Informing plasmid compatibility with bacterial hosts using protein-protein interaction data

Research Article

Tim Downing[1,2], Min Jie Lee[1], Conor Archbold[1], Adam McDonnell[1], Alexander Rahm[3].

[1] School of Biotechnology, Dublin City University, Dublin, Ireland.
[2] The Pirbright Institute, UK. Corresponding author. Email: Tim.Downing@pirbright.ac.uk Orcid: 0000-0002-8385-6730.
[3] GAATI Lab, University of French Polynesia, Tahiti, French Polynesia. Orcid: 0000-0002-5534-2716.

Highlights

- We tested for protein interactions between key plasmids and 4,363 bacterial hosts
- 23 host species had high rates of protein interactions with four *bla*OXA-48 plasmids
- *Pseudomonas* species showed high rates of interactions with these plasmids
- Host-plasmid protein-protein interaction rates may be associated with compatibility

Abstract

The compatibility of plasmids with new host cells is significant given their role in spreading antimicrobial resistance (AMR)[1] and virulence factor genes. Evaluating this using *in vitro* screening is laborious and can be informed by computational analyses of plasmid-host compatibility through rates of protein-protein interactions (PPIs) between plasmid and host cell proteins. We identified large excesses of such PPIs in eight important plasmids, including pOXA-48, using most known bacteria (n=4,363). 23 species had high rates of interactions with four *bla*OXA-48-positive plasmids. We also identified 48 species with high interaction rates with plasmids common in *Escherichia coli*. We found a strong association between one plasmid and the fimbrial adhesin operon *pil*, which could enhance host cell adhesion in aqueous environments. An excess rate of PPIs could be a sign of host-plasmid compatibility, which is important for AMR control given that plasmids like pOXA-48 move between species with ease.

Keywords: Fimbriae, plasmid, adhesin, protein-protein interaction, network.

[1] **Abbreviations**
Antimicrobial resistance (AMR)
Carbapenemase-producing Enterobacteriaceae (CPE)
Extraintestinal pathogenic E. coli (ExPEC)
Horizontal gene transfer (HGT)
Protein-protein interactions (PPIs)
Type IV pilus (T4P)

1. Introduction

Plasmids can provide new genes to host cells, enhancing properties like antimicrobial resistance (AMR) and adhesion to host molecules [1,2]. Plasmids are short extrachromosomal DNA elements that are usually circular, such that conjugative plasmids can move between different host cells sharing the same ecological niche [3,4]. DNA transformation, extracellular vesicles and bacteriophages also permit plasmid gene transfer, albeit less often than conjugation [5-8]. Collectively, these horizontal gene transfer (HGT) processes meaning that the role of plasmids in enabling adaptation to new environments is a key concern in addressing the spread of infectious bacteria. Additionally, plasmids typically induce a small fitness cost in their host cells [9,10]. To overcome this, many plasmids possess genes to parasitise host cells via post-segregation killing and stable inheritance: this ensures their propagation in daughter cells [11] and facilitates vertical plasmid transmission. Many plasmids also endow their host cells with enhanced AMR [12], thereby enhancing their fitness and subsequently the plasmid's survival. Consequently, clonal radiations of emerging pathogens are often associated with the HGT-mediated acquisition of plasmids encoding key AMR genes (like *bla*OXA-48) that enable survival in antibiotic-rich niches.

Plasmid-encoded AMR genes help drive the transmission of carbapenemase-producing *Enterobacteriaceae* (CPE) including *Escherichia coli* and *Klebsiella.* These genes are concerning because they can be transmitted to other bacteria. These two priority pathogens impair human health according to the World Health Organisation (WHO) [13] and were the most common CPEs in England in 2014-16 at 27% (*E. coli*) and 52% (*Klebsiella*) where the most common carbapenemase gene was *bla*OXA-48 [14]. More recent surveys of CPEs in England in 2020-22 confirmed this pattern, showing *E. coli* and *Klebsiella* were responsible for 26% and 33% of cases, respectively [15]. Such plasmid-driven AMR gain by extraintestinal pathogenic *E. coli* (ExPEC) has enhanced their capacity to infect epithelial cells beyond their normal gut niche [16-25].

Highly transmissible pOXA-48-like plasmids are a compelling example of challenges driven by mobile genetic elements [26]. The *bla*OXA-48 gene encodes an Ambler class D beta-lactamase first identified in *Klebsiella pneumoniae* from Turkey [27]. It is usually on a highly mobile 62 Kb conjugative plasmid named pOXA-48a [28] that has spread the *bla*OXA-48 gene globally, causing substantial outbreaks in many countries, including Ireland [29-32]. An estimated 87% of *bla*OXA-48-positive isolates are MDR, and about half of CPE isolated in Europe have this group II OXA carbapenemase gene [33], including *Serratia maracescens* (since 2021). This gene has been found regularly in community and hospital settings across diverse host species, most typically in the *Enterobacterales* order. Host-pOXA-48a co-evolution has occurred, for example through host cell wall permeability changes to alleviate an initial low level of *bla*OXA-48 carbapenemase activity [34].

Beyond pOXA-48, IncF plasmids' roles in spreading AMR genes are well known [12,35-36], including non-conjugative plasmids pEK499 and pEK516 in MDR *E. coli* [37-40]. These plasmids were first identified in England in *E. coli* and have ~75% similarity to each other [37]. Additionally, a distinct conjugative IncI1 *bla*CTX-M-3-*bla*TEM-1b-positive plasmid pEK204 was found in epidemiologically related hosts in Belfast [37-38]. Plasmid pEK204 encodes a type IV pilus (T4P) through its *pil* operon which encodes a thin pilus for epithelial cell adhesion or biofilm formation [41]. Plasmids encoding such fimbrial adhesins could allow bacterial cells to survive in new environments because adherence to human epithelial cells facilitates bacterial colonisation and the formation of biofilms [42]. These adhesins include type 1, P, F1C/S and AFA fimbriae/pili that are typically assembled by the chaperone-usher secretion pathway in gram-negative bacteria [43-47]. Certain pathogens retain multiple different fimbriae that can be expressed depending on the environmental conditions [48-49]. This indicates that possessing a repertoire of fimbriae could be adaptively beneficial. T4P are thin connections that can be elongated or retracted by ATPases [50] and are associated with biofilm formation [51], virulence [52] and bacteriophage infection [53]. The *pil* operon (particulary *pilS*) may allow differential epithelial cell or abiotic surface adhesion with a thin flexible pilus [41] that can elevate conjugation rates in liquid environments [54]. It can also permit biofilm formation such that a T4P-negative cell can adhere to

other T4P-expressing cells bound with a host cell [41]. Although the *pil* operon is somewhat rare, its widespread nature across bacteria implies a role in human infection [55].

These and other findings imply that a plasmid's capacity to interact successfully with a host cell is important and stems more from its genetic profile more so than the host genome [4]. This suggests host fitness increases following plasmid gain depend not only on the environment, but also on plasmid-host protein-protein interactions (PPIs) arising from their collective genetic profiles. These PPIs must be beneficial for plasmids to persist because plasmid-encoded genes must be replicated and expressed in the host cell. This presents the hypothesis that higher rates of plasmid-host PPIs are a signature of higher compatibility between the bacterial cell and the plasmid.

Limited work has examined plasmid-host PPIs across all bacteria. Previously, PPI network analysis has identified AMR genes in *E. coli* [55,56], but has not been applied at the level of whole plasmids. Here, we examined PPIs between plasmid and bacterial hosts to explore the potential compatibility between the host cell and plasmid. We focused on four pOXA-48-related isoforms and four *E. coli* plasmids (pEK204, pEK499, pEK516, pEC958). We assessed these features initially in 4,363 samples representative of most bacteria. We found that it was possible that elevated sample-plasmid PPI levels could be symptomatic of host-plasmid compatibility, and that this observation may stem from operons associated with environments where those bacterial samples have been found.

2. Methods

2.1 Protein-protein interaction data extraction for 24 *Enterobacterales* and all bacteria

We retrieved PPI information for two key datasets from STRING database v12 [57] using R packages Rentrez v1.2.2 [58], STRINGdb v2.0.1 [59], and StringR v1.4.0 [60] (Figure 1), in a working environment using Rv4.0.1 [61], RStudio v2022.2.3.492 [62], Bioconductor v3.11 [63]. The dataset first was for three genera from the *Enterobacterales* order containing: *Escherichia* and *Shigella* (n=11), *Klebsiella* (n=2) and *Serratia* (n=11) (Table S1). *Serratia* belongs to the *Yersiniaceae* family, whereas the others are all *Enterobacteriaceae*. Three other potential samples in StringDB belonging to these genera were excluded because they lacked any pOXA-48-related plasmid-sample PPIs and so had no information. The second dataset was composed of 4,363 StringDB bacterial samples (including the 24 *Enterobacterales* from above) with valid PPI data as outlined previously [79] (Figure 1). Using a score threshold of 400 as per previous work [79], the number of proteins and their PPIs were recorded. These StringDB scores come from a range of sources, including: text mining, pathway databases, co-expression research, small-scale and high-throughput experiments, and computational predictions (chromosomal proximity, phylogenetic context and gene fusion) [57]. The data was processed, collated and visualised using Dplyr v1.0.8 [64], Forcats v0.5.1 [65], Ggplot2 v3.3.5 [66], Ggrepel v0.9.1 [67], ReadR v2.1.2 [68], Readxl v1.4.0 [69], Tibble v3.1.6 [70], TidyR v1.2.0 [71], Tidyverse v1.3.0 [72] and VennDiagram v1.7.3 [73].

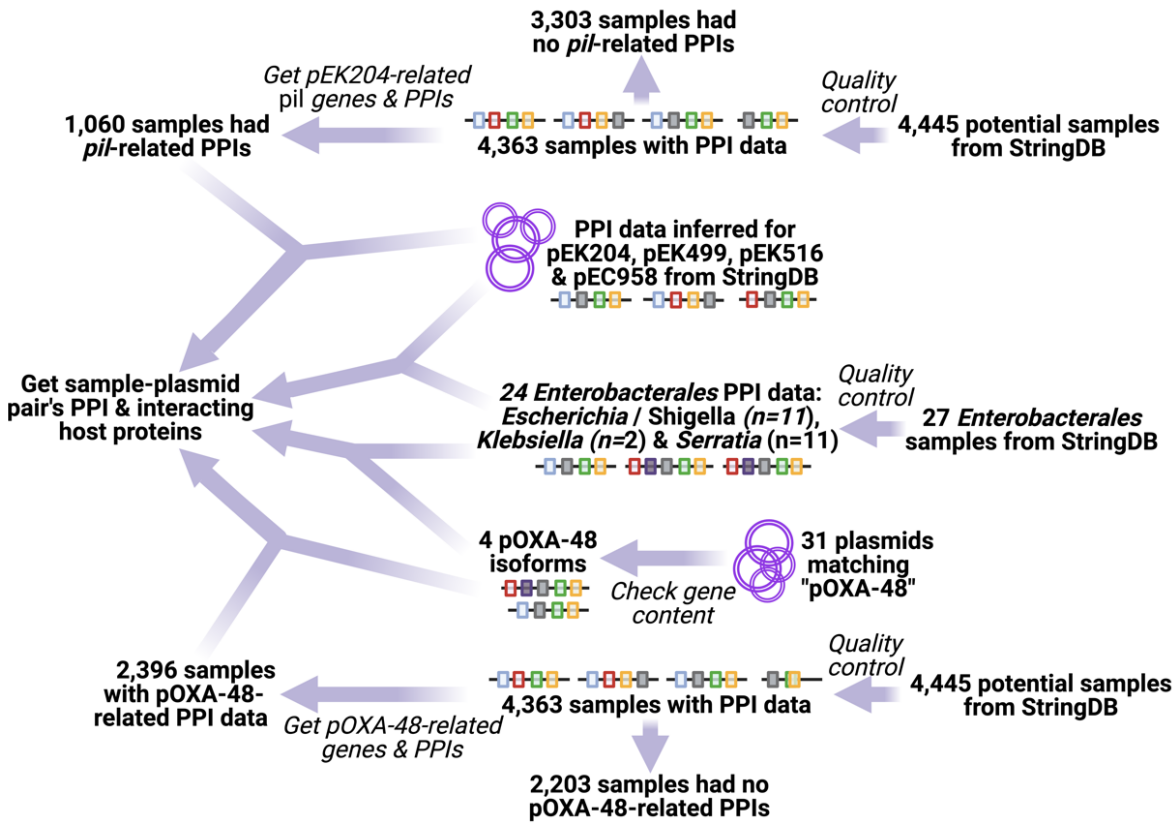


Fig 1. The data sources and methods summary, read from right to left. We used PPI data from StringDB related to two datasets: one with 24 *Enterobacterales* (middle) and another with 4,363 diverse bacterial samples (top and bottom). Top: We compared the interacting host proteins and associated PPIs for four plasmids pEK204, pEK499, pEK516 and pEC958 compared to the 24 *Enterobacterales* samples' PPI data to assess their PPIs. These plasmids were also compared to the 4,363 samples' PPI data. Bottom: We examined 31 initial pOXA-48-related plasmids that resolved as four main isoforms whose interacting host proteins and associated PPIs we compared to the 24 *Enterobacterales'* PPI and the 4,363 samples in the same way. Created with BioRender.com.

2.2 Plasmid data retrieval and quality control

The IncL pOXA-48a plasmid has been strongly associated with *E. coli* and *K. pneumoniae* [28]. To explore this plasmid type further, we downloaded accession data for all (n=31) NCBI NucCore entries containing "pOXA-48". Plasmid pKP112 possessed a *bla*OXA-48 gene and therefore was considered a valid pOXA-48 plasmid in this case. These came from the three *Enterobacterales* species: *E. coli* (n=7), *K. pneumoniae* (n=22), and *Serratia marcescens* (n=2). Among these 31 sequences, there were four genetically distinct pOXA-48 isoforms, all isolated from *K. pneumoniae* (strains named Kp11978, KP112, Kpn-30715/15 and K8) (Table 1). Among these four isoforms, pOXA-48_3071 had a *bla*OXA-48 gene that was not properly annotated in the GenBank entry, so it was considered as a pOXA-48-positive plasmid here. We also obtained the unique proteins of four important plasmids linked to antibiotic resistance in *E. coli*: pEK204 (IncI1), pEK499 (IncFII-FIA), pEK516 (IncF) and pEC958 (IncF) (Table S2). Plasmids pEC958 and pEK499 have high similarity (~92%) [74].

Plasmid group	Plasmid name	Strain of origin	Accession number	Length (bp)	Genes	CDSs
pOXA-48	pOXA-48	Kp11978	JN626286.1	61,881	40	74
	pKP112	KP112	LN864819.1	84,252	62	102
	pOXA-48_30715	Kpn-30715/15	NZ_KX523901	65,488	84	84
	pOXA-48_k8	K8	MT441554.1	65,499	87	87
Entero-bacterales	pEK204	Strain C	EU935740	93,732	89	95
	pEK499	Strain A	EU935739	117,536	103	141
	pEK516	Strain D	EU935738	64,471	65	79
	pEC958	EC958	NZ_HG941719	135,602	153	153

Table 1. The plasmids examined came from two groups. The four pOXA-48-related plasmids were originally identified in *K. pneumoniae* (Strain of origin), and the four linked to *Enterobacterales* were initially isolated in *E. coli*.

To assess plasmid protein similarity, we measured the amino acid sequence similarity of the translated genes of the eight plasmids in a pairwise manner using R package Biostrings v2.60.2 [75]. Among the proteins shared by the pOXA-48-related plasmid pairs, 94% (132 out of 138) had a match (*bitscore* < -100). Among the proteins shared by pEK499, pEK516 and pEC958, 94% (130 out of 140) had a match, but pEK204's comparisons with pEK499, pEK516 and pEC958 showed less similarity (18%). Comparisons between each group (pOXA-48-related vs pEK204, pEK499, pEK516 and pEC958) showed only 22% of protein-level similarity for shared proteins.

The PPI information for these eight plasmids was compared with StringDB PPI data for the 24 *Enterobacterales* and 4,363 bacterial samples above using the above packages in R v4.0.1 and Genbankr v1.20.0 [76]. We computed the number of host proteins interacting with each plasmid protein, the number of host-plasmid PPIs, and the Jaccard index of the fraction of interacting host proteins compared to the combined total of sample and plasmid proteins. The number of genes per plasmid was recorded and analysed using R package UpSetR [77]. All p values reported in this study were corrected for multiple testing using the Benjamini-Hochberg method in R.

2.3 Plasmid comparison and estimation of an excess of protein-protein interactions

2,021 samples (out of 4,363) had PPIs with at least one of these four pOXA-48 isoforms, of which 650 had PPIs with all four isoforms. Samples without PPIs were not examined because these provided no information. There were 6,181 sample-plasmid pairs in total with ≥10 interacting proteins and ≥10 PPIs, coming from: pOXA-48 (1,820 pairs), pKP112 (1,868), pOXA-48_30715 (673) and pOXA-48_k8 (1,820). For the four *E. coli* plasmids (pEK204, pEK499, pEK516 and pEC958), all 24 *Enterobacterales*

188 and 1,604 other samples had ≥1 PPI with at least one of these plasmids. There were 3,828 sample-
189 plasmid combinations with ≥10 interacting host proteins and ≥10 PPIs across these samples, coming
190 from: pEK204 (856), pEK499 (1,324), pEK516 (852) and pEC958 (796).
191
192 To quantify instances of high host-plasmid PPI rates for each plasmid set above, we examined on the
193 association between the numbers of interacting host proteins per sample and the number of sample-
194 plasmid PPIs using a linear model predicting the number of PPIs per sample-plasmid pair. We based the
195 prediction on the numbers of: interacting host proteins, plasmid proteins, proteins per sample, and
196 associated pairwise interactions for each plasmid set (Table S3). This showed that the only informative
197 variables were the number of interacting proteins and the number of PPIs per sample and not the other
198 variables. Thus, we used these parameters in a linear model to get the residual numbers of PPIs per
199 sample-plasmid pair, reflecting instances where the observed number of PPIs was far in excess of the
200 expected number. These residuals from the resid R function were normalised such that the average
201 excess of PPIs was zero. We did not explore deficits in the observed PPI numbers.
202
203 **2.4 Interactions between fimbrial adhesin operons**
204
205 We focused further on pEK204's *pil* operon, which is associated with adhesion and has 14 genes, *pilI-V*
206 [37]. We used pEK499 as a baseline because it had the most samples with PPIs (1,324 out of 4,363)
207 among the other three plasmids. We found 1,060 samples had detected PPIs with both pEK204 and
208 pEK499. We examined them using a normalised log2-scaled ratio of the Jaccard indices of pEK204
209 versus pEK499. The pairwise Pearson correlation between 14 *pil* genes was determined using ppcor
210 v1.1 [78] based on presence-absence data for 4,377 samples from across bacteria [79]. Genes with
211 multiple synonyms were collapsed into a single representative. These 14 *pil* genes were also compared
212 to 14 *fim* genes and four *pap* genes in the same manner. We retrieved the *pil, fim* and *pap* gene
213 synonyms and associated Pil, Fim and Pap PPIs in the gene name and PPI data for the 4,363 samples.
214

3 Results

To test for excess PPI rates with different hosts and plasmids, we compared the PPIs of four pOXA-48 plasmid isoforms and four key *E. coli* plasmids with two sample collections whose PPI data was retrieved from StringDB. The first sample collection consisted of 24 *Escherichia*, *Klebsiella*, *Serratia* and *Shigella* samples to examine genus-level plasmid links, and the second collection had 4,363 samples from across bacteria to inform on wider patterns.

3.1 No genus-level association of pOXA-48 within *Enterobacterales*

We examined the PPI levels of four known pOXA-48 isoforms (pOXA-48, pKP112, pOXA-48_30715, pOXA-48_k8, Table 1) with 24 *Enterobacterales* from *Escherichia* and *Shigella* (n=11), *Klebsiella* (n=2) and *Serratia* (n=11) (Table S1). There was a wide range of PPI levels but no genus-level association for any of these plasmids, consistent with previous work [80]. To resolve this finding, we examined the genetic composition of these plasmids. These pOXA-48 isoforms shared just four genes including *ssb*, *pemI/K* and *bla*OXA-48 (Figure 2A) (*bla*OXA-48 was present but not annotated on pOXA-48_3071). PemI is the antitoxin of a type II toxin-antitoxin (TA) system for post-segregational killing that can neutralise its endoribonuclease toxin, PemK [81]. These differences suggested that the plasmids' effects on host cell PPI networks would vary.

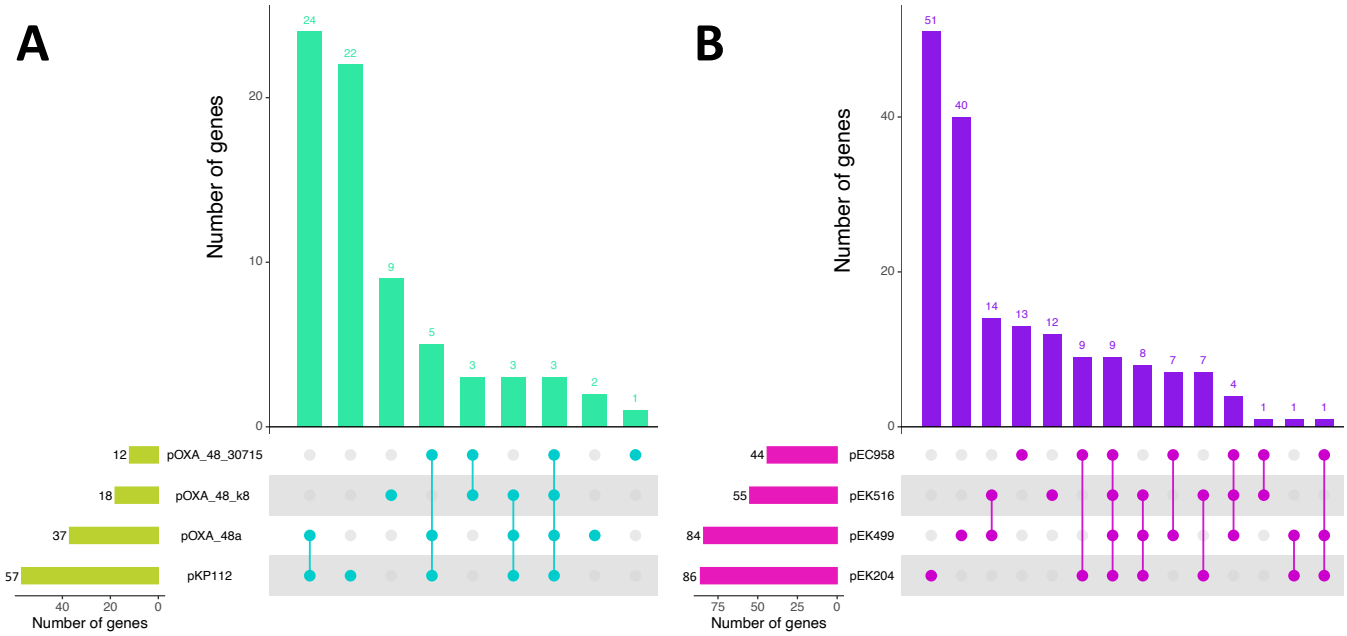


Fig 2. (A) The overlap of unique annotated genes between four pOXA-48-related plasmids (pOXA-48, pKP112, pOXA-48_30715, pOXA-48_k8). Top: The intersection sizes (y-axis) and the genes per set (green) showed that only four genes were shared by all four plasmids: *ssb*, *pemI/K* and *bla*OXA-48 (*bla*OXA-48 was not annotated properly on pOXA-48_3071 and so the intersection of all four plasmids shows three genes but contains *bla*OXA-48 too). Bottom: The numbers of genes per sample (left, lime-green, set size) with the corresponding sets (right: cyan circles). The most divergent isoform was pKP112, which had 22 unique genes. (B) The overlap of genes between the four common *E. coli* plasmids (pEK204, pEK499, pEK516, pEC958). Top: The intersection sizes (y-axis) and the genes per set (purple) showed that nine genes were shared by all four plasmids: *ssb*, *psiB* and *traA/E/L/K/R/V/Y*. Bottom: The numbers of genes per sample (left, pink, set size) with the corresponding sets (right: purple circles). The most divergent isoform was pEK204 (51 unique genes).

3.2 Incomplete associations of pEK204, pEK499, pEK516 and pEC958 with *Escherichia* species

253 Next, we assessed the level of PPIs between these 24 *Enterobacterales* and four major plasmids
254 (pEK204, pEK499, pEK516, pEC958, Table 1). These plasmids possess numerous AMR genes and are
255 prevalent, particularly in certain *E. coli* phylogroups. As above, the numbers of interacting host proteins
256 per sample and PPIs had no clear association (Table S1). Again, there were elevated rates of interacting
257 proteins and PPIs for nine of ten *Escherichia* samples, but *Klebsiella* and *Serratia* were more
258 heterogeneous (Table S1). Consequently, the plasmids' proteomes appeared to be the major factors, like
259 pOXA-48. The gene content of pEK204, pEK499, pEK516 and pEC958 showed nine shared genes: *ssb*,
260 *psiB* and *traA*/*E*/*L*/*K*/*R*/*V*/*Y* (Figure 2B). The *tra* region is associated with conjugation, *psiB* encodes a
261 plasmid SOS inhibition protein that stops induction of the SOS pathway by binding RecA [82], and *ssb*
262 encodes a ssDNA-binding protein involved in DNA replication, recombination and repair [83].
263
264 **3.3 Varied patterns of pOXA-48 interactions across bacteria**
265
266 To shed light on these complex patterns, we extended our analysis of the four pOXA-48 plasmids to all
267 bacterial samples with valid PPI data from StringDB (n=4,363), focusing on 6,181 sample-plasmid
268 combinations with ≥10 interacting host proteins and ≥10 PPIs. Plasmid pOXA-48_30715 had the most
269 interacting host proteins (53±23, median±SD) and PPIs (53±26), followed by pKP112 (interacting
270 proteins, 39±45; PPIs, 43±65), pOXA-48 (interacting proteins, 35±45; PPIs, 37±58), and then pOXA-
271 48_k8 (interacting proteins, 30±33; PPIs, 31±37) (Figure S1). Notably, the interacting host proteins per
272 plasmid had the largest association with the number of PPIs per pair ($0.957 \leq r \leq 0.995$, Figure S2).
273
274 Only 0.6% (40 out of 6,401) of the sample-plasmid pairs had high levels of association, measured as an
275 excess of plasmid-linked PPIs ≥100, ≥100 interacting host proteins and ≥20% excess PPIs after using a
276 model to correct for the number of interacting host proteins per sample-plasmid pair (Figure S3, Table
277 S4). 22 of these 40 were linked to pKP112, 13 to pOXA-48, three to pOXA-48_k8 and two to pOXA-
278 48_30715 (Figure 3). *P. aeruginosa* had the most interacting host proteins of these 40: it shared 19
279 proteins on pKP112 (Figure S4) that had PPIs with 450 other *P. aeruginosa* proteins (Figure S5) with
280 684 PPIs, an excess of 27%. *P. aeruginosa* also had >300 interacting host proteins, >340 PPIs and an
281 excess of PPIs for pOXA-48. The sample with the 2nd highest number of interacting host proteins was
282 *Aromatoleum aromaticum* EbN1, which had 32 proteins on pOXA-48 (Figure S6) interacting with 252
283 *A. aromaticum* proteins with 400 PPIs, an excess of 34% (Figure S7). This sample also had 297
284 interacting proteins linked to pKP112 with 452 PPIs, an excess of 28%. *Enterobacter cloacae subsp.*
285 *cloacae* ATCC 13047 had 125 other host proteins interacting with pKP112 (Figure S8) with 173 PPIs
286 (an excess of 20%, Figure S9). Two *Legionella* and a *Citrobacter* species also had high associations
287 with pKP112 and pOXA-48. *Klebsiella oxytoca* had 173 interacting host proteins with pKP112 that had
288 221 PPIs (an excess of 9%), but no such association with the other three plasmids.
289
290 **3.4 Varied protein-level interactions of bacteria with pEK204, pEK499, pEK516 and pEC958**
291
292 Assessing pEK204, pEK499, pEK516 and pEC958 in the same manner with the larger collection of
293 4,363 samples, we identified 1,604 samples with ≥1 PPI with at least one of these plasmids. We focused
294 further on 3,828 sample-plasmid pairs with ≥10 interacting host proteins and ≥10 PPIs in 1,410 samples,
295 of which 48.4% (683 out of 1,410) had PPIs with all four plasmids. The number of interacting host
296 proteins per sample was comparable across all four plasmids, and highest for pEK204 (median of 64 vs
297 54 for the others, Figure S10). The distributions of numbers of PPIs per sample were similar, but with a
298 much higher variance for pEK204 relative to the other three (13.8 vs 8.2), highlighting that samples had
299 elevated PPI rates (Figure S11).
300
301 Here, 1.5% (56 out of 3,828) of sample-plasmid pairs had strong associations, measured as an excess of
302 plasmid-linked PPIs ≥300, ≥100 interacting host proteins and an excess of PPIs of ≥20% (Table S5). 48
303 were associated with pEK204, five with pEC958, two with pEK499, and one with pEK516 (Figure 4).
304 In particular, the number of pEK204 PPIs per pair varied much more than the other plasmids (Figure
305 S12). Nine of the total of 56 sample-plasmid pairs were from the *Pseudomonas* genus, a pathogen

existing in diverse environments [84]. *P. fluorescens* F113 (aka *P. ogarae* F113) had the most interacting host proteins with pEK204 (500) linked to 1,215 PPIs, an excess PPI rate of 22%. As with pKP112 and pOXA-48, *P. aeruginosa* had 24 pEK204-encoded proteins (including 12 from *pil*) (Figure S13) interacting with 596 other *P. aeruginosa* proteins (Figure S14) with 1,187 PPIs (an excess of 4%). Like pKP112 and pOXA-48 too, *A. aromaticum* EbN1 had 44 proteins on pEK204 including those from the *pil* operon and *tra* region (Figure S15) interacting with 582 other *A. aromaticum* EbN1 proteins (Figure S16) with 1,191 PPIs (an excess of 6%).

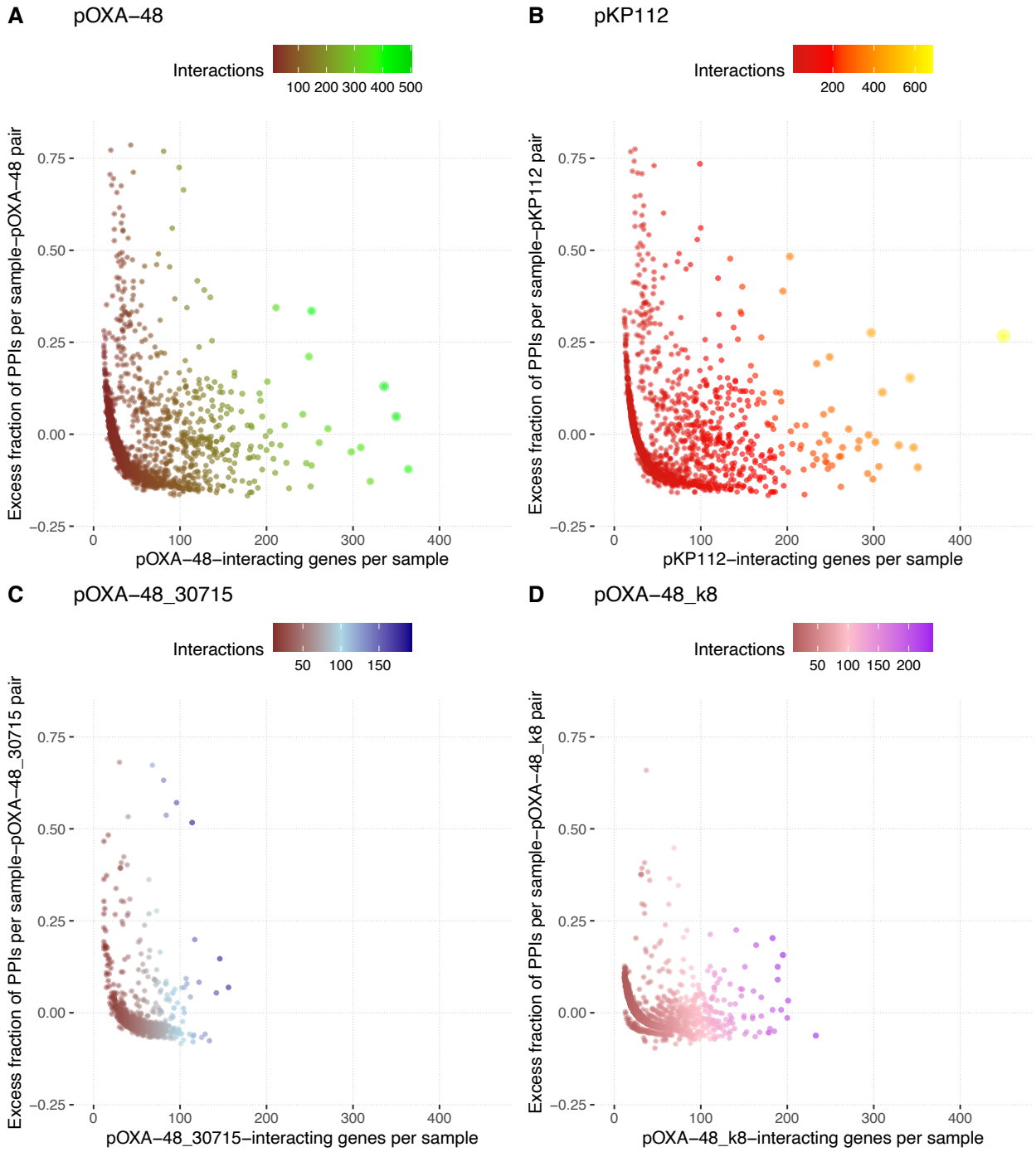


Fig 3. The numbers of interacting host proteins per sample (x-axis) versus the excess fraction of PPIs (y-axis) for each sample-plasmid pair for the four pOXA-48 isoforms (A) pOXA-48 (red-green), (B) pKP112 (red-yellow), (C) pOXA-48_30715 (brown-blue) and (D) pOXA-48_k8 (mauve-purple) using data from 6,181 samples' plasmids with >10 interacting host proteins and >10 PPIs each. The number of PPIs per sample-plasmid pair is denoted by the point size and by the shade based on the scale: larger means fewer PPIs, smaller means more PPIs. Note this scale differs for each plasmid. In (B), *P. aeruginosa* had the most interacting proteins with pKP112 (yellow).

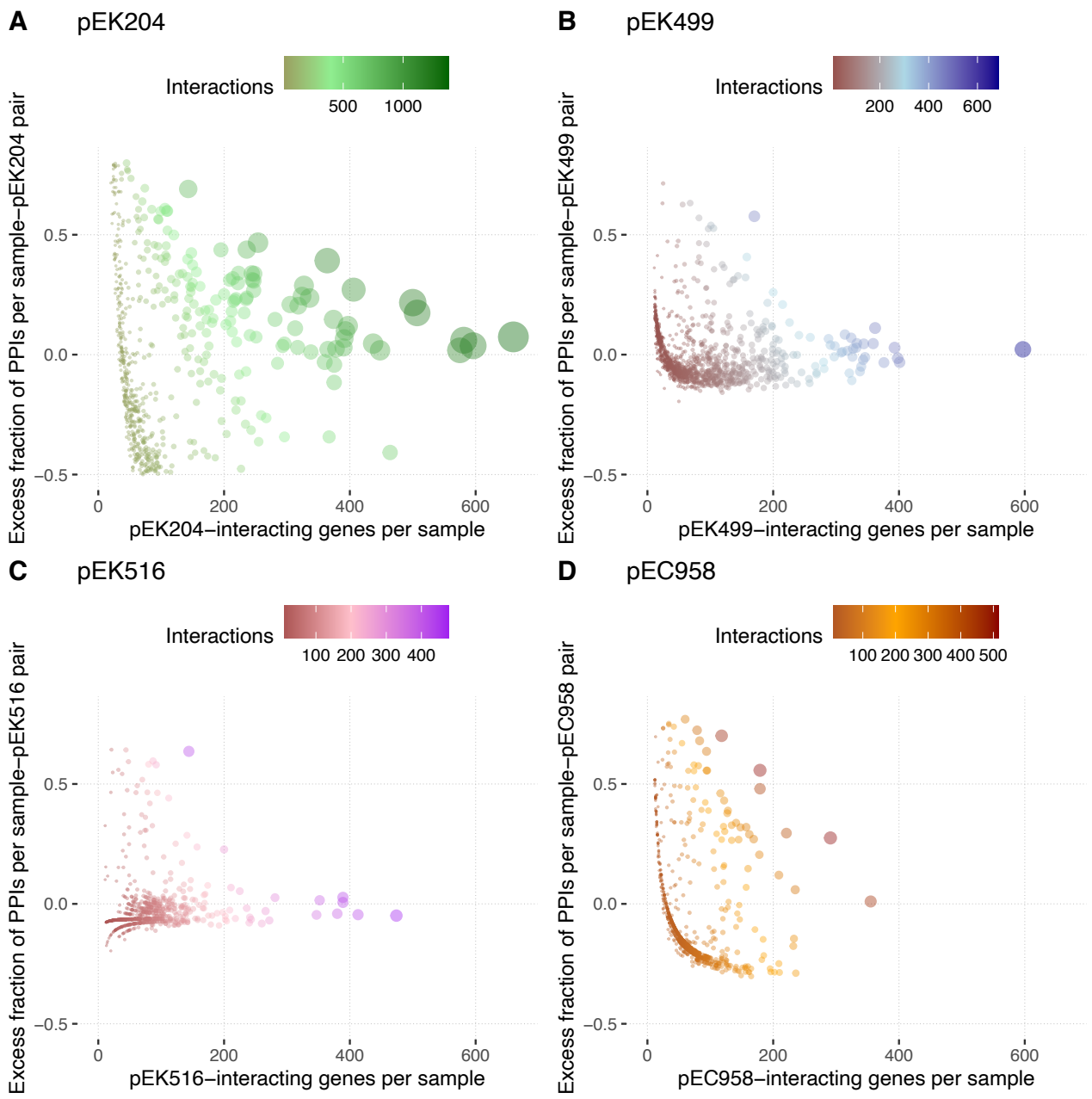


Fig 4. The numbers of interacting host proteins per sample (x-axis) versus the excess fraction of PPIs (y-axis) for each sample-plasmid pair for (A) pEK204 (brown-green), (B) pEK499 (brown-blue), (C) pEK516 (red-mauve) and (D) pEC958 (orange-brown) using data from 3,828 samples' plasmids with 10+ interacting proteins and 10+ PPIs each. The number of PPIs per sample-plasmid pair is denoted by the point size and by the shade based on the scale: larger means fewer PPIs, smaller means more PPIs. Note this scale differs for each plasmid. In (A) for pEK204, the five samples with the most interacting proteins per sample were *P. fluorescens* F113 (aka *P. ogarae* F113), *P. aeruginosa, Aromatoleum aromaticum* EbN1, *Rubrivivax gelatinosus* IL144 and *Cellvibrio* sp. BR were the six rightmost samples (all green). *E. coli* K12 MG1655 had the most interacting proteins with pEK499 (B, blue) and pEK516 (C, mauve). In (D), *Ralstonia eutropha* H16 had the most interacting proteins with pEC958 (brown).

3.5 The *pil* operon caused elevated protein interaction signals in pEK204

We examined the conjugative plasmid pEK204 further given the *pil* operon's role in conjugation. 316 samples out of 1,060 with PPIs for both pEK204 and pEK499 had ≥1 of pEK204's 14 *pilI-pilV* genes: the most common were *pilQ* (encoding a T4P secretin) and *pilT* (encoding a T4P twitching motility

ATPase that helps retract the pilus). We scaled the Jaccard index for pEK204 by that for pEK499 to create a log2-scaled ratio that was positively correlated with the number of interacting host proteins for pEK204. This ratio identified samples with a higher potential for pEK204-specific PPIs, including several *Pseudomonas* samples (*P. aeruginosa, P. fluorescens* F113 as above) (Figure 5, Table S6). Some samples had >11 *pil* operon genes, including *Cellvibrio sp. BR*, whose 13 pEK204 proteins (Figure S17) had the highest number of pEK204-related PPIs of all samples (1,383, an excess of 7.4%), interacting with 660 other *Cellvibrio sp. BR* proteins (Figure S18). Samples with the highest pEK204-linked PPIs were associated with urine or water environments based on environmental niche information from BacDive [85], such as *Aromatoleum aromaticum* EbN1, a Betaproteobacterium from a river source. Similarly, water-associated *Rubrivivax gelatinosus* IL144 had 14 pEK204-matching proteins (Figure S19) that had PPIs with 507 other *Rubrivivax gelatinosus* IL144 proteins with 1,127 PPIs (an excess of 2%, Figure S20).

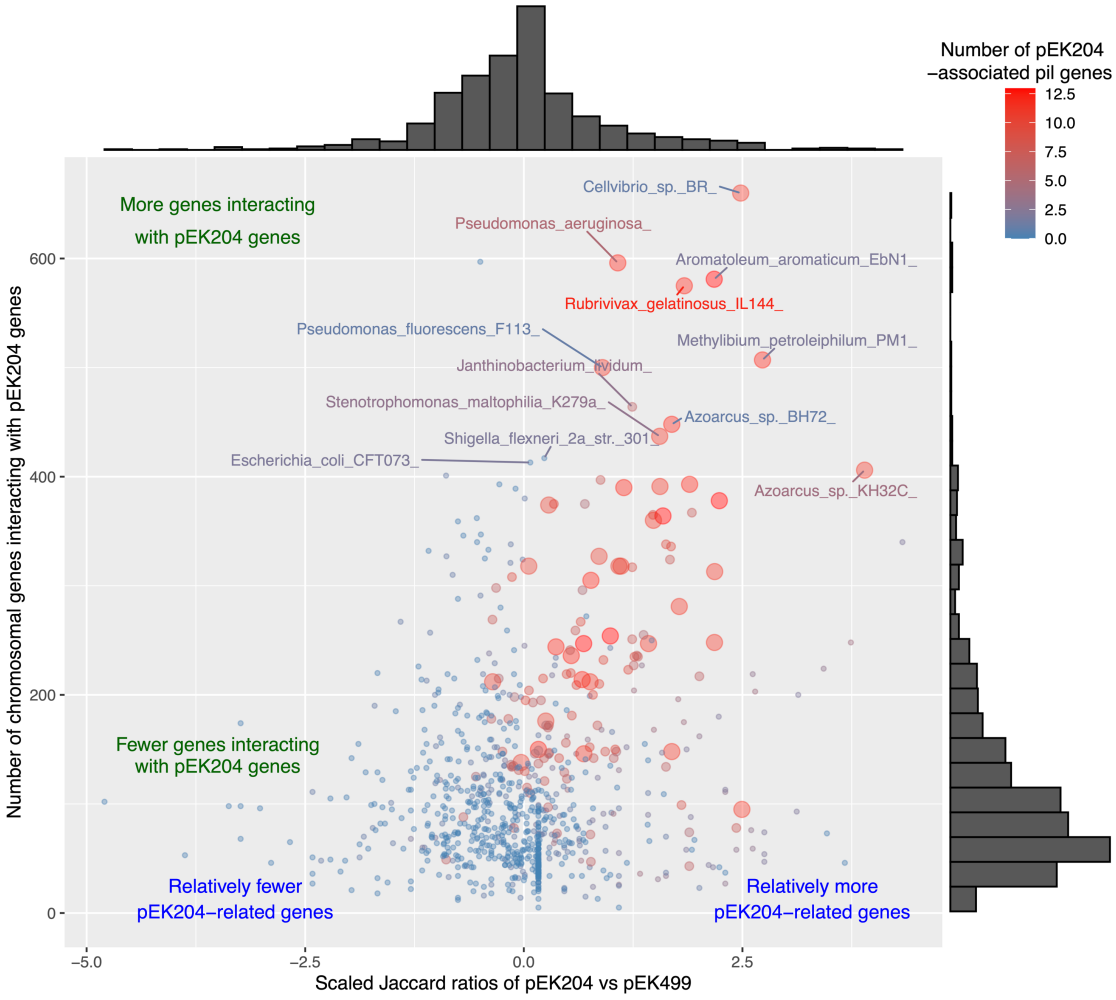


Fig 5. A plot of the normalised log2-scaled ratio of the Jaccard indices of pEK204 divided by that for pEK499 (x-axis) versus the number of sample proteins interacting with pEK204 proteins (y-axis). The data plotted is for 1,060 bacterial samples, of which 316 had ≥1 pEK204-linked *pil* gene(s) (14 genes, *pilIJKLMNOPQRSTUV*. The 12 labelled samples had ≥400 interacting host proteins and a log2-scaled ratio for pEK204 >0. A larger point size and a redder colour signifies the presence of more pEK204-related *pil* genes: blue indicates fewer genes. The histograms indicate the marginal densities of the scatterplots for both axes.

3.6 Interactions of *pil* operon gene products with other fimbrial adhesin proteins

The pairwise correlations of the 14 pEK204 *pil* genes' frequencies across the 4,363 bacterial samples (Table S7) showed that most were positively or negatively correlated with one another, with the

371 exceptions of *pilL*, *pilQ* and *pilV*. Previous work has shown that PilK-PilV, but not PilI or PilJ, are
372 essential for R64 pilus generation [86]. Adding four *pap* genes and 14 *fim* genes to this identified five
373 sets of genes, of which three involved *pil* genes. The first two sets were negatively correlated within
374 their own groups, but were positively correlated with the other set (*pilO/P/R/T* and
375 *pilA/I/J/M/N/S/U/W*/V). The 3rd gene set was largely uncorrelated and stemmed from from FimT's (a
376 T4P biogenesis protein) and FimV's (a motility protein) interactions with PilQ (a ComQ homolog), PilL
377 and PilV (for FimT) and PilF/G/J (for FimV) in *P. aeruginosa* PAO1 and *Acinetobacter baumannii*
378 A118 [87]. Testing for PPIs between *pil, fim* and *pap* proteins in the 4,363 samples found that *Ralstonia*
379 *eutropha* H16 was the single sample with a Fim-Pil PPI across its 13 *pil* genes and four *fim* genes
380 (though major structural protein FimA with PilM-O and pilQ, Figure S22A). Notably, both *E. coli* K12
381 MG1655 (Figure S22B) CFT073 (Figure S22C) had Fim and Pap PPIs, suggesting scope for Pil proteins
382 to interact or evolve such PPIs.
383
384
385 **4. Discussion**
386
387 To explore the hypothesis that an elevated rate of plasmid-host PPIs might be associated with higher
388 compatibility between a plasmid and a host cell, we tested for an excess of host-plasmid PPIs using
389 eight key plasmids in representatives from most bacteria (n=4,363). We achieved this by modelling data
390 from StringDB. We saw that these plasmids had extensive heterogeneity in their PPI levels across 24
391 *Enterobacterales* (*Escherichia*, *Klebsiella*, *Serratia* and *Shigella*) samples. This absence of a genus-level
392 association was consistent with previous work [80].
393
394 To illuminate this issue, we explored the four genetically distinct *bla*OXA-48-positive plasmids in a
395 larger set of 4,363 samples and found that these plasmids had ≥1 PPI with most bacteria, allowing us to
396 identify instances where there was a large excess of PPIs in 40 (0.6%) sample-plasmid pairs. This
397 identified *P. aeruginosa, A. aromaticum* EbN1 and 21 other common animal and plant pathogens as
398 more likely to have PPIs with ≥1 of these *bla*OXA-48-positive plasmids, particularly with the pKP112
399 and pOXA-48 isoforms, suggesting those plasmids could become common in these species. Consistent
400 with these predictions, *bla*OXA-48-positive human clinical isolates from Bangladesh [88] and Sudan
401 [89] have been found in *P. aeruginosa*, even though previous *in vitro* experiments suggested
402 incompatibility [34]. This is concerning because *bla*OXA-48 is the most common AMR gene in CPE
403 and is prevalent in *Pseudomonas* based on carbapenemase gene rates in England from 2020-22 [15].
404 Given their already frequent retention of these plasmids, the high levels of carbapenems in clinical
405 environments may allow these hosts to obtain, co-evolve with and compensate for the metabolic cost of
406 possessing these plasmids [90].
407
408 We used the same approach to assess the other group of four genetically related plasmids (pEK204,
409 pEK499, pEK516, pEC958) that likely originated in ExPEC [37]. ExPEC pose a significant public
410 health problem and change their plasmid composition to improve their fitness [91]. Like the pOXA-48-
411 related plasmids, these plasmids had ≥1 PPI with many bacteria, and we found large excess rates of PPIs
412 in 48 species. 19% of these were from the *Pseudomonas* genus, where *P. fluorescens* F113 and *P.*
413 *aeruginosa* had strong associations with pEK204. This was concerning because pEK204 has spread
414 across pathogenic bacteria [26], and IncI1 plasmids like pEK204 promote AMR due to the presence of
415 *bla*CTX-M, *bla*TEM, *bla*CMY and other ESBL genes [92-93].
416
417 *P. fluorescens* F113 and *P. aeruginosa* were among 12 samples with high rates of PPIs with pEK204,
418 particularly in samples whose proteins interacted with *pil* operon proteins. This operon encodes a T4P
419 with two main known isoforms [94], had 14 genes on pEK204, and is a virulence factor for *P.*
420 *aeruginosa* [95-96]. A ten-gene *pil* operon is involved in conjugation of pathogenicity island PAPI-1 in
421 *P. aeruginosa* PA14 independently of pEK204 [95] that has been linked to *P. aeruginosa* ST309
422 infection outbreaks [97], and *pil* genes form part of an integrative conjugative element Pac_ICE1 in *P.*
423 *syringae* pv actinidiae [98]. Notably, *pil* may be more associated with aqueous environments, and the

loss of *pil* genes from PAPI-1 has been associated with better acute human lung [99] and mouse gut [100] infection, perhaps due to Pil proteins acting as cell surface antigens [101]. Consistent with this connection to water, our approach identified bacteria associated with rivers like *Cellvibrio sp. BR* and others found in wastewater, sludge, sewage and freshwater *Rubrivivax gelatinosus* IL144 [100]. We identified three distinct groups of correlated *pil* gene rates in these 4,363 samples, suggesting that pEK204's *pil* repertoire differed from other *pil* operon compositions. We found evidence of PPIs between *pil, fim* and *pap* gene products in *E. coli* and *R. eutropha*, which is important because ExPEC can possess type 1, IV and P fimbriae and tend to express a single fimbrial type at any time depending on the host niche, indicating coordinated expression [103-105]. In this way, the gain of *pil* via pEK204 could facilitate superior host survival, conjugation or biofilm formation in water environments [41].

A limitation of this study was the inconsistent nomenclature of plasmid and samples, which meant that certain protein matches were not detected or misclassified, and affecting PPI identification. For instance, functions were assigned to ~46% of genes in previous work [106]. In addition, we did not map reads or align sequences here to detect genes and relied on the accuracy of existing gene annotation tools based on protein sequence patterns. Such DNA sequence analysis would yield diverse and likely ambiguous gene cluster alignment scores that also lead to matches being unclear or imprecisely categorised. Future work is needed to improve genome and plasmid annotation, explore the strength and specificity of PPIs, and assess other highly transmissible plasmids.

Conclusions

Given that no study yet has explored PPI between key plasmids across all bacteria to date, so we developed a model testing for excess rates of PPIs between 4,363 bacterial proteomes and the proteomes of eight plasmids of public health concern. 23 potential bacterial host cells had high PPI rates with four *bla*OXA-48-positive plasmids, including *P. aeruginosa*. Similarly, 48 species had elevated PPI rates with four common *E. coli* plasmids, including nine from the *Pseudomonas* genus. One of the latter plasmids (pEK204), is associated with host cell binding and conjugative transfer in aqueous environments, and appeared to be a major driver of host-plasmid PPIs in the species with high rates of PPIs with pEK204. These results overcome a limitation that the gain of a new plasmid may have heterogenous effects on different host cells, and this can depend on genetic differentiation levels, as has been illustrated for pOXA-48 [80]. Consequently, our approach using excess rates of PPIs between plasmid and host cell proteins added results on the potential compatibility of plasmids with particular species. Our study has implications for controlling emerging problems, such as *bla*OXA-48-carrying plasmids that can spread among numerous diverse host genera.

Author Contributions

TD – Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualisation, Project administration. MJL – Methodology, Formal analysis, Investigation, Data Curation, Visualisation. CA – Formal analysis, Investigation, Data Curation, Visualisation. AMcD – Formal analysis, Investigation, Data Curation, Visualisation. AR – Conceptualization, Methodology, Writing - Review & Editing.

Acknowledgements

The authors acknowledge School of Biotechnology (Dublin City University) student funding. We thanks the reviewers and others with comments for their constructive input.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Data Summary

Processing files and relevant code associated with this paper are available on Github at https://github.com/downingtim/plasmids_2022.

S. Kurtz, A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes, Genome Biol. 5 (2004) 1–9.

References

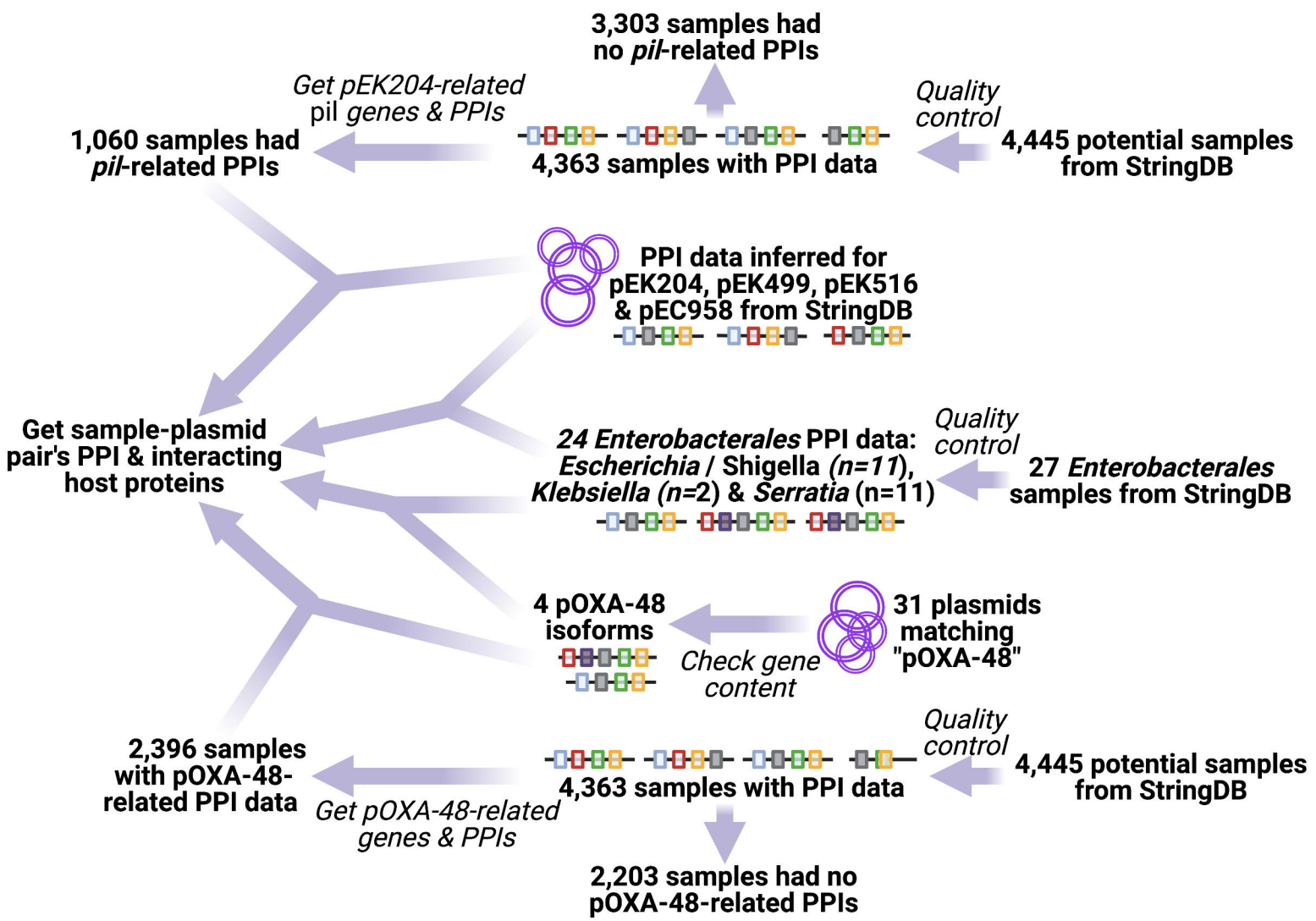
1. S. Partridge, S. R., Kwong, S. M., Firth, N. & S. O. Jensen, Mobile genetic elements associated with antimicrobial resistance, Clin. Microbiol. Rev. 31 (2018) e00088-17.
2. A. San Millan, Evolution of plasmid-mediated antibiotic resistance in the clinical context, Trends Microbiol. 26 (2018) 978–985.
3. Y. Che, et al, Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes, Proc Natl Acad Sci U S A, 118(6) (2021) e2008731118. doi: 10.1073/pnas.2008731118
4. C. S. Smillie, et al, Ecology drives a global network of gene exchange connecting the human microbiome, Nature 480 (2011) 241–244.
5. U. Canosi, G. Lüder, T. A. Trautner, SPP1-mediated plasmid transduction, J Virol. 44(2) (1982) 431-6. doi: 10.1128/JVI.44.2.431-436.1982
6. S. Erdmann, B. Tschitschko, L. Zhong, M. J. Raftery, R. Cavicchioli, A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells, Nat Microbiol. 2(10) (2017) 1446-1455. doi: 10.1038/s41564-017-0009-2
7. X. Zhang, et al, Improvement in the efficiency of natural transformation of *Haemophilus parasuis* by shuttle-plasmid methylation, Plasmid, 98 (2018) 8-14. doi: 10.1016/j.plasmid.2018.07.001
8. T. Wein, T. Dagan, Plasmid evolution, Curr Biol, 30(19) (2020) R1158-R1163. Doi: 10.1016/j.cub.2020.07.003
9. A. San Millan, R. C. MacLean, Fitness Costs of Plasmids: a Limit to Plasmid Transmission. Microbiol Spectr. 5(5) (2017). doi: 10.1128/microbiolspec.MTBP-0016-2017
10. J. P. J. Hall, et al, Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation, PLoS Biol. 19(10) (2021) e3001225. doi: 10.1371/journal.pbio.3001225
11. M. H. Nicolas-Chanoine, X. Bertrand, J. Y. Madec. *Escherichia coli* ST131, an intriguing clonal group. Clin Microbiol Rev 27 (2014) 543–574.
12. L. Neffe, L. Abendroth, W. Bautsch, S Häussler, J. Tomasch, High plasmidome diversity of extended-spectrum beta-lactam-resistant *Escherichia coli* isolates collected during one year in one community hospital, Genomics 114(3) (2022) 110368. doi: 10.1016/j.ygeno.2022.110368
13. E. Tacconelli, et alm WHO Pathogens Priority List Working Group. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis, Lancet Infect Dis. 18(3) (2018) 318-327. doi: 10.1016/S1473-3099(17)30753-3
14. K. L. Hopkins, et al, Diversity of carbapenemase-producing *Enterobacterales* in England as revealed by whole-genome sequencing of isolates referred to a national reference laboratory over a 30-month period, J Med Microbiol. 71(5) (2022). doi: 10.1099/jmm.0.001518
15. UK Health Security Agency, Quarterly laboratory surveillance of acquired carbapenemase-producing Gram-negative bacteria in England: October 2020 to March 2022, Health Protection Report 16(5) (2022).
16. J. J. Jacob, A. K. Pragasam, K. Vasudevan, B. Veeraraghavan, G. Kang, J. John, V. Nagvekar, A. Mutreja, *Salmonella* Typhi acquires diverse plasmids from other Enterobacteriaceae to develop cephalosporin resistance, Genomics 113(4) (2021) 2171-2176. doi: 10.1016/j.ygeno.2021.05.003.
17. V. Calhau, G. Ribeiro, N. Mendonça, G. J. Da Silva, Prevalent combination of virulence and plasmidic-encoded resistance in ST 131 *Escherichia coli* strains, Virulence 4 (2013) 726–729.
18. J. T. Poolman, M. Wacker, Extraintestinal pathogenic *Escherichia coli*, a common human pathogen: challenges for vaccine development and progress in the field, J Infect Dis 213 (2016) 6–13.

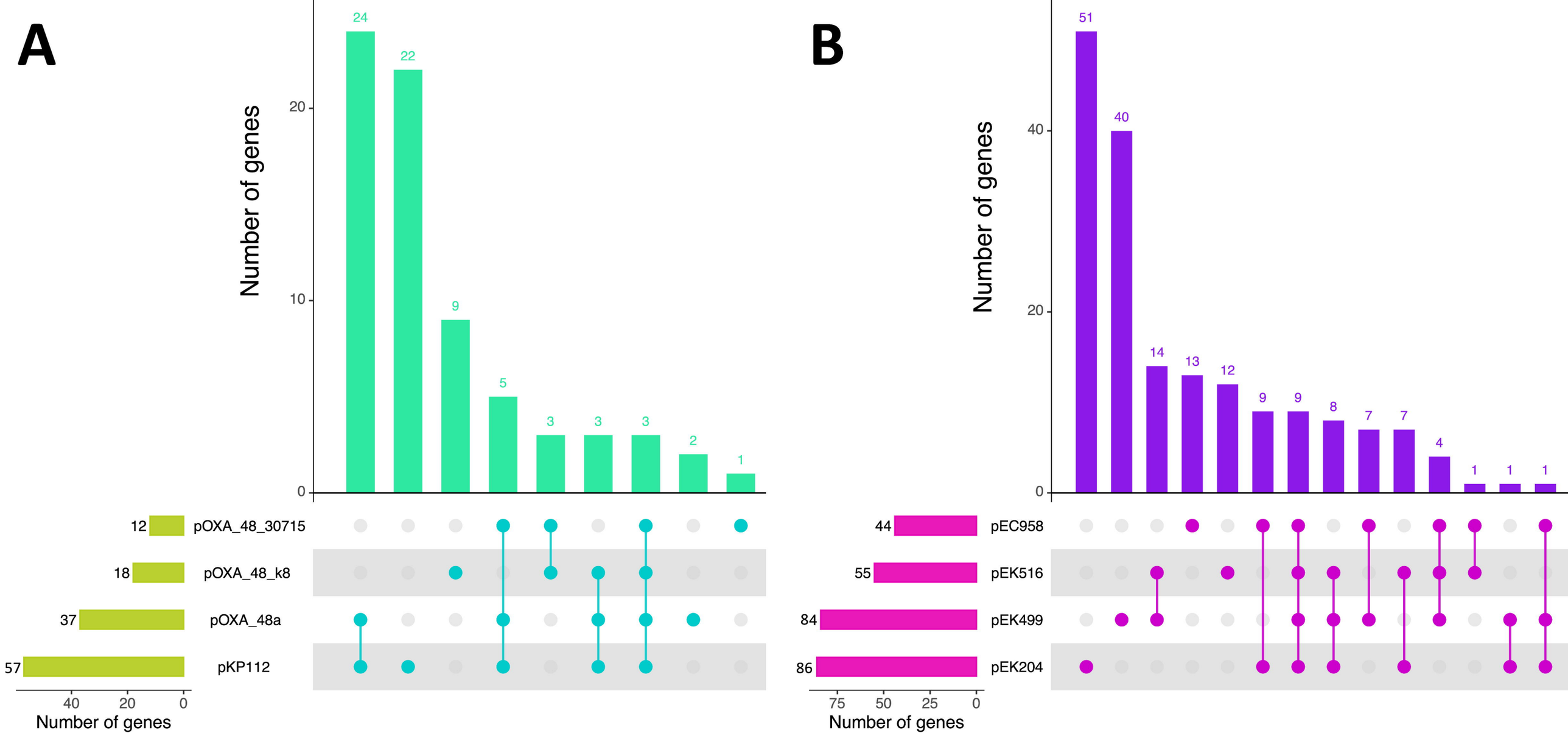
532 19. N. L. Ben Zakour, et al, Sequential acquisition of virulence and fluoroquinolone resistance has
533 shaped the evolution of *Escherichia coli* ST131, mBio 7 (2016) e00347.
534 20. A. McNally, et al, Combined analysis of variation in core, accessory and regulatory genome
535 regions provides a super-resolution view into the evolution of bacterial populations, PLoS Genet
536 12 (2016) 1006280.
537 21. J. R. Johnson, et al, Household Clustering of *Escherichia coli* Sequence Type 131 Clinical and
538 Fecal Isolates According to Whole Genome Sequence Analysis, Open Forum Infect Dis 3 (2016)
539 ofw129.
540 22. T. Kallonen, et al, Systematic longitudinal survey of invasive *Escherichia coli* in England
541 demonstrates a stable population structure only transiently disturbed by the emergence of ST131,
542 Genome Res 27 (2017) 1437–1449.
543 23. C. Goswami, et al, Genetic analysis of invasive *Escherichia coli* in Scotland reveals determinants
544 of healthcare-associated versus community-acquired infections. Microb Genom 4 (2018).
545 24. A. G. Decano, et al, Complete assembly of *Escherichia coli* ST131 genomes using long reads
546 demonstrates antibiotic resistance gene variation within diverse plasmid and chromosomal
547 contexts, mSphere 4 (2019) e00130–19.
548 25. S. Ny, L. Sandegren, M. Salemi, C. G. Giske. Genome and plasmid diversity of extended-
549 spectrum β-lactamase-producing *Escherichia coli* ST131 – tracking phylogenetic trajectories with
550 Bayesian inference, Sci Rep 9 (2019) 10291.
551 26. R. León-Sampedro, et al, Pervasive transmission of a carbapenem resistance plasmid in the gut
552 microbiota of hospitalized patients, Nat Microbiol. 6(5) (2021) 606-616. doi: 10.1038/s41564-
553 021-00879-y.
554 27. L. Poirel, T. Naas, P. Nordmann, Diversity, epidemiology, and genetics of class D beta-
555 lactamases, Antimicrob Agents Chemother. 54(1) (2010) 24-38. doi: 10.1128/AAC.01512-08
556 28. L. Poirel, C. Héritier, V. Tolün, P. Nordmann P., Emergence of oxacillinase-mediated resistance
557 to imipenem in *Klebsiella pneumoniae*, Antimicrob Agents Chemother. 48(1) (2004) 15-22. doi:
558 10.1128/AAC.48.1.15-22.2004
559 29. D. J. O'Brien, et al, First isolation and outbreak of OXA-48-producing *Klebsiella pneumoniae* in
560 an Irish hospital, March to June 2011, Euro Surveill. 16(29) (2011) 19921.
561 30. C. Wrenn, et al, Investigation of the first outbreak of OXA-48-producing *Klebsiella pneumoniae*
562 in Ireland, J Hosp Infect 87(1) (2014) 41-6. doi: 10.1016/j.jhin.2014.03.001.
563 31. D. Morris, et al, Dissemination of clonally related multidrug-resistant *Klebsiella pneumoniae* in
564 Ireland, Epidemiol Infect. 144(2) (2016) 443-8. doi: 10.1017/S0950268815001041.
565 32. C. Brehony, et al, Molecular Epidemiology of an Extended Multiple-Species OXA-48 CPE
566 Outbreak in a Hospital Ward in Ireland, 2018–2019, Antimicrobial Stewardship & Healthcare
567 Epidemiology 1(1) (2021) e54. doi: https://doi.org/10.1017/ash.2021.206
568 33. B. D. Johnston, et al, Global molecular epidemiology of carbapenem-resistant *Escherichia coli*
569 (2002-2017), Eur J Clin Microbiol Infect Dis. 2021 doi: 10.1007/s10096-021-04310-6.
570 34. L. Poirel, R. A. Bonnin, P. Nordmann, Genetic features of the widespread plasmid coding for the
571 carbapenemase OXA-48, Antimicrob Agents Chemother. 56(1) (2012) 559-62. doi:
572 10.1128/AAC.05289-11
573 35. T. Downing, Tackling Drug Resistant Infection Outbreaks of Global Pandemic *Escherichia coli*
574 ST131 Using Evolutionary and Epidemiological Genomics, Microorganisms 3(2) (2015) 236-67.
575 doi: 10.3390/microorganisms3020236
576 36. C. Goswami, et al, Origin, maintenance and spread of antibiotic resistance genes within plasmids
577 and chromosomes of bloodstream isolates of *Escherichia coli,* Microb Genom. 6(4) (2020)
578 e000353. doi: 10.1099/mgen.0.000353
579 37. N. Woodford, et al, Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516,
580 encoding CTX-M enzymes in three major *Escherichia coli* lineages from the United Kingdom, all
581 belonging to the international O25:H4-ST131 clone, Antimicrob Agents Chemother 53 (2009)
582 4472–4482.
583 38. H. Dhanji, et al, Molecular epidemiology of fluoroquinolone-resistant ST131 *Escherichia coli*
584 producing CTX-M extended-spectrum beta-lactamases in nursing homes in Belfast, UK, J
585 Antimicrob Chemother. 66(2) (2011) 297-303. doi: 10.1093/jac/dkq463.

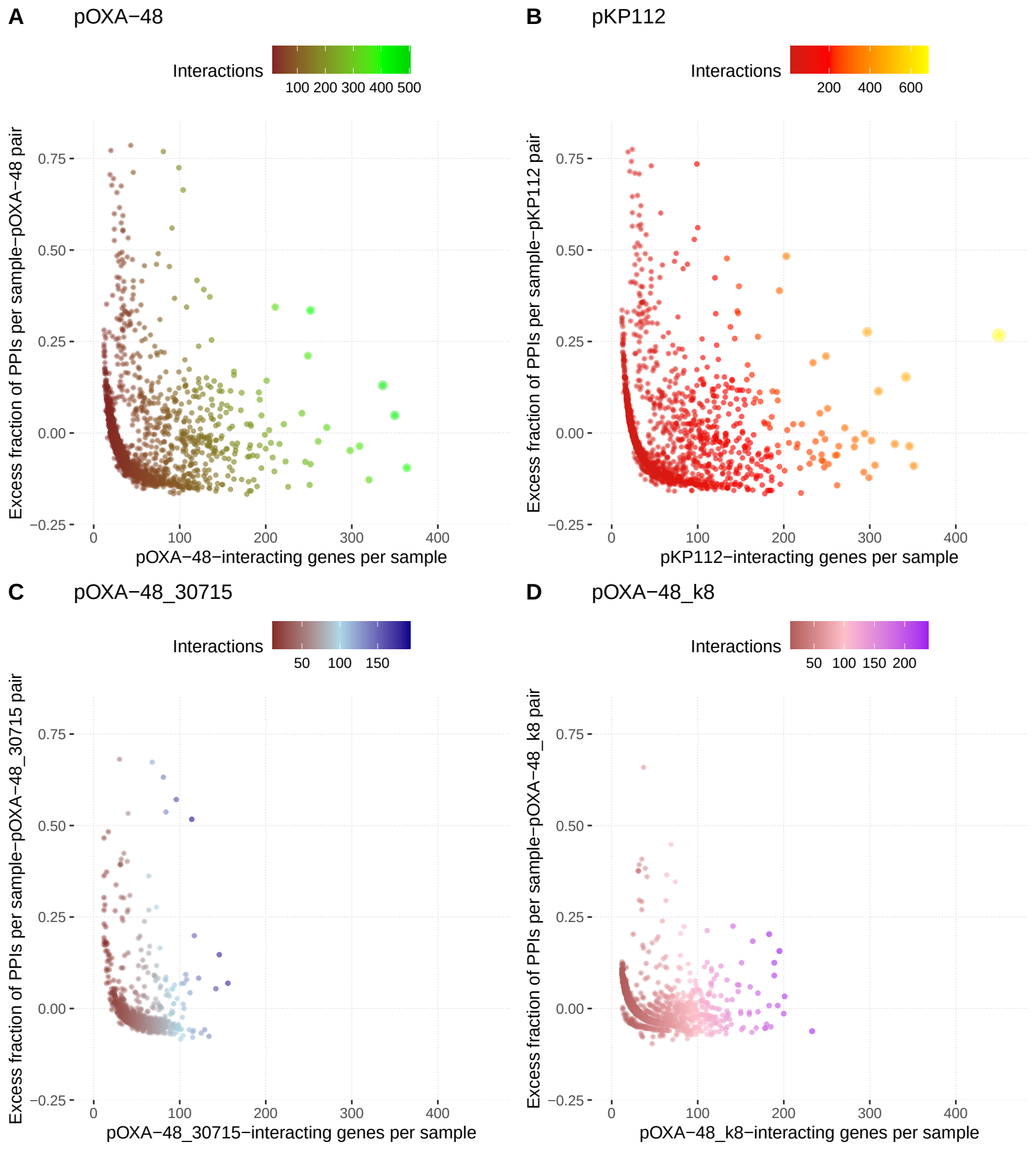
586 39. M. Doumith, et al, Characterization of plasmids encoding extended-spectrum β-lactamases and
587 their addiction systems circulating among *Escherichia coli* clinical isolates in the UK, J
588 Antimicrob Chemother. 67(4) (2012) 878-85. doi: 10.1093/jac/dkr553
589 40. A. G. Decano, T. Downing, An *Escherichia coli* ST131 pangenome atlas reveals population
590 structure and evolution across 4,071 isolates, Sci Rep. 9(1) (2019) 17394. doi: 10.1038/s41598-
591 019-54004-5
592 41. E. G. Dudley, et al, An IncI1 plasmid contributes to the adherence of the atypical
593 enteroaggregative *Escherichia coli* strain C1096 to cultured cells and abiotic surfaces, Infect
594 Immun 74 (2006) 2102–211.
595 42. S. Sarkar, D. Vagenas, M. A. Schembri, M. Totsika, Biofilm formation by multidrug resistant
596 *Escherichia coli* ST131 is dependent on type 1 fimbriae and assay conditions, Pathog Dis 74
597 (2016) ftw013.
598 43. M. J. Kuehn, J. Heuser, S. Normark, S. J. Hultgren, P pili in uropathogenic *E. coli* are composite
599 fibres with distinct fibrillar adhesive tips, Nature 356 (1992) 252–255.
600 44. I. Connell, et al, Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary
601 tract, Proc Natl Acad Sci U S A 93 (1996) 9827–9832.
602 45. J. J. Martinez, M. A. Mulvey, J. D. Schilling, J. S. Pinkner, S. J. Hultgren, Type 1 pilus-mediated
603 bacterial invasion of bladder epithelial cells, Embo J 19 (2000) 2803–2812.
604 46. R. Selvarangan, et al, Interaction of Dr adhesin with collagen type IV is a critical step in
605 *Escherichia coli* renal persistence, Infect Immun 72 (2004) 4827–4835.
606 47. S. P. Nuccio, A. J. Baumler, Evolution of the chaperone/usher assembly pathway: fimbrial
607 classification goes Greek, Microbiol Mol Biol Rev 71(4) (2007) 551-575.
608 48. G. T. Werneburg, D. G. Thanassi, Pili Assembled by the Chaperone/Usher Pathway in
609 *Escherichia coli* and *Salmonella*, EcoSal Plus 8(1) (2018).
610 49. D. J. Wurpel, S. A. Beatson, M. Totsika, N. K. Petty, M. A. Schembri, Chaperone-usher fimbriae
611 of *Escherichia coli*, PLoS One 8(1) (2013) e52835.
612 50. L. Craig, K. T. Forest, B. Maier, Type IV pili: dynamics, biophysics and functional consequences,
613 Nat. Rev. Microbiol. 17 (2019) 429–440.
614 51. G. A. O'Toole, R. Kolter, Flagellar and twitching motility are necessary for *Pseudomonas*
615 *aeruginosa* biofilm development, Mol Microbiol 30 (1998) 295–304.
616 52. J. C. Comolli, et al, *Pseudomonas aeruginosa* gene products PilT and PilU are required for
617 cytotoxicity in vitro and virulence in a mouse model of acute pneumonia, Infect Immun 67 (1999)
618 3625–3630.
619 53. A. K. Schmidt, et al, A Filamentous Bacteriophage Protein Inhibits Type IV Pili To Prevent
620 Superinfection of *Pseudomonas aeruginosa*, mBio e0244121 (2022). doi: 10.1128/mbio.02441-21
621 54. J. L. Cottell, M. A. Webber, L. J. L. Piddock, Persistence of transferable extended-spectrum-β-
622 lactamase resistance in the absence of antibiotic pressure, Antimicrob. Agents Chemother. 56(9)
623 (2012) 4703-4706.
624 55. A. G. Decano, et al, Plasmids shape the diverse accessory resistomes of *Escherichia coli* ST131,
625 Access Microbiol. 3(1) (2020) acmi000179. doi: 10.1099/acmi.0.000179
626 56. S. K. Miryala, S. Ramaiah, Exploring the multi-drug resistance in *Escherichia coli* O157:H7 by
627 gene interaction network: A systems biology approach, Genomics 111 (2019) 958–965. doi:
628 10.1016/j.ygeno.2018.06.002
629 57. D. Szklarczyk, et al, The STRING database in 2021: customizable protein-protein networks, and
630 functional characterization of user-uploaded gene/measurement sets, Nucleic Acids Res. 49(D1)
631 (2021) D605-D612. doi: 10.1093/nar/gkaa1074.
632 58. D. J. Winter., rentrez: an R package for the NCBI eUtils API, The R Journal 9(2) (2017) 520-526.
633 59. D. Szklarczyk, et al, STRING v11: protein-protein association networks with increased coverage,
634 supporting functional discovery in genome-wide experimental datasets, Nucleic acids research
635 47.D1 (2018) D607-D613.
636 60. H. Wickham, stringr: Simple, Consistent Wrappers for Common String Operations, R package
637 version 1.4.0 (2019) https://CRAN.R-project.org/package=stringr
638 61. R Core Team, R: A language and environment for statistical computing, R Foundation for
639 Statistical Computing, Vienna, Austria (2021), URL https://www.R-project.org

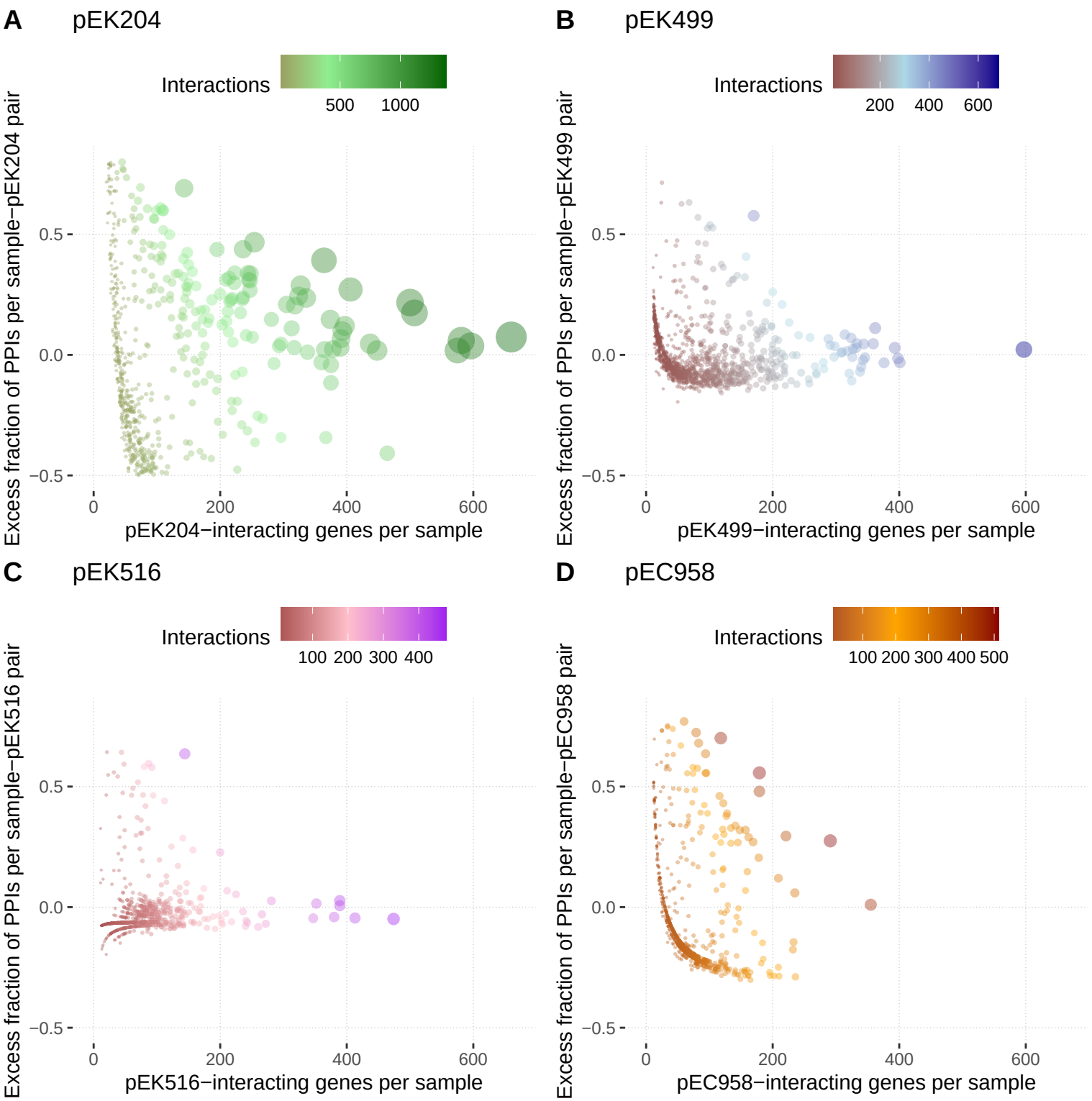
62. RStudio Team, RStudio: Integrated Development Environment for R, RStudio, PBC, Boston, MA (2022), URL http://www.rstudio.com
63. R. Ihaka, R. Gentleman, R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics, (5)3 (1996) 299-314.
64. H. Wickham, R. François, L. Henry, K. Müller., dplyr: A Grammar of Data Manipulation, R package version 1.0.8 (2022), https://CRAN.R-project.org/package=dplyr
65. H. Wickham, forcats: Tools for Working with Categorical Variables (Factors), R package version 0.5.1 (2021), https://CRAN.R-project.org/package=forcats
66. H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York (2016).
67. K. Slowikowski, ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2', R package version 0.9.1 (2021), https://CRAN.R-project.org/package=ggrepel
68. H. Wickham, J. Hester, J. Bryan, readr: Read Rectangular Text Data, R package version 2.1.2 (2022), https://CRAN.R-project.org/package=readr
69. H. Wickham, J. Bryan, readxl: Read Excel Files, R package version 1.4.0 (2022), https://CRAN.R-project.org/package=readxl
70. K. Müller, H. Wickham, tibble: Simple Data Frames, R package version 3.1.6 (2021), https://CRAN.R-project.org/package=tibble
71. H. Wickham, M. Girlich, tidyr: Tidy Messy Data, R package version 1.2.0 (2022), https://CRAN.R-project.org/package=tidyr
72. H. Wickham, et al, Welcome to the tidyverse, Journal of Open Source Software 4(43) (2019) 1686 doi: https://doi.org/10.21105/joss.01686.
73. H. Chen, VennDiagram: Generate High-Resolution Venn and Euler Plots, R package version 1.7.3 (2022), https://CRAN.R-project.org/package=VennDiagram
74. M. Totsika, et al, Insights into a multidrug resistant *Escherichia coli* pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms, PLoS One 6(10) (2011) e26578. doi: 10.1371/journal.pone.0026578
75. H. Pagès, P. Aboyoun, R. Gentleman, S. DebRoy, Biostrings: Efficient manipulation of biological strings, R package version 2.60.2 (2021), https://bioconductor.org/packages/Biostrings
76. G. Becker, M. Lawrence, genbankr: Parsing GenBank files into semantically useful objects, R package version 1.24.0 (2022).
77. J. R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties, Bioinformatics 33(18) (2017) 2938-2940 doi: 10.1093/bioinformatics/btx364
78. S. Kim, ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients, Communications for Statistical Applications and Methods 22(6) (2015) 665-674.
79. T. Downing, A. D. Rahm, Bacterial plasmid-associated and chromosomal proteins have fundamentally different properties in protein interaction networks, BioRxiv (2022) doi: https://doi.org/10.1101/2022.04.29.490008
80. A. Alonso-del Valle, et al, Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities, Nat Commun. 12(1) (2021) 2653, doi: 10.1038/s41467-021-22849-y.
81. S. Tsuchimoto, H. Ohtsubo, E. Ohtsubo, Two genes, pemK and pemI, responsible for stable maintenance of resistance plasmid R100, J Bacteriol. 170(4) (1988) 1461-6, doi: 10.1128/jb.170.4.1461-1466.1988.
82. A. Bailone, et al, PsiB polypeptide prevents activation of RecA protein in *Escherichia coli*, Mol Gen Genet. 214(3) (1988) 389-95, doi: 10.1007/BF00330471.
83. R. R. Meyer, P. S. Laine, The single-stranded DNA-binding protein of *Escherichia coli*, Microbiol Rev. 54(4) (1990) 342-80, doi: 10.1128/mr.54.4.342-380.1990.
84. A. Peix, M. H. Ramírez-Bahena, E. Velázquez. The current status on the taxonomy of *Pseudomonas* revisited: An update, Infect Genet Evol. 57 (2018) 106-116, doi: 10.1016/j.meegid.2017.10.026
85. L. C. Reimer, et al, BacDive in 2019: bacterial phenotypic data for High-throughput biodiversity analysis., Nucleic Acids Research database issue (2019).
86. T. Yoshida, S. R. Kim, T. Komano, Twelve pil genes are required for biogenesis of the R64 thin pilus, J Bacteriol. 181(7) (1999) 2038-43, doi: 10.1128/JB.181.7.2038-2043.1999.

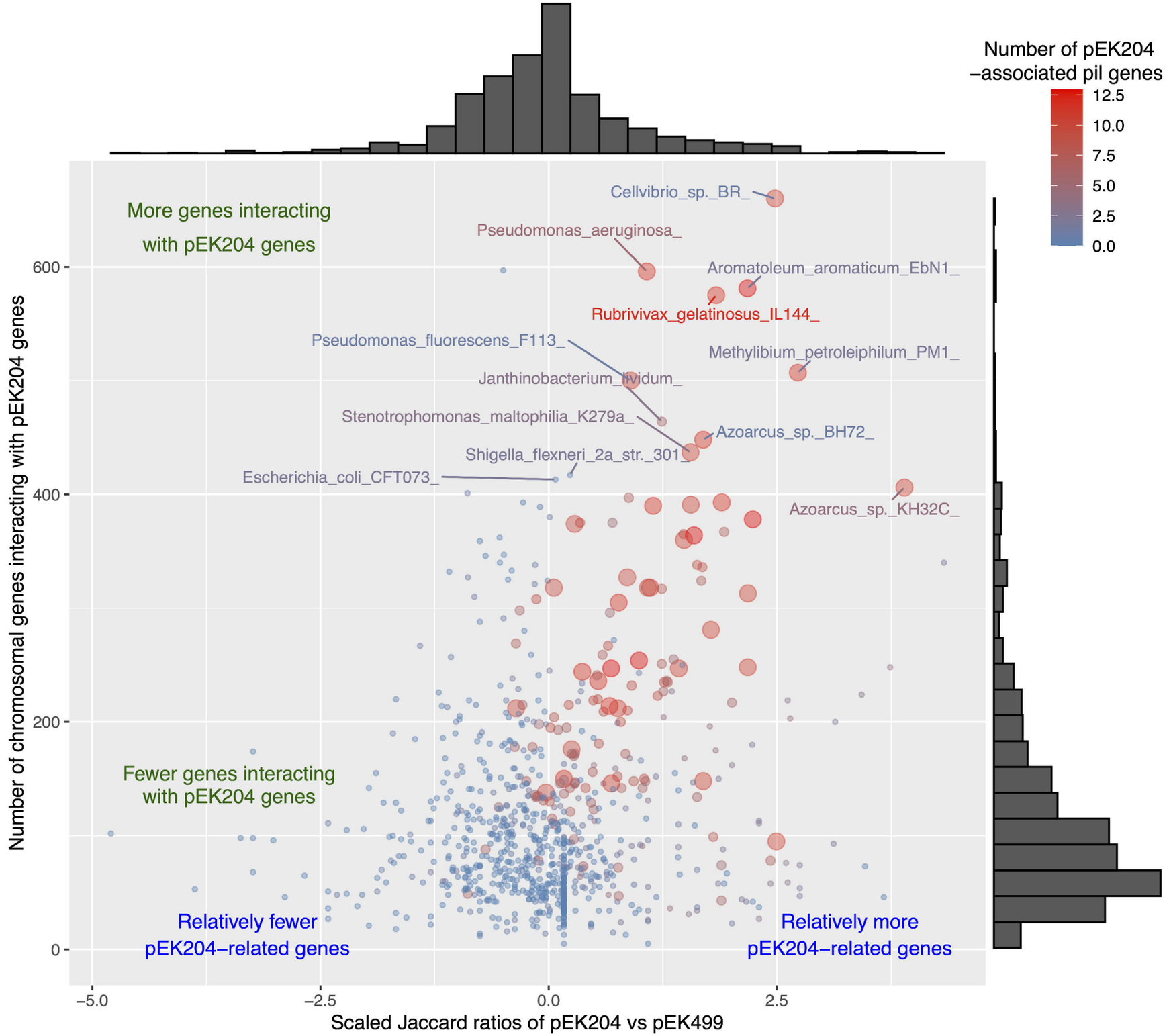
87. R. A. Alm, J. S. Mattick, Identification of two genes with prepilin-like leader sequences involved in type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*, J Bacteriol. 178(13) (1996) 3809-17, doi: 10.1128/jb.178.13.3809-3817.1996
88. D. O. Ali, M. H. Arbab, N. M. Admed, Molecular Detection of bla OXA-48 Gene Encoding Carbapenem Resistance Pseudomonas aeruginosa Clinical Isolates from Khartoum State Hospitals, Sudan, Gezira Journal of Health Sciences 2(16) (2021) 18-26, doi: http://journals.uofg.edu.sd/index.php/gjhs/article/view/1690
89. N. Hoque, et al, Genomic diversity and molecular epidemiology of a multidrug resistant Pseudomonas aeruginosa DMC30b isolated from hospitalized burn patient in Bangladesh, BioRxiv (2022), doi: https://doi.org/10.1101/2022.07.06.498939
90. A. Newbury, et al, Fitness effects of plasmids shape the structure of bacteria-plasmid interaction networks, Proc Natl Acad Sci U S A. 119(22) (2022) e2118361119, doi: 10.1073/pnas.2118361119
91. C. Ludden, A. G. Decano, et al, Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones, Microb Genom. 6(4) (2020) e000352, doi: 10.1099/mgen.0.000352
92. S. L. Foley, P. R. Kaldhone, S. C. Ricke, J. Han, Incompatibility Group I1 (IncI1) Plasmids: Their Genetics, Biology, and Public Health Relevance, Microbiol Mol Biol Rev. 85(2) (2021) e00031-20, doi: 10.1128/MMBR.00031-20
93. A. Carattoli, L. Villa, D. Fortini, A. García-Fernández, Contemporary IncI1 plasmids involved in the transmission and spread of antimicrobial resistance in *Enterobacteriaceae*, Plasmid 118 (2021) 102392. doi: 10.1016/j.plasmid.2018.12.001
94. A. Siryaporn, et al, Surface attachment induces *Pseudomonas aeruginosa* virulence, Proc. Natl Acad. Sci. USA. 111 (2014) 16860–16865 .
95. A. Persat, et al, Type IV pili mechanochemically regulate virulence factors in *Pseudomonas aeruginosa*, Proc Natl Acad Sci U S A 112(24) (2015) 7563-8, doi: 10.1073/pnas.1502025112
96. M. Q. Carter, J. Chen, S. Lory, The *Pseudomonas aeruginosa* pathogenicity island PAPI-1 is transferred via a novel type IV pilus, J Bacteriol. 192(13) (2010) 3249-58, doi: 10.1128/JB.00041-10.
97. E. L. Fonseca, et al, Emergence of XDR high-risk *Pseudomonas aeruginosa* ST309 in South America: a global comparative genomic analysis, BioRxiv (2021) doi: https://doi.org/10.1101/2021.01.21.427610.
98. R. T. M. Poulter, J. Ho, T. Handley, G. Taiaroa, M. I. Butler, Comparison between complete genomes of an isolate of *Pseudomonas syringae* pv. actinidiae from Japan and a New Zealand isolate of the pandemic lineage, Sci Rep. 8(1) (2018) 10915. doi: 10.1038/s41598-018-29261-5
99. A. Lorenz, et al, Importance of flagella in acute and chronic *Pseudomonas aeruginosa* infections, Environ Microbiol. 21(3) (2019) 883-897, doi: 10.1111/1462-2920.14468.
100. D. Skurnik, et al, A comprehensive analysis of in vitro and in vivo genetic fitness of *Pseudomonas aeruginosa* using high-throughput sequencing of transposon libraries, PLoS Pathog. 9(9) (2013) e1003582, doi: 10.1371/journal.ppat.1003582.
101. C. S. Arlehamn, T. J. Evans, *Pseudomonas aeruginosa* pilin activates the inflammasome, Cell Microbiol. 13(3) (2011) 388-401, doi: 10.1111/j.1462-5822.2010.01541.x
102. S. Nagashima, et al, Complete genome sequence of phototrophic betaproteobacterium *Rubrivivax gelatinosus* IL144, J Bacteriol. 194(13) (2012) 3541-2, doi: 10.1128/JB.00511-12
103. Y. Xia, D. Gally, K. Forsman-Semb, B. E. Uhlin, Regulatory cross-talk between adhesin operons in *Escherichia coli*: inhibition of type 1 fimbriae expression by the PapB protein, Embo J 19 (2000) 1450–1457.
104. N. J. Holden, B. E. Uhlin, D. L. Gally, PapB paralogues and their effect on the phase variation of type 1 fimbriae in *Escherichia coli*, Mol Microbiol 42 (2001) 319–330.
105. N. J. Holden, D. L. Gally, Switches, cross-talk and memory in *Escherichia coli* adherence, J Med Microbiol 53 (2004) 585–593.
106. H. Zhou H, et al, Functions predict horizontal gene transfer and the emergence of antibiotic resistance, Sci Adv. 7(43) (2021) eabj5056, doi: 10.1126/sciadv.abj5056


3,303 samples had no *pil*-related PPIs
Get pEK204-related pil genes & PPIs
Quality control
1,060 samples had *pil*-related PPIs
4,363 samples with PPI data
4,445 potential samples from StringDB
PPI data inferred for pEK204, pEK499, pEK516 & pEC958 from StringDB
Get sample-plasmid pair's PPI & interacting host proteins
24 Enterobacterales PPI data: *Escherichia* / Shigella *(n=11)*, *Klebsiella (n=2)* & *Serratia* (n=11)
Quality control
27 *Enterobacterales* samples from StringDB
4 pOXA-48 isoforms
Check gene content
31 plasmids matching "pOXA-48"
Quality control
2,396 samples with pOXA-48-related PPI data
4,363 samples with PPI data
4,445 potential samples from StringDB
Get pOXA-48-related genes & PPIs
2,203 samples had no pOXA-48-related PPIs


A
Number of genes
0
10
20
24
22
9
5
3
3
3
2
1
12
pOXA_48_30715
18
pOXA_48_k8
37
pOXA_48a
57
pKP112
40
20
0
Number of genes
B
Number of genes
0
20
40
51
40
14
13
12
9
9
8
7
7
4
1
1
1
44
pEC958
55
pEK516
84
pEK499
86
pEK204
75
50
25
0
Number of genes


A
pOXA−48
Interactions
100 200 300 400 500
Excess fraction of PPIs per sample−pOXA−48 pair
pOXA−48−interacting genes per sample
B
pKP112
Interactions
200 400 600
Excess fraction of PPIs per sample−pKP112 pair
pKP112−interacting genes per sample
C
pOXA−48_30715
Interactions
50 100 150
Excess fraction of PPIs per sample−pOXA−48_30715 pair
pOXA−48_30715−interacting genes per sample
D
pOXA−48_k8
Interactions
50 100 150 200
Excess fraction of PPIs per sample−pOXA−48_k8 pair
pOXA−48_k8−interacting genes per sample


A
pEK204
Interactions
500
1000
Excess fraction of PPIs per sample−pEK204 pair
pEK204−interacting genes per sample
B
pEK499
Interactions
200
400
600
Excess fraction of PPIs per sample−pEK499 pair
pEK499−interacting genes per sample
C
pEK516
Interactions
100
200
300
400
Excess fraction of PPIs per sample−pEK516 pair
pEK516−interacting genes per sample
D
pEC958
Interactions
100
200
300
400
500
Excess fraction of PPIs per sample−pEC958 pair
pEC958−interacting genes per sample
0.5
0.0
−0.5
0
200
400
600


Number of pEK204
–associated pil genes
12.5
10.0
7.5
5.0
2.5
0.0
More genes interacting
with pEK204 genes
Fewer genes interacting
with pEK204 genes
Relatively fewer
pEK204–related genes
Relatively more
pEK204–related genes
Cellvibrio_sp._BR_
Pseudomonas_aeruginosa_
Aromatoleum_aromaticum_EbN1_
Rubrivivax_gelatinosus_IL144_
Pseudomonas_fluorescens_F113_
Methylibium_petroleiphilum_PM1_
Janthinobacterium_lividum_
Stenotrophomonas_maltophilia_K279a_
Azoarcus_sp._BH72_
Shigella_flexneri_2a_str._301_
Escherichia_coli_CFT073_
Azoarcus_sp._KH32C_
Number of chromosomal genes interacting with pEK204 genes
0
200
400
600
–5.0
–2.5
0.0
2.5
Scaled Jaccard ratios of pEK204 vs pEK499