patterns of interaction between DSs and their cognate inhibitors, where anti-CRISPR
430 systems associate with CRISPR-Cas Type I-E but dissociate from CRISPR-Cas Type I-F
431 and anti-CBASS systems associate with CBASS systems.

432 Although we again applied a D-value threshold of >0 to reduce phylogenetic
433 dependence, the overall D-vales were relatively low (highest = 0.5), indicating that some
434 phylogenetic structure may remain in these data. To further evaluate these associations,
435 we re-analysed the dataset using Goldfinder, which more explicitly controls for
436 phylogeny. Goldfinder identified 14 of the 36 filtered AGEs as positively interacting with
437 each other and detected 19 significant associations among them (with D-values of
438 between 0.02 and 0.4, Supplementary Table 3 & 7). Of these pairs, 11 were also identified
439 by Coinfinder, indicating some concordance between the two methods. No
440 dissociations were found by Goldfinder, and notably, in line with our AGE interaction
441 networks (i.e. Figures 4A & B) being dominated by DSs and anti-defence systems,
442 GoldFinder only found interactions among DSs and anti-defences. The recovery of
443 mainly DS and anti-defence in our interactions supports this strategy as having detected
444 putatively sub-genomic ecological interactions among AGEs and prompted us to further
445 explore the broader trends among AGE groups in terms of quantitating interactions
446 among AGEs.

447 To characterise the broader dynamics and relative importance of interactions
448 among AGEs, we collapsed our AGEs into feature classes (i.e. ARGs, DS, phages, etc)
449 according to the respective feature databases and correlated this to the detected
450 association and dissociation interactions (with associations again predominating
451 [90.4% n=534/591 interactions]). This revealed the highest numbers of associations
452 between ARGs and DSs (31.1%, n=166/534 interacting pairs) followed by DSs and DSs
453 (27% n=144/534) and anti-defence systems and DSs (n=11% 60/534, Figure 4C).
454 Conversely, ARGs and DSs (35.1% n= 20/57) along with DSs and DSs (45.6% n=26/57)
455 had the most dissociations (Figure 4C). However, these results do not account for
456 discovery bias that may have arisen from differences in detection of AGEs attributable to
457 the respective size of the databases. Hence, to use our interaction findings to bring

458 meaningful insights on the relative importance of AGE interactions, we calculated how
459 many interactions arose between AGE classes as a proportion of the total number of
460 theoretical pairs that could occur between each pair of AGE classes (e.g. for ARG to DS
461 interactions, there were a total of 166 significant associations among a theoretical total
462 of 28,665 possible interactions, enumerated from the detection of 441 unique ARGs and
463 130 unique DS, giving a proportion of 0.006 significant pairs (Methods, Figure 4D).
464 Overall, this revealed that DS association interactions with themselves and with anti-
465 defence systems and ICEs were comparatively overrepresented (Figure 4D), and that
466 self-interactions between anti-defence systems were also high. Dissociation
467 interactions, although lower in abundance, were more frequently detected between DS-
468 DS, anti-defence-DS, and plasmid-DS pairs (Figure 4D). Hence, our results indicate that
469 interactions between DSs and anti-defence systems are common within the accessory
470 genome and may influence bacterial genome composition. While such interactions are
471 consistent with an ongoing interaction among DSs and anti-defence systems, they do not
472 necessarily imply direct functional synergy, and these patterns will be further influenced
473 by other ecological factors. Instead, they highlight the complex interplay of factors
474 shaping genetic diversity, AMR and bacterial evolution (3, 66, 67).


A
blaOXA.494
CBASS II
Lamassu Fam
RM Type II
blaPDC.5
bph.sal
Rosmer TA
Shedu
Pycsar
Shango
crpP
CRISPR Type I-E
blaPDC.3
Wadjet I
Gabija
RloC
RM Type IIG
Anti-CBASS acb2
AbiE
qacEdelta1
parC_S87L
Anti-CRISPR acrie1
CBASS III
Anti-CRISPR acrif5
Zorya Type I
Anti-CRISPR acrif12
sul1
RM Type IV
CRISPR Type I-F
Septu
RM Type I
RM Type III
Anti-CRISPR acrie3
Druantia III
gyrA_T83I
Anti-Dnd Adnd
Anti-CRISPR acrif3
Mokosh Type I
Pseudo Dobby NC_048109 Phage
Association
B
pAgo
Druantia III
qacEdelta1
gyrA_T83I
Shango
parC_S87L
RM Type II
Mokosh Type I
blaOXA.488
Shedu
Lamassu Fam
IncQ2_1
blaOXA.50
RM Type IV
Pseudo Dobby NC_048109 Phage
RM Type IIG
Wadjet I
Septu
Gabija
blaPDC.35
CRISPR Type I-F
AbiE
Druantia II
Pycsar
CRISPR Type I-E
RM Type I
PsyrTA
Tiamat
blaPDC.3
blaOXA.395
Anti-CRISPR acrif12
blaOXA.494
CBASS III
Zorya Type I
blaOXA.486
sul1
Anti-CRISPR acrie1
Dissociation
C
ARG
Anti-defence
Defence
ICE
Phage
Plasmid
Association
Dissociation
Number of interactions
40
80
120
160
D
Proportion of total theoretical pairs
0.030
0.020
0.010
0
Association
Dissociation
ARG
Anti-defence
Defence
ICE
AMR
Pair Type
Defence system
ARG
ICE
Phage
Plasmid
Anti-defence system

475

476 **Figure 4. Interactions among accessory genome elements in *P. aeruginosa*. (A)**
477 Significant associations network among individual accessory genome elements. Each
478 node depicts a genomic element coloured by type according to the inlaid key. Nodes are
479 scaled by the number of significant associations the element has. Edges represent
480 significant associations between elements with thickness scaled by effect size.
481 Elements have been triaged using a Phylogenetic independence value of ≥0 and being
482 found in >300 genomes. **(B)** Significant dissociations network among individual
483 accessory genome elements. Each node depicts a genomic element coloured by type
484 according to the inlaid key. Nodes are scaled by the number of significant dissociations
485 the element has. Edges represent significant dissociations between elements with
486 thickness scaled by effect size. **(C)** Bubble plots representing significant associations
487 (upper) or dissociations (lower) interactions between pairs of accessory genome feature
488 categories (axes). Bubbles are scaled by the number of significant interactions between
489 those categories. **(D)** The relative importance of AGE interactions is shown as bar plots
490 showing the proportion of feature category pair interactions that were found to be
491 significant associations (dark grey) or dissociations (light grey) with the denominator of
492 the total number of theoretical pairs (to account for database size). Feature category
493 pairs are indicated by groups above the bar plots (coloured according to the inlaid key)
494 and x-axis labels.

495 **Discussion**

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

499 The comprehensive analysis of 2,940 *P. aeruginosa* genomes revealed 82% of
500 detected DSs were significantly associated with bacterial niche, highlighting the role of
501 environmental pressures in shaping genome defence strategies. Notably, the CF niche
502 isolates were enriched for DarTG and CRISPR-Cas Type I-F, while non-CF niche isolates
503 exhibited a higher prevalence of Druantia II and RM-Type I systems. These findings
504 suggest that CF-adapted *P. aeruginosa* may rely on distinct defence mechanisms,
505 potentially reflecting selection for systems that counteract host-adapted phages or
506 environmental stress (respectively). Different bacterial DSs may be favoured under

507 distinct ecological conditions, yet identifying the environmental drivers of their
508 distribution in complex environments remains challenging. Although, there are studies
509 that have previously linked some ecological factors such as force of infection to the
510 evolution of immune strategies, where bacteria in lower infection risk environments can
511 favour CRISPR immunity compared to surface modification immunity (68, 69). This is
512 supported by increased abundance of CRISPR systems in thermophiles that reside in low
513 infection risk environments (70, 71), which may partly explain the enrichment of CRISPR-
514 Cas Type I-F in *P. aeruginosa* isolates from the CF niche in this study. Furthermore, pan-
515 genome differences between different *P. aeruginosa* phylogroups suggest that variation
516 in DSs could influence homologous recombination dynamics. However, these
517 associations may also reflect ecological or phylogenetic factors rather than direct
518 functional effects. While more diverse or abundant systems might limit the uptake of
519 exogenous DNA, this remains a hypothesis requiring further validation (72). Additionally,
520 CF isolates exhibited significant genome reduction and contained fewer DSs, anti -
521 defence systems, plasmids, ARGs and phages per kilobase of genome length, supporting
522 the hypothesis that long-term adaptation to the host-lung environment drives genomic
523 streamlining, through loss of genes linked to motility, virulence and environmental
524 adaptability due to focused selection pressures in chronic infection. However, as these
525 were overall weak effects, we sought to determine whether ecological relationships
526 among the sub-compartments of the vast accessory genome could be driving genome
527 composition.

528 We found a high density of DSs in *P. aeruginosa* genomes, consistent with
529 previous studies, and extended this work by illuminating the many significant
530 interactions among DSs. Associating interactions were far more common than
531 dissociating ones, supporting the hypotheses that DSs function in a complementary
532 manner, working alongside each other to provide a multi-layered approach against
533 invading AGEs, providing protection if other defence mechanisms fail (6, 14, 18, 56-58,
534 73, 74). Distinctly however, we reveal that associations are unlikely due to co-
535 localisations in defence islands.

536 Conversely, CRISPR-Cas Type I-F had the highest number of dissociations,
537 particularly with RM systems and Pycsar. Given its well documented role in blocking
538 HGT, including the acquisition of other DSs (60, 61, 75), this may explain why genomes

539 containing CRISPR-Cas Type I-F harboured significantly fewer DSs than those without it
540 (Supplementary Figure 2). Together, these findings highlight the complex interactions
541 shaping DS organisation and suggests that while many DSs complement one another,
542 CRISPR-Cas Type I-F may act as a gatekeeper, limiting genomic plasticity in *P.*
543 *aeruginosa* living in potentially nutrient deficient, highly competitive or stressful
544 environments. Hence, the interactions we show among defence systems here represent
545 plausible meaningful mechanistic or ecological relationships beyond simple co-
546 inheritance on defence islands. While these patterns are consistent with potential
547 interactions, they do not necessarily impact direct functional synergy and could also
548 arise from shared environmental or phylogenetic factors (7, 58, 76).

549 Overall, our study of the accessory genome interactions in *P. aeruginosa* revealed
550 a complex network of interactions among AGE classes that will be partly shaped by
551 ecological interactions. However, we also found relationships among AGEs that were
552 consistent with ecological interactions among AGEs (e.g. direct sub-genomic conflicts)
553 shaping the distribution of AGEs. For example, the most frequent interactions were
554 between ARGs and DSs, consistent with anti-plasmid activity, followed by DS-DS
555 interactions and DS-anti-defence system interactions, consistent with putative
556 mechanistic interactions among these AGEs and suggesting that DSs play a central role
557 in structuring the *P. aeruginosa* accessory genome. Many defence genes are found in the
558 accessory genome and are a key contributor to intraspecific variation of bacteria (77, 78).
559 However, their high turnover rate, co-acquisition with AGEs and persistence in the
560 accessory genome all affect the amount of protection that they offer against HGT (69, 79,
561 80), as will shifting environments, as highlighted above.

562 At a finer scale, our data revealed strong associations between specific AGEs,
563 such as anti-CRISPR AcrIF3 and AcrIF5, as well as the ARG pair *qacE*Δ and *sul1* and the
564 DS AbiE with ICE bhp-sal. These findings reinforce the notion that AGEs interact to
565 promote dissemination and survival in the environment, where AcrIF5 works closely with
566 AcrIF3 to block Cas 2/3 recruitment to ensure that CRISPR-Cas Type I-F cannot degrade
567 foreign DNA (73). The spread of ARGs is facilitated somewhat by DSs, where systems
568 such as Abi, Gabija, CBASS and RM have been found to be more frequently located within
569 or next to AGEs like ICEs, to affect AGE propagation, stability and HGT to confer
570 resistance in bacteria (3, 58, 79, 81, 82). Our study reveals key associations and

571 dissociations that appear largely independent of phylogeny (though this should be
572 explicitly considered alongside the results here e.g. D-values, the respective findings
573 between Coinfinder and Goldfinder). These patterns may be further shaped by multiple
574 factors including ecological context, and differential selective pressures.

575 Which brings us to the limitations of this study. Firstly, although our dataset
576 includes isolates from diverse ecological niches, many isolates originate from studies on
577 AMR, which may bias genome composition. Our findings are also reliant on the content
578 and curation of various AGE databases, so may have failed to detect significant
579 interactions with yet undescribed AGEs. Furthermore, our association interactions
580 appeared to be robust against incidental co-localisation but ascertaining the full impact
581 of co-localisation would be better evaluated with complete genomes assembled from
582 long-read sequencing data (e.g. as in Wu et al., 2024). Another limitation is that we relied
583 solely on DefenseFinder (6) to identify DSs, which may not fully capture the diversity of
584 all DSs. As a result, we may have missed relevant interactions due to incomplete or
585 imperfect annotations. A further consideration is that despite trying to account for
586 phylogenetic independence using the Coinfinder D-value output, the overall values were
587 relatively low, indicating that some phylogenetic dependence likely remains. To address
588 this, we re-analysed our data using Goldfinder (47), which accounts for shared ancestry
589 by mapping trait distributions onto the phylogeny and identifying interactions that are
590 independent of evolutionary relatedness. Despite identifying fewer total interactions,
591 Goldfinder recovered many of the same associations between DSs and AGEs as
592 Coinfinder, supporting the robustness of our main findings although these were often
593 highly co-located, so each interacting pair should be interpreted alongside markers of
594 co-localisation and phylogenetic independence. Our *in-silico* investigation into putative
595 synergism or antagonism among AGEs in *P. aeruginosa* does not offer explanations of
596 the mechanistic or ecological basis for those interactions. However, the approach is
597 robust to many confounding factors (e.g. phylogenetic dependence and the impact of
598 genome size and co-localisation), and many had a plausible mechanistic basis, so we
599 put forward these interactions as strong candidates for future research on biologically
600 meaningful interactions among the accessory genome milieu in *P. aeruginosa*.

601 The results of this work highlight the importance and complexity of accessory
602 genome interactions in bacteria, particularly the expanding array of DSs and anti-

defence systems in shaping accessory genome interactions. Multiple association and dissociation interactions between DSs, and other accessory genome features were found and suggest adaptation of *P. aeruginosa* is driven by the selective forces in their environment. Future research should aim to explore and validate these putative interactions both phenotypically and mechanistically, and to continue to explore other possible drivers of this apparent synergy and antagonism.

Supplementary Material

Supplementary Table 1. *Pseudomonas aeruginosa* isolates, associated metadata and accessory genome feature content used in this study.

Supplementary Table 2. DefenseFinder and Anti-defenseFinder output from *P. aeruginosa* dataset.

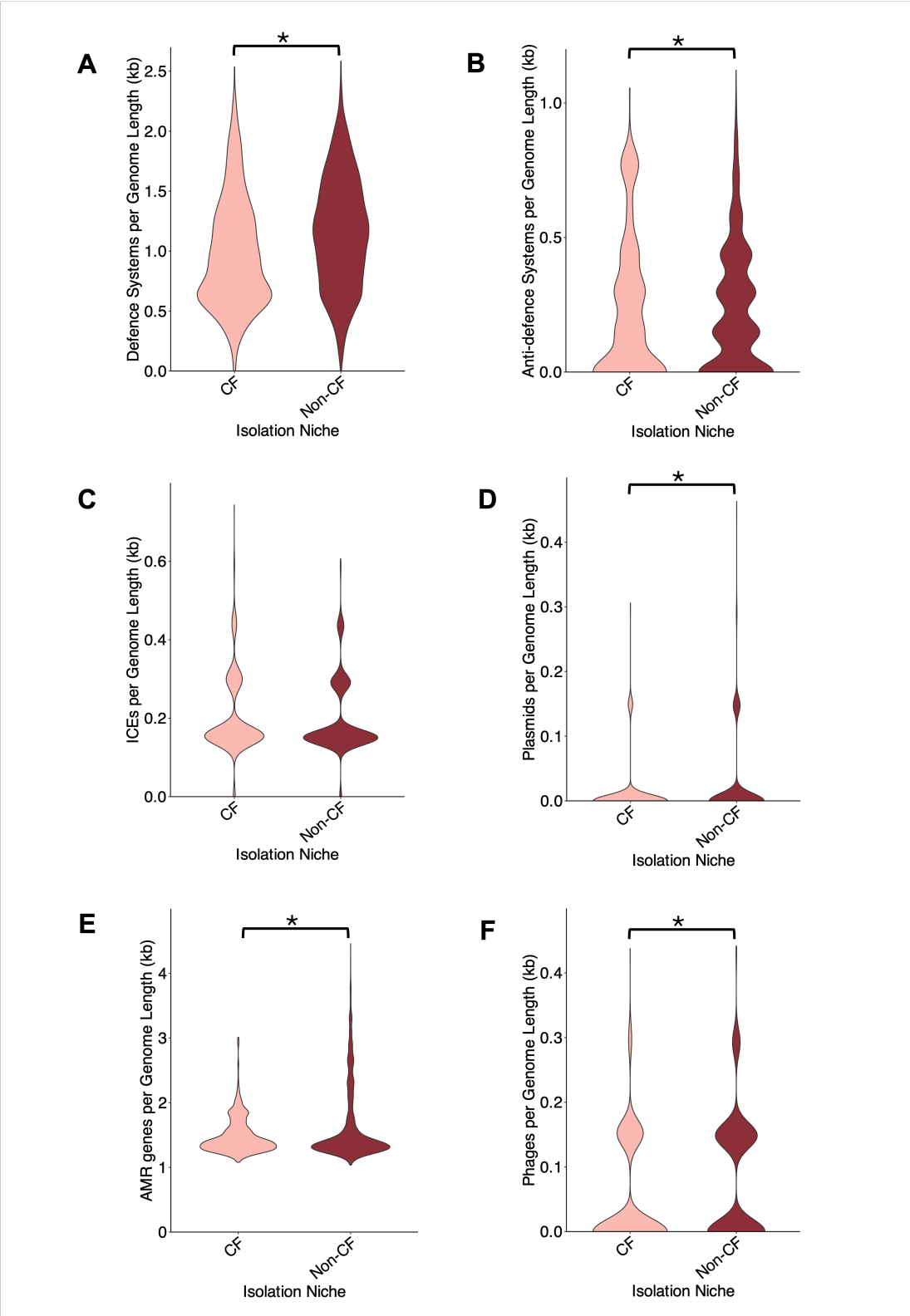
Supplementary Table 3. Accessory genome elements found to significantly interact by Coinfinder, their associated D-value, frequencies and interactions.

Supplementary Table 4. List of DSs pairs found to associate and dissociate by Coinfinder.

Supplementary Table 5. List of accessory genome feature pairs found to associate and dissociate by Coinfinder.

Supplementary Table 6. List of DS pairs found to associate and dissociate by Goldfinder.

Supplementary Table 7. List of accessory genome feature pairs found to associate and dissociate by Goldfinder**.**

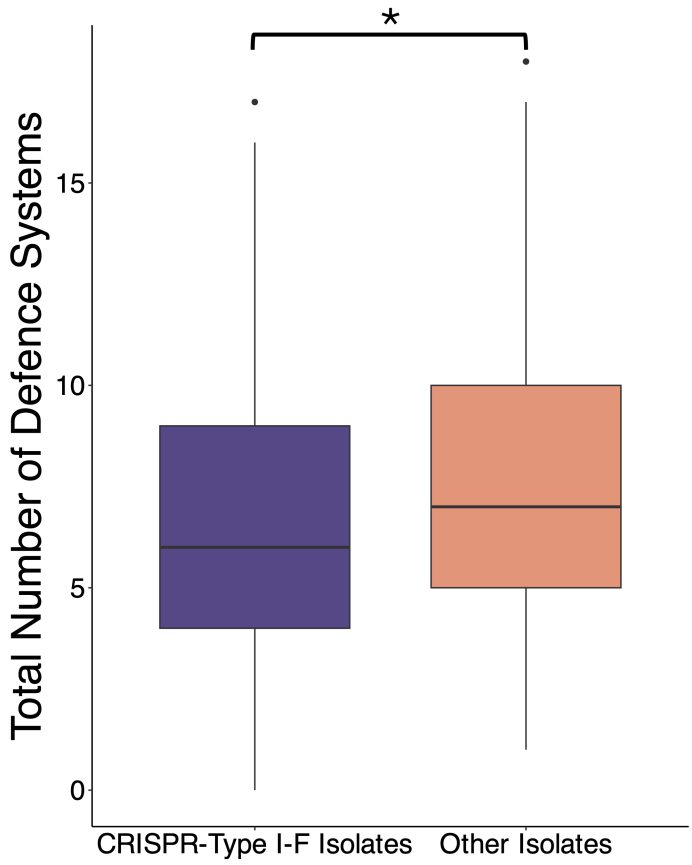


625

626 **Supplementary Figure 1**. Number of defence systems **(A)**, anti-defence systems **(B)**,

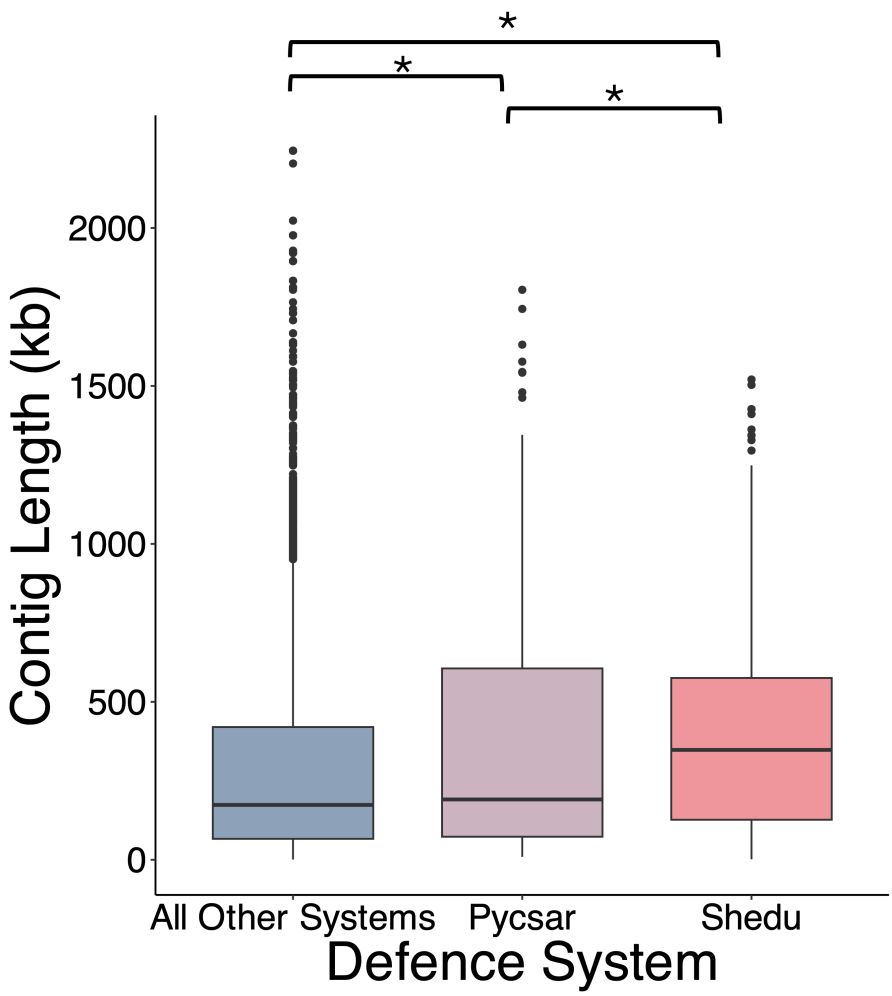
627 ICEs **(C)**, plasmids **(D)**, ARGs **(E)** and phage **(F)** present in *P. aeruginosa* per genome

628 length (kb) of genomes isolated from CF patients and non-CF sources. Significant
629 differences in number of accessory genome elements per genome length (kb) between
630 genomes isolates from CF and non-CF niches is denoted with a * (*t*-test $p<0.05$). ICE =
631 Integrated Conjugative Element ARGs = Antimicrobial Resistance Genes CF = cystic
632 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, $p<0.05$).
633

634 **Supplementary Figure 2**. Comparison of the number of DSs found in *P. aeruginosa*
635 isolates containing the DS CRISPR-Type I-F (n=1591) and isolates that did not (n=2683).
636 The plot shows the median (line), interquartile ranges (box), and range (whiskers) of the
637 number for each group. Outliers are shown as individual markers. The asterisk
638 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 393 kb and 349 kb respectively) were slightly longer than contigs containing other DSs (average of 275 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.

640 **Supplementary Figure 3**. Comparison of length in kilobases of contiguous sequences
641 containing either Pycsar (n=539), Shedu (n=586) or all other DS (n=23471). The plot
642 shows the median (line), interquartile ranges (box), and range (whiskers) of the number
643 for each group. Outliers are shown as individual markers. The asterisk indicates a
644 statistically significant difference ($p<0.05$) as measured by a Kruksall-Wallis rank test
645 followed by a Dunn's post hoc test for pairwise comparisons.

646 **Funding Information**

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

650 **Author contributions**

651 C.E.C and K.S.B contributed to the writing of the original draft. All authors (C.E.C,
652 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
653 and editing of subsequent drafts. C.E.C was responsible for formal analysis,
654 investigation, visualisation and methodology. A.W, J.P, R.A.F, and C.E.C. contributed to
655 data curation. K.S.B, M.A.B, M.D.S, C.E.C, and E.R.W contributed to funding acquisition
656 and conceptualisation. K.S.B and E.R.W were responsible for project administration.
657 K.S.B was responsible for supervision.

658 **Acknowledgements**

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

661 **Competing Interests Statement**

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

663 **Data availability**

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

666

667 **References**

668 1. Vos M, Buckling A, Kuijper B, Eyre-Walker A, Bontemps C, Leblond P, et al. Why
669 do mobile genetic elements transfer DNA of their hosts? Trends in Genetics.
670 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 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.

762 33. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using
763 de Bruijn graphs. Genome research. 2008;18(5):821-9.
764 34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
765 SPAdes: a new genome assembly algorithm and its applications to single-cell
766 sequencing. Journal of computational biology. 2012;19(5):455-77.
767 35. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, et al. Robust high-
768 throughput prokaryote de novo assembly and improvement pipeline for Illumina data.
769 Microbial genomics. 2016;2(8):e000083.
770 36. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.
771 2014;30(14):2068-9.
772 37. Liu M, Li X, Xie Y, Bi D, Sun J, Li J, et al. ICEberg 2.0: an updated database of
773 bacterial integrative and conjugative elements. Nucleic Acids Research.
774 2018;47(D1):D660-D5.
775 38. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et
776 al. In silico detection and typing of plasmids using PlasmidFinder and plasmid
777 multilocus sequence typing. Antimicrob Agents Chemother. 2014;58(7):3895-903.
778 39. Seemann T. Abricate. Github.
779 40. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of
780 mobile genetic elements with geNomad. Nature Biotechnology. 2024;42(8):1303-12.
781 41. Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, et al. PHASTEST: faster
782 than PHASTER, better than PHAST. Nucleic Acids Res. 2023;51(W1):W443-w50.
783 42. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et
784 al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the
785 genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep.
786 2021;11(1):12728.
787 43. Whelan FJ, Rusilowicz M, McInerney JO. Coinfinder: detecting significant
788 associations and dissociations in pangenomes. Microb Genom. 2020;6(3).
789 44. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood
790 Trees for Large Alignments. PLOS ONE. 2010;5(3):e9490.
791 45. Bastian M, Heymann S, Jacomy M, editors. Gephi: an open source software for
792 exploring and manipulating networks. Proceedings of the international AAAI conference
793 on web and social media; 2009.
794 46. Fritz SA, Purvis A. Selectivity in Mammalian Extinction Risk and Threat Types: a
795 New Measure of Phylogenetic Signal Strength in Binary Traits. Conservation Biology.
796 2010;24(4):1042-51.
797 47. Gavriilidou A, Paulitz E, Resl C, Ziemert N, Kupczok A, Baumdicker F. Goldfinder:
798 Unraveling Networks of Gene Co-occurrence and Avoidance in Bacterial Pangenomes.
799 bioRxiv. 2024:2024.04.29.591652.
800 48. Camus L, Briaud P, Vandenesch F, Moreau K. How Bacterial Adaptation to Cystic
801 Fibrosis Environment Shapes Interactions Between Pseudomonas aeruginosa and
802 Staphylococcus aureus. Front Microbiol. 2021;12:617784.
803 49. Sousa AM, Pereira MO. Pseudomonas aeruginosa Diversification during
804 Infection Development in Cystic Fibrosis Lungs-A Review. Pathogens. 2014;3(3):680-
805 703.
806 50. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, Parsons BN, et al.
807 Patterns of genome evolution that have accompanied host adaptation in *Salmonella*.
808 Proceedings of the National Academy of Sciences. 2015;112(3):863-8.

51. 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.
52. Grote A, Earl AM. Within-host evolution of bacterial pathogens during persistent infection of humans. Current Opinion in Microbiology. 2022;70:102197.
53. 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.
54. 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.
55. 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.
56. 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.
57. 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.
58. 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.
59. 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.
60. Wheatley RM, MacLean RC. CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. The ISME Journal. 2021;15(5):1420-33.
61. 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).
62. Payne LJ, Hughes TCD, Fineran PC, Jackson SA. New antiviral defences are genetically embedded within prokaryotic immune systems. bioRxiv. 2024:2024.01.29.577857.
63. 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.
64. 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.
65. 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.
66. 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.
67. Murtazalieva K, Mu A, Petrovskaya A, Finn RD. The growing repertoire of phage anti-defence systems. Trends in Microbiology. 2024;32(12):1212-28.

68. 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.
69. van Houte S, Buckling A, Westra Edze R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. Microbiology and Molecular Biology Reviews. 2016;80(3):745-63.
70. 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.
71. 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.
72. 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.
73. 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.
74. 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.
75. 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.
76. 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.
77. Smith WPJ, Wucher BR, Nadell CD, Foster KR. Bacterial defences: mechanisms, evolution and antimicrobial resistance. Nature Reviews Microbiology. 2023;21(8):519-34.
78. Meaden S, Fineran PC. Bacterial defense islands limit viral attack. Science. 2021;374(6566):399-400.
79. 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.
80. 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.
81. 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.
82. 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.