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

Research Article

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

[1] School of Biotechnology, Dublin City University, Dublin, Ireland.
[2] GAATI Lab, University of French Polynesia, Tahiti, French Polynesia. Orcid: 0000-0002-5534-2716
Corresponding author. Email: Tim.Downing@dcu.ie Orcid: 0000-0002-8385-6730

Running head: host-plasmid compatibility based on protein interactions

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

Abstract

The compatibility of plasmids in a new host cell is of broad significance, largely due to these plasmids' role in spreading antimicrobial resistance (AMR) and virulence factor genes. Such horizontal gene transfer can transform the host cell phenotypes, but only if the plasmid genes can be replicated, expressed and work synergistically in the new host cell. *In vitro* screening to evaluate this is laborious and can be informed by efficient computational analyses of plasmid-host compatibility through rates of protein-protein interactions (PPIs) between plasmid and host cell proteins. We identified instances of large excesses of such PPIs in eight plasmids of public health concern, including pOXA-48, using the PPI data of samples representative of most known bacteria (n=4,363). This identified 23 species with high rates of PPIs with four *bla*OXA-48-positive plasmids, such as *Pseudomonas aeruginosa.* This approach also identified 48 species with high compatibility with four plasmids common in *Escherichia coli* infections. We found a strong association of one plasmid with the fimbrial adhesin operon *pil*, which could enhance host cell adhesion in water environments. Our work on eight plasmids demonstrated that an excess rate of PPIs could be a sign of host-plasmid compatibility, which is important for AMR control because plasmids like pOXA-48 can jump between species with apparent ease. This compatibility could also be a sign of long-term co-evolution and therefore compensation by the host cell, implying the plasmids could be retained even in the absence of antibiotic pressure.

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 mean the role of plasmids in enabling adaptation to new environments is a key concern to address the spread of infectious bacteria. 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. This implies that changing the assortment of host genes and proteins interacting with plasmid proteins could improve host fitness in relevant environments.

Plasmids' capacity to interact successfully with the host cell is important and stems from their genetic profiles more so than the host genome [4,9]. Plasmid-host protein-protein interactions (PPIs) must be synergistic for plasmids to persist because plasmid-encoded genes must be replicated and expressed in the host cell. However, plasmids typically induce a small fitness cost in their host cells [10,11 and thus possess genes for post-segregational killing and stable inheritance to parasitise the host cell and ensure their propagation in daughter cells [12], facilitating the vertical plasmid transmission.

Plasmid-encoded AMR genes help drive the transmission of carbapenemase-producing *Enterobacteriaceae* (CPE) including *Escherichia coli* and *Klebsiella,* and are a key concern because these genes 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 was encoded by *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 a diverse assortment of host species, most typically in the *Enterobacterales* order. Host-pOXA-48a co-evolution has occurred, such as 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 [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 similar 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]. Such plasmids encoding such fimbrial adhesins could allow bacterial cells to survive in new environments because adherence to human epithelial cells helps bacterial colonisation and in forming 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], suggesting 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 [41,54] that can elevate conjugation rates in liquid environments [55]. 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 - it was present in ~11% of all *E. coli* ST131 from predominantly human sources where 27% of these lacked a *pap* operon (P fimbriae) but had a complete *pil* operon, implying a role for T4P in human infection [56].

Limited work has examined potential plasmid-host co-evolutionary information embedded in PPIs. Here, we examined interactions between plasmid proteins and bacterial host proteins to infer 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 24 *Enterbacterales* genomes, and then more widely 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 these may stem from operons associated with environments where those bacterial samples have been found.

Methods

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 v4.0.1 [58], RStudio v2022.2.3.492 [59], Bioconductor v3.11 [60] and R packages Rentrez v1.2.2 [61], STRINGdb v2.0.1 [62], and StringR v1.4.0 [63] (Figure 1). 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. The second dataset was composed of 4,363 StringDB bacterial samples with valid PPI data as outlined previously (Downing & Rahm 2022) (Figure 1). Using a score threshold of 400, the number of proteins and their PPIs were recorded. 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].

Plasmid data retrieval and quality control

The IncL pOXA-48a plasmid has been strongly associated with *E. coli* and *K. pneumoniae* [28]. To explore it further, we downloaded accession data for all (n=31) NCBI NucCore entries containing "pOXA-48" that came from the three *Enterobacterales* species: *E. coli* (n=7), *K. pneumoniae* (n=22), and *Serratia marcescens* (n=2). The protein similarity per plasmid indicated that all four genetically distinct pOXA-48 isoforms were from *K. pneumoniae* (strains named Kp11978, KP112, Kpn-30715/15 and K8) (Table 1). Plasmid pKP112 possessed a bla*OXA-48* gene and so was considered a valid pOXA-48 plasmid here. We 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].

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 the bitscore value from 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 were recorded and analysing using R package UpSetR [77]. All p values reported in this study were corrected for multiple testing using the Benjamini-Hochberg method in R.

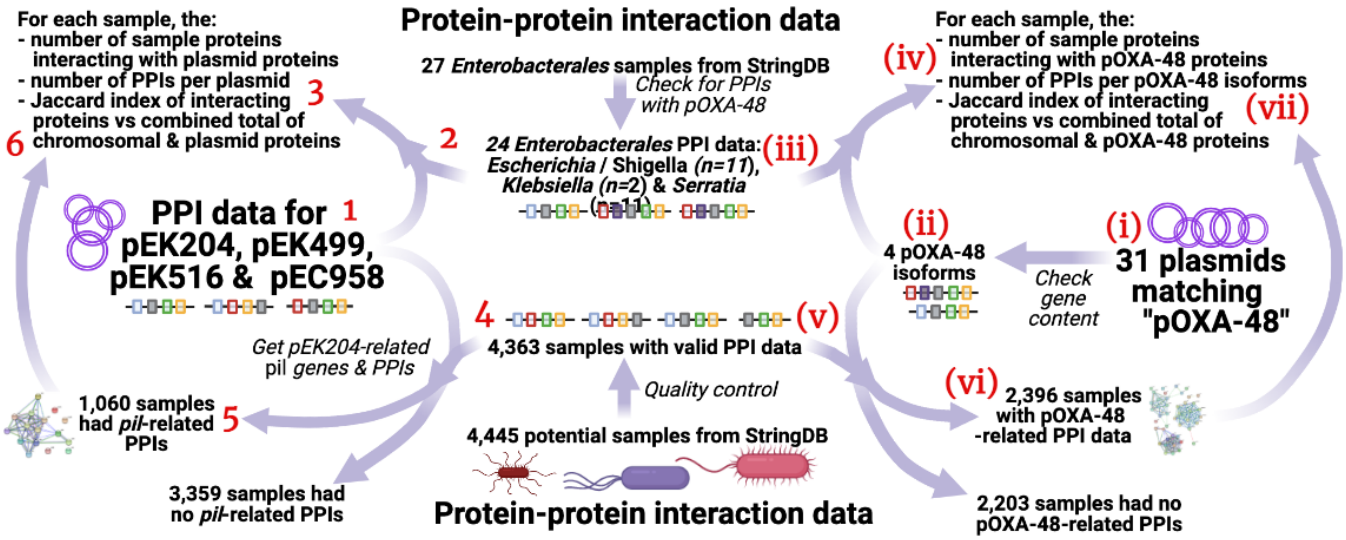


Fig 1. The data sources and methods summary. We used PPI data from StringDB related to two datasets: one with 24 *Enterobacterales* (middle, top) and another with 4,363 diverse bacterial samples (middle, bottom). Left: We compared the interacting host proteins and associated PPIs for (1) four plasmids pEK204, pEK499, pEK516 and pEC958 compared to the (2) 24 *Enterobacterales* samples' PPI data to (3) assess their potential (in)combability. These plasmids were compared to the (4) 4,363 samples' PPI data to get (5) their pEK204-related *pil* genes and again (6) quantify their potential (in)combability. Right panel: We examined (i) 31 initial pOXA-48-related plasmids to (ii) identify four main isoforms whose interacting host proteins and associated PPIs we compared to the (iii) 24 *Enterobacterales'* PPI data to (iv) examine their potential (in)combability. Likewise, these four pOXA-48 isoforms were compared to the (v) 4,363 samples' PPI data to get (vi) their pOXA-48-related genes and (vii) computed their potential (in)combability. Created with BioRender.com.

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*.

Comparing pOXA-48-related plasmids and quality control

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. There were 6,181 sample-plasmid pairs in total with more than 10 interacting proteins and more than 10 PPIs, coming from: pOXA-48 (1,820 pairs), pKP112 (1,868), pOXA-48_30715 (673) and pOXA-48_k8 (1,820). The number of interacting host proteins per sample was only slightly associated with the numbers of genes per sample (r^2=0.017) and per plasmid (r^2=0.005). To quantify instances of high host-plasmid PPI rates, we first examined on the association of the number of interacting host proteins per sample with the number of sample-plasmid PPIs. We

evaluated this using a linear model predicting the number of PPIs per combination in these 6,181 pairs based on the numbers of interacting host proteins, plasmid proteins, proteins per sample, and associated pairwise interactions. This showed that the number of interacting host proteins per sample had the largest effect (p=4.6e-15), followed by the pre-existing number of PPIs per sample (p=1.9e-14), then the interaction between the number of interacting host proteins with the number of plasmid proteins (p=7.6e-9), and lastly the number of plasmid proteins (p=0.0123, Table S3). The number of proteins per sample had no detectable effects, nor did its interactions with the numbers plasmid proteins and interacting host proteins. The same model applied to each plasmid showed that the number of interacting proteins (2.3e-15 ≤ p ≤ 4.6e-15, 0.960 ≤ r ≤ 0.995) and the baseline number of PPIs per sample (2.3e-15 ≤ p ≤ 0.001) were useful predictors, whereas the number of proteins per sample and its interaction with the number of interacting host proteins were not – except for pOXA-48_k8 where the latter two variable had slight effects (Table S3). Consequently, we focused on the number of sample proteins and PPIs linked to each plasmid. For each individual plasmid we used a linear model of the number of PPIs per sample based on the number of interacting host proteins to get the residual numbers of PPIs per sample, reflecting sample-plasmid pairs whether the association between the numbers of interacting host proteins and PPIs deviated from the expected rate. To obtain the expected number of PPIs per sample-plasmid pair, we used a model based on the number of interacting sample proteins, number of sample genes, number of specific plasmid genes, and their interactions. The scaled difference between the observed and expected numbers of PPIs per sample-plasmid pair was calculated as a fraction where the Z-normalised ratio suggested a difference of ±0.2 (or ±20%) was equivalent to ±2, such that the average excess PPIs was zero. Deficits (difference < -0.2) in the observed PPI numbers were not explored here.

Escherichia coli plasmid comparison and quality control

For the four *E. coli* plasmids (pEK204, pEK499, pEK516 and pEC958), all 24 *Enterobacterales* and 1,604 other samples had at least one PPI with at least one of these plasmids. There were 3,828 sample-plasmid combinations with over 10 interacting host proteins and more than 10 PPIs across these samples, coming from: pEK204 (856 combinations), pEK499 (1,324), pEK516 (852) and pEC958 (796). To quantify instances of high host-plasmid PPI rates as above, we found that the number of interacting host proteins per sample was dependent on the number of proteins per sample (p=4e-16) but much less so on the number of proteins per plasmid (p=0.07). As above, we assessed on the association of the number of interacting host proteins per sample with the number of sample-plasmid PPIs. A linear model predicting the number of PPIs per pair in these 3,828 plasmid-sample combinations showed that the number of interacting proteins per sample had the largest effect (p=4.6e-15), followed by the interaction between the interacting host protein and plasmid protein numbers (7.1e-12), and the number of plasmid proteins (p=2.4e-9, Table S3). The pre-existing number of PPIs per sample, and the number of proteins per sample had no detectable effects, nor did the latter's interactions with the numbers plasmid proteins and interacting proteins. The same model applied to each plasmid showed that the number of interacting host proteins (p=4.6e-15 for all, 0.84 ≤ r ≤ 0.98) was a useful predictor, whereas the number of proteins per sample and its interaction with the number of interacting host proteins were not, and the baseline number of PPIs per sample was useful for pEK204, pEK499 and pEK516 (8.3e-8 ≤ p ≤ 2.6e-2, Table S3). The expected number of PPIs and scaled difference between the observed and expected numbers of PPIs per sample-plasmid pair were calculated as above.

Interactions between fimbrial adhesin operons

We focused further on the *pil* operon associated with adhesion encoded by pEK204, which has 14 genes *pilI-V* [37]. We used pEK499 as a baseline comparator because it had the largest number of samples with which it had PPIs (1,324 out of 4,363). We found 1,060 samples had detected PPIs with both pEK204 and pEK499 and these were examined using a normalised log2-scaled ratio of the Jaccard indices of pEK204 versus pEK499. The pairwise Pearson correlation between 14 *pil* genes was determined using ppcor v1.1 [78] based on presence-absence data for 4,377 samples from across bacteria [79]. Genes with multiple synonyms were collapsed into a single representative. These 14 *pil*

239 genes were also compared to 14 *fim* genes and four *pap* genes in the same manner. Next, we aggregated
240 PPI data across the 4,363 samples to search for *pil, fim* and *pap* gene synonyms and then for Pil, Fim
241 and Pap PPIs. The gene name synonyms included: F7-2 as PapA and FooG as PapG in *E. coli* CFT073;
242 and FooB as PapB in 536; and the Pap proteins for *E. coli* 536 were: ECP_4533 (PapG), ECP_4536
243 (PapK), ECP_4538 (PapD), ECP_4539 (PapC), ECP_4540 (PapH), ECP_4541 (PapA) and ECP_4532
244 (PapE/PapF); and for *E. coli* K12 MG1655, atpG was PapC, atpA was PapA, aptH was PapE, aptC was
245 PapG, aptE was PapH, atpF was PapF, atpB was PapD and atpD was PapB. Other synonyms were the
246 normal name followed by an underscore and/or a digit.

Results

To test for compatibility and signals of co-evolution between plasmids with different hosts, 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 was 24 *Escherichia*, *Klebsiella*, *Serratia* and *Shigella* samples to examine genus-level plasmid links, and the second collection was 4,363 samples from across bacteria to inform on wider patterns.

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

We examined the level of PPIs 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]. This implied that these plasmids' protein products may interact with proteins encoded by accessory genes, which are fluid in these species. This heterogeneity can also be explained by the observation that these plasmids shared just four unique genes including *ssb*, *pemI/K* and *bla*OXA-48 (Figure 2A) (*bla*OXA-48 was not annotated properly 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 genetic differences suggested that the plasmids' effects on host cell PPI networks could be divergent as a consequence.

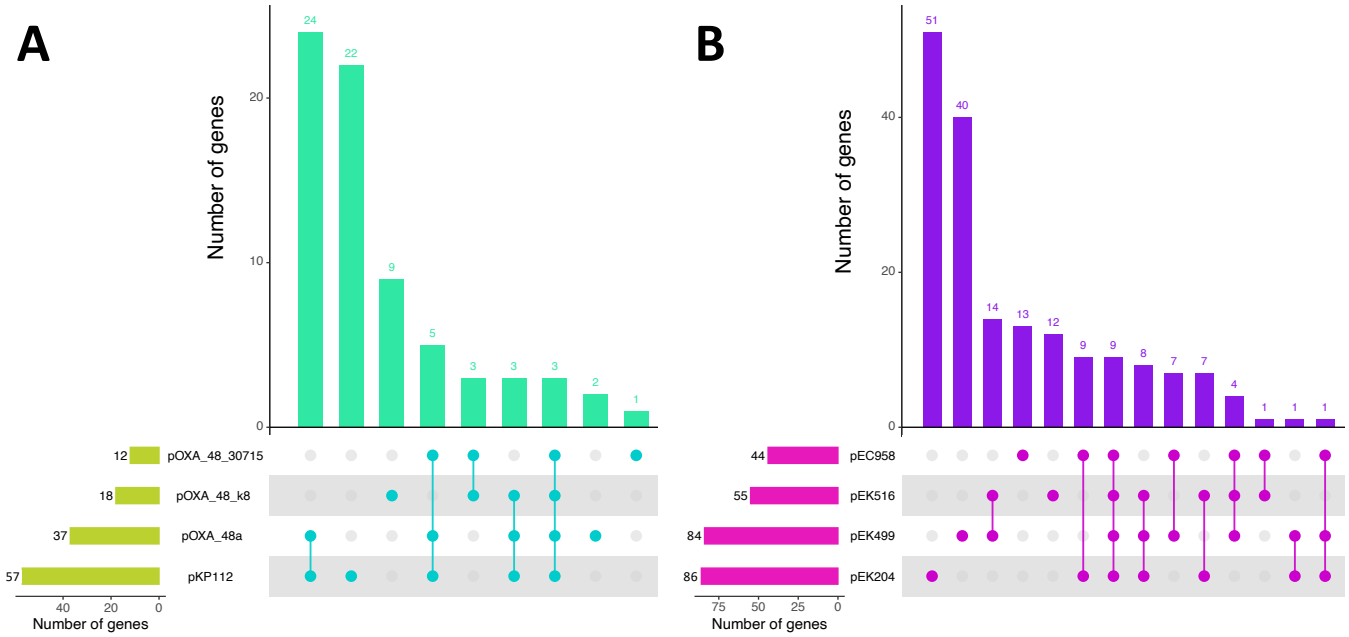


Fig 2. (A) The overlap of unique annotated genes between the four pOXA-48-related plasmids (pOXA-48, pKP112, pOXA-48_30715, pOXA-48_k8). Top: The intersection sizes (y-axis) and the genes per set (green) showed that only four genes were shared by all four plasmids: *ssb*, *pemI/K* and *bla*OXA-48. Bottom: 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 only 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).

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

Next, we assessed the level of PPIs between four major plasmids (pEK204, pEK499, pEK516, pEC958, Table 1) and the same 24 *Enterobacterales* samples because these plasmids possess numerous AMR genes and are prevalent, particularly in certain *E. coli* phylogroups. As above, the numbers of interacting host proteins per sample and PPIs had no clear positive association (Table S1). Even though the levels of sample-plasmid association were affected by the number of unique genes per plasmid, there were elevated rates of interacting proteins and PPIs for nine of ten *Escherichia* samples, whereas *Klebsiella* and *Serratia* were more heterogeneous. Consequently, the plasmids' genetic differences appeared to be the major determinants of this, like pOXA-48. The gene content of pEK204, pEK499, pEK516 and pEC958 showed nine shared genes: *ssb*, *psiB* and *traA*/*E*/*L*/*K*/*R*/*V*/*Y* (Figure 2B). The *tra* region is associated with conjugation, *psiB* encodes a plasmid SOS inhibition protein that stops induction of the SOS pathway by binding RecA [82], and *ssb* encodes a ssDNA-binding protein involved in DNA replication, recombination and repair [83]. As anticipated, pEK204 (with 51 unique genes) was the most divergent sample in this group.

Varied patterns of pOXA-48 interactions across bacteria

To shed light on these complex patterns, we extended our analysis of the four pOXA-48 plasmids to all bacterial samples with valid PPI data from StringDB (n=4,363), focusing on 6,181 sample-plasmid combinations with more than 10 interacting host proteins and over 10 PPIs. Plasmid pOXA-48_30715 had the most interacting host proteins (53±23, median±SD) and PPIs (53±26), followed by pKP112 (interacting proteins, 39±45; PPIs, 43±65), pOXA-48 (interacting proteins, 35±45; PPIs, 37±58), and then pOXA-48_k8 (interacting proteins, 30±33; PPIs, 31±37) (Figure S1). Although the number of proteins per plasmid and per sample had small effects on the number of plasmid-host PPIs, the interacting host proteins per plasmid had the biggest effect ($0.957 \leq r \leq 0.995$, Figure S2).

Only 0.6% (40 out of 6,401) of sample-plasmid pairs had high levels of association, measured as an excess of plasmid-linked PPIs of over 100, more than 100 interacting host proteins and at least a 20% excess PPIs after using a model to correct for the number of interacting host proteins per sample-plasmid pair (Figure S3, Table S4). 22 of these 40 were linked to pKP112, 13 to pOXA-48, three to pOXA-48_k8 and two to pOXA-48_30715 (Figure 3). *P. aeruginosa* had the most interacting host proteins of these 40: it shared 19 proteins on pKP112 (Figure S4) that had PPIs with 450 other *P. aeruginosa* proteins (Figure S5) with 684 PPIs, an excess of 27%. *P. aeruginosa* also had >300 interacting host proteins, >340 PPIs and an excess of PPIs for pOXA-48. The sample with the 2nd highest number of interacting host proteins was *Aromatoleum aromaticum* EbN1, which had 32 proteins on pOXA-48 (Figure S6) interacting with 252 *A. aromaticum* proteins with 400 PPIs, an excess of 34% (Figure S7). This sample also had 297 interacting proteins linked to pKP112 with 452 PPIs, an excess of 28%. *Enterobacter cloacae subsp. cloacae* ATCC 13047 had 125 other host proteins interacting with pKP112 (Figure S8) with 173 PPIs (an excess of 20%, Figure S9). Two *Legionella* and a *Citrobacter* species also had high associations with pKP112 and pOXA-48. *Klebsiella oxytoca* had 173 interacting host proteins with pKP112 that had 221 PPIs (an excess of 9%), but no such associated with other three plasmids.

Varied protein-level interactions of bacteria with pEK204, pEK499, pEK516 and pEC958

Assessing pEK204, pEK499, pEK516 and pEC958 in the same manner with the larger collection of 4,363 samples identified 1,604 samples with at least one PPI with at least one of these plasmids. We focused further on 3,828 sample-plasmid pairs with more than 10 interacting host proteins and over 10 PPIs in 1,410 samples, of which 48.4% (683 out of 1,410) had PPIs with all four plasmids. The number of interacting host proteins per sample was comparable across all four plasmids, and highest for pEK204 (median of 64 vs 54 for the others, Figure S10). The distributions of numbers of PPIs per sample were similar, but with a much higher variance for pEK204 relative to the other three (13.8 vs 8.2), highlighting that samples had elevated PPI rates (Figure S11).

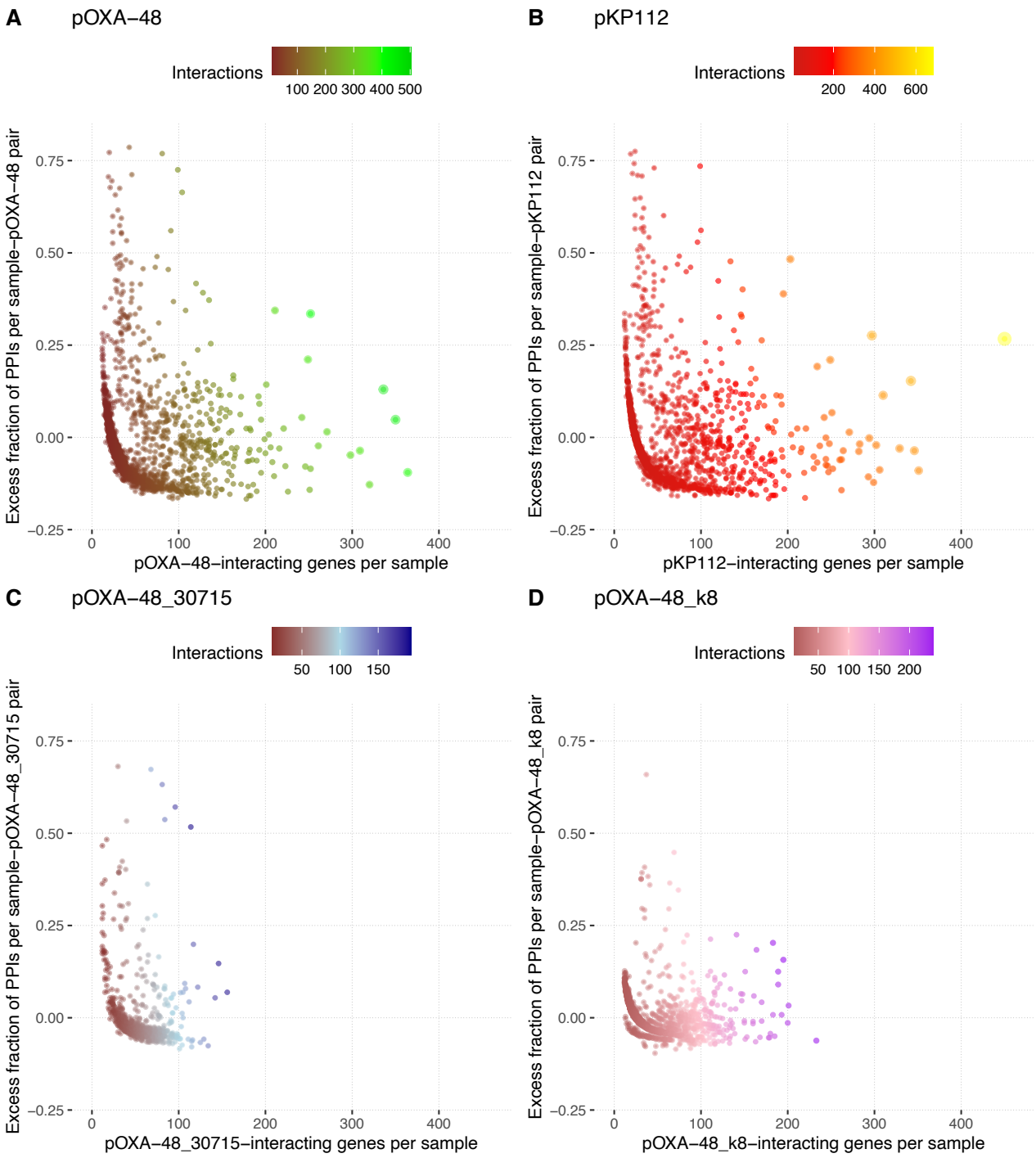


Fig 3. The numbers of interacting host proteins per sample (x-axis) versus the excess fraction of PPIs (y-axis) for each sample-plasmid pair for the four pOXA-48 isoforms (A) pOXA-48 (red-green), (B) pKP112 (red-yellow), (C) pOXA-48_30715 (brown-blue) and (D) pOXA-48_k8 (mauve-purple) using data from 6,181 samples' plasmids with over 10 interacting host proteins and at least 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).

Here, 1.5% (56 out of 3,828) of sample-plasmid pairs had strong associations, measured as an excess of plasmid-linked PPIs of over 300, more than 100 interacting host proteins and an excess of PPIs of at least 20% (Table S5). 48 were associated with pEK204, five with pEC958, two with pEK499, and one with pEK516 (Figure 4). In particular, the host-pEK204 PPIs per pair varied much more than the other plasmids (Figure S12). Nine of these 56 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%. Like for 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%).

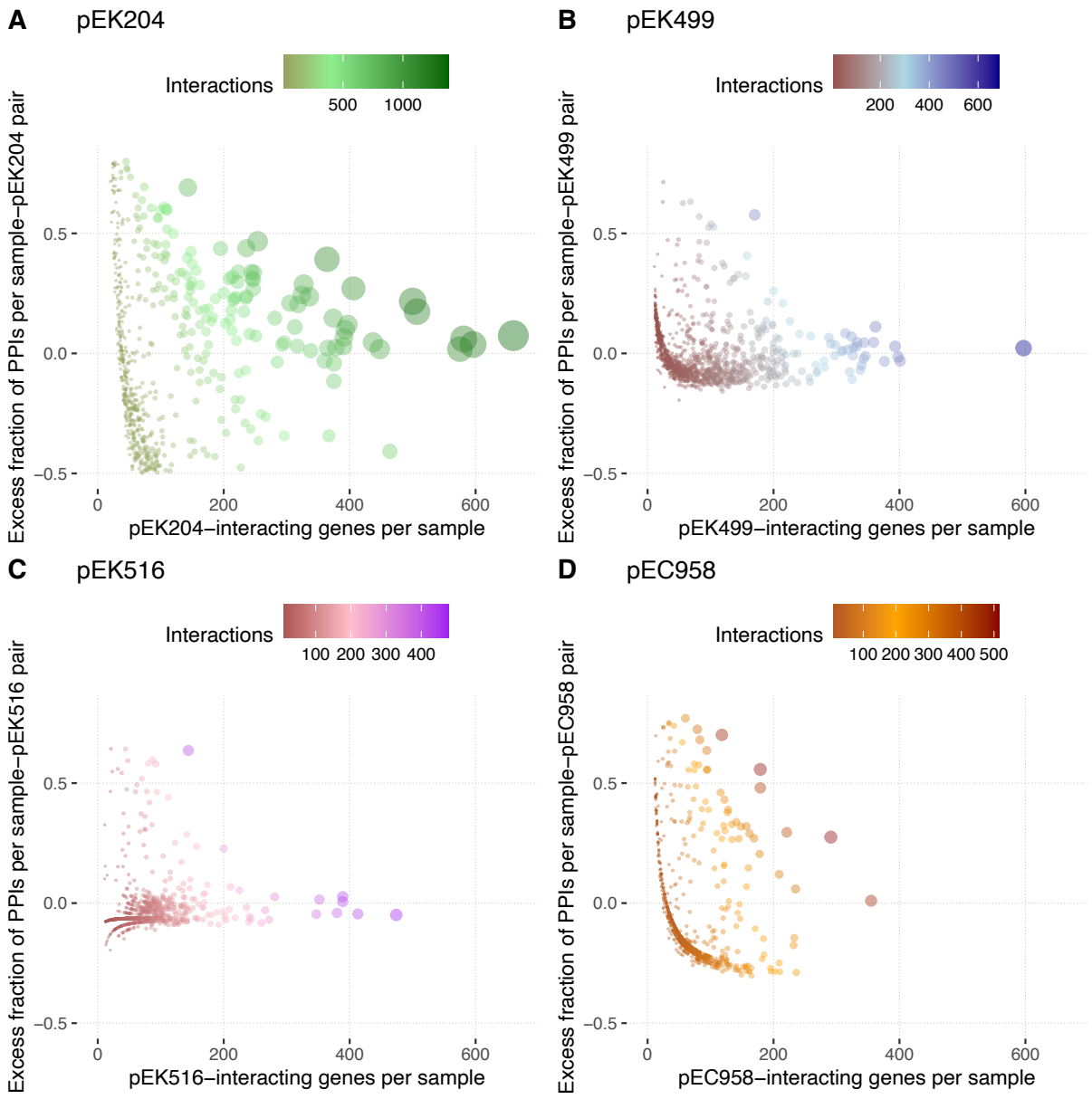


Fig 4. The numbers of interacting host proteins per sample (x-axis) versus the excess fraction of PPIs (y-axis) for each sample-plasmid pair for plasmids (A) pEK204 (brown-green), (B) pEK499 (brown-blue), (C) pEK516 (red-mauve) and (D) pEC958 (orange-brown) using data from 3,828 samples' plasmids with 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), the five samples with the most interacting proteins per sample with pEK204 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).

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

We examined conjugative plasmid pEK204 further given its elevated association with certain hosts, focusing on the *pil* operon because the *tra* region is well-studied. 316 samples out of 1,060 with PPIs for both pEK204 and pEK499 had at least one 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 compared scaled the Jaccard index of pEK204 by pEK499 as a baseline given the similar numbers of unique genes per plasmid (Table 1). This log2-scaled ratio of the Jaccard indices was positively correlated with the number of interacting host proteins for pEK204 and so pointed to samples with a higher potential for pEK204-specific PPIs, including several *Pseudomonas* samples like *P. aeruginosa* and *P. fluorescens* F113 as above (Figure 5, Table S6). Some samples had 12 or more *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).

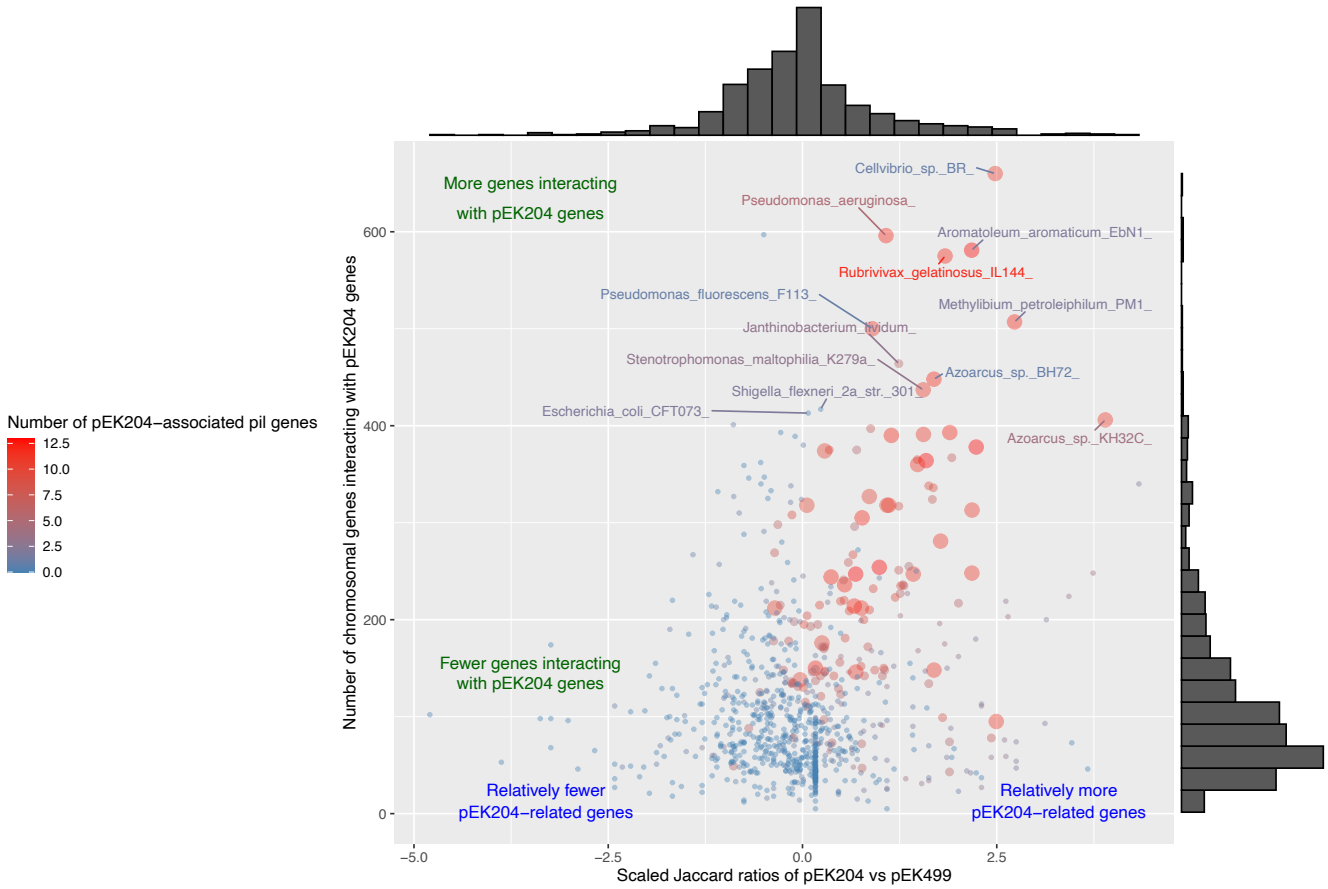


Fig 5. A plot of the normalised log2-scaled ratio of the Jaccard indices of pEK204 divided by that for pEK499 (x-axis) and the number of sample proteins interacting with pEK204 proteins (y-axis) for 1,060 bacterial samples, of which 316 had at least one pEK204-linked *pil* gene(s). The 14 *pil* genes on pEK204 were *pilIJKLMNOPQRSTUV*. The 12 samples labelled had at least 400 interacting host proteins and a log2-scaled ratio for pEK204 > 0. A larger point size and colour signifies the presence of more pEK204-related *pil* genes (where blue is fewer genes and red is more). The histogram indicates the marginal density of the scatterplot for both axes.

407 **Interactions of *pil* operon gene products with other fimbrial adhesin proteins**
408
409 The pairwise correlations of the 14 pEK204 *pil* genes' frequencies across the 4,363 bacterial samples
410 (Table S7) showed that most were positively or negatively correlated with one another, with the
411 exceptions of *pilL*, *pilQ* and *pilV*. Previous work has shown that PilK-PilV, but not PilI or PilJ, are
412 essential for R64 pilus generation [86]. Adding four *pap* genes and 14 *fim* genes to this assessment
413 identified five sets of genes, of which three involved *pil* genes. The first two sets were negatively
414 correlated within their own group, but were positively correlated with the other set (*pilO/P/R/T* and
415 *pilA/I/J/M/N/S/U/W*/V). The 3rd gene set was largely uncorrelated and stemmed from from FimT's (a
416 T4P biogenesis protein) and FimV's (a motility protein) interactions with PilQ (a ComQ homolog), PilL
417 and PilV (for FimT) and PilF/G/J (for FimV) in *P. aeruginosa* PAO1 and *Acinetobacter baumannii*
418 A118 [87]. Testing for PPIs between *pil, fim* and *pap* proteins in these 4,363 samples found that
419 *Ralstonia eutropha* H16 was the single sample with a Fim-Pil PPI across its 13 *pil* genes and four *fim*
420 genes (though major structural protein FimA with PilM-O and pilQ, Figure S22A). Notably, previous
421 work [37,56] has observed pEK204 in *E. coli*, and here both *E. coli* K12 MG1655 (Figure S22B)
422 CFT073 (Figure S22C) had Fim and Pap PPIs, suggesting scope for Pil proteins to interact or evolve
423 such PPIs.
424
425
426 **Discussion**
427
428 To evaluate potential compatibility and signals of co-evolution between diverse bacterial speciments and
429 plasmids, this study tested for an excess of host-plasmid PPIs using eight key plasmids in
430 representatives from most bacteria (n=4,363) by modelling data from StringDB. We initially
431 demonstrated that these plasmids had extensive heterogeneity in their PPI levels across 24
432 *Enterobacterales* (*Escherichia*, *Klebsiella*, *Serratia* and *Shigella*) samples, which were correlated with
433 the number of interacting host proteins. This absence of a genus-level association was consistent with
434 previous work [80].
435
436 To illuminate this opacity, we explored the four genetically distinct *bla*OXA-48-positive plasmids in a
437 larger set of 4,363 samples and found that these plasmids had at least one PPI with most bacteria,
438 allowing us to identify instances where there was a large excess of PPIs in 40 (0.6%) of informative
439 sample-plasmid pairs. This identified *P. aeruginosa, A. aromaticum* EbN1 and 21 other common animal
440 and plant pathogens as more likely to be compatible with at least one of these *bla*OXA-48-positive
441 plasmids, particularly with the pKP112 and pOXA-48 isoforms, suggesting those plasmids could
442 become common in these species. This is concerning because *bla*OXA-48 is the most common AMR
443 gene in CPE and is prevalent in *Pseudomonas* as well, based on carbapenemase gene rates in England
444 from 2020-22 (UK Health Security Agency 2022). Given their high compatibility with these plasmids
445 already, the high levels of carbapenems in clinical environments may allow these hosts to obtain, co-
446 evolve with and compensate for the metabolic cost of possessing these plasmids [88].
447
448 We used the same approach to assess the other group of four genetically related plasmids (pEK204,
449 pEK499, pEK516, pEC958) that likely originated in ExPEC [37]. ExPEC pose a significant public
450 health problem and change their plasmid composition to improve their fitness [89]. Like the pOXA-48-
451 related plasmids, these plasmids had at least one PPI with many bacteria, and we found large excess
452 rates of PPIs in 56 (1.5%) informative sample-plasmid pairs. 19% of the species (nine out of 48)
453 observed in the 56 sample-plasmid pairs were from the *Pseudomonas* genus, where *P. fluorescens* F113
454 and *P. aeruginosa* had the strong associations with pEK204. This was concerning because pEK204 has
455 spread across pathogenic bacteria [26], and IncI1 plasmids like pEK204 promote AMR due to the
456 presence of *bla*CTX-M, *bla*TEM, *bla*CMY and other ESBL genes [90-91].
457
458 We applied a novel approach of scaling the rate of pEK204-host PPIs by the level the pEK499-host
459 PPIs, which identified numerous samples, especially ones with high numbers of *pil* operon genes. The
460 diverse *pil* operon encodes a T4P with two main known isoforms [92] and had 14 genes on pEK204.

This T4P is virulence factor for *P. aeruginosa* [93-94], consistent with the compelling associations with *P. fluorescens* F113 and *P. aeruginosa* observed here. A ten-gene *pil* operon is involved in conjugation of pathogenicity island PAPI-1 in *P. aeruginosa* PA14 independently of pEK204 [93] that has been linked to *P. aeruginosa* ST309 infection outbreaks [95], and *pil* genes form part of an integrative conjugative element Pac_ICE1 in *P. syringae* pv actinidiae [96]. 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 [97] and mouse gut [98] infection, perhaps due to Pil proteins acting as cell surface antigens [99]. Consistent with this water link, 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 *pilI-V* gene rates in these 4,363 samples, suggesting that pEK204's *pil* repertoire differed from other *pil* operon compositions. We also found evidence of PPIs between *pil, fim* and *pap* gene products in *E. coli* and *R. eutropha.* This was important because fimbrial adhesins have key roles in binding host molecules and are common in bacterial pathogens, such as *fimH* found in >90% of CPEs globally [33]. 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 [101-103]. In this way, the gain of *pil* via pEK204 could facilitate superior host survival, conjugation or biofilm formation in water environments [41].

These results overcome a limitation that 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 helped to add more information on the potential compatibility of a plasmid and a particular species. 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 [104]. In addition, we did not map reads or align sequences here to detect genes, and relied on the accuracy of existing gene annotation tools. However, the resulting gene cluster alignment scores would result in a spectrum of ambiguous similarities that also lead to matches being unclear or imprecisely categorised. Future work is needed to improve genome and plasmid annotation, to explore the strength and specificity of PPIs, and to assess other highly transmissible plasmids.

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.

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.

References

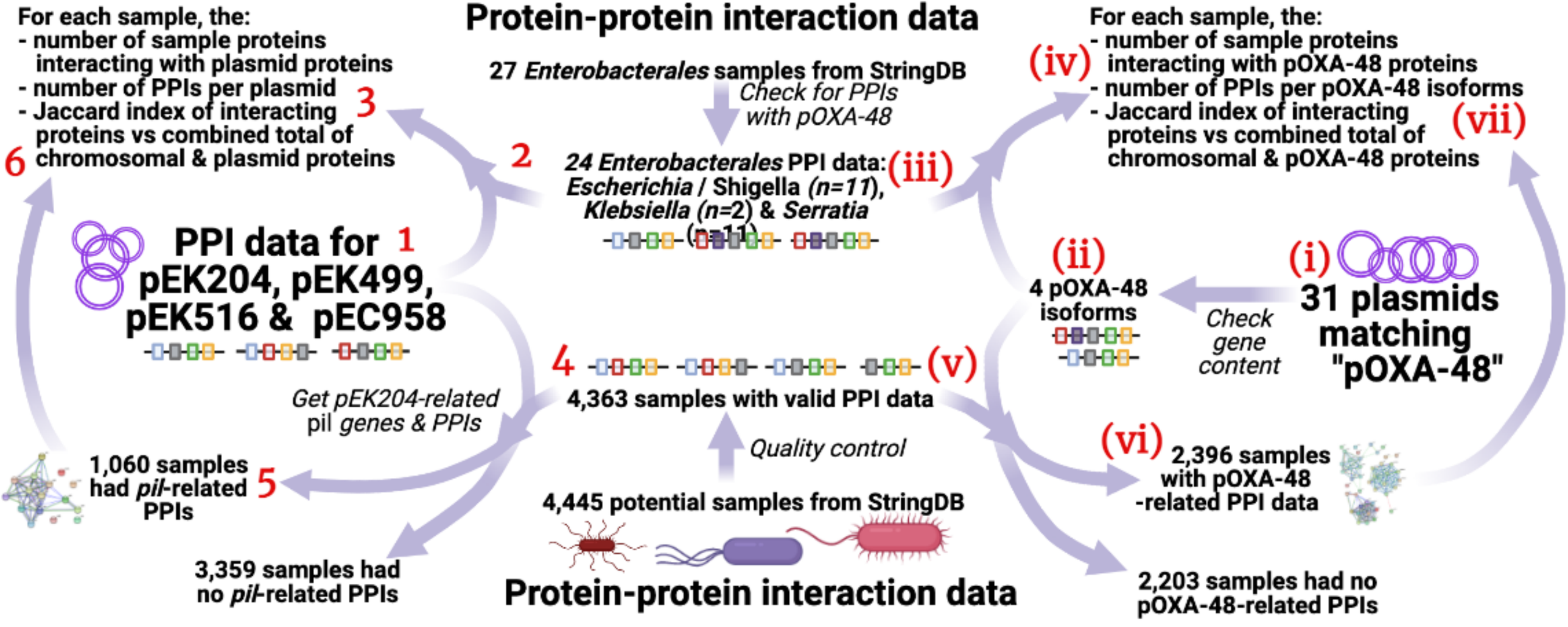
1. Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile genetic elements associated with antimicrobial resistance. Clin. Microbiol. Rev. 31, e00088-17 (2018).
2. San Millan, A. Evolution of plasmid-mediated antibiotic resistance in the clinical context. Trends Microbiol. 26, 978–985 (2018).
3. Che Y, et al. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. Proc Natl Acad Sci U S A. 2021 118(6):e2008731118. doi: 10.1073/pnas.2008731118
4. Smillie CS, et al. Ecology drives a global network of gene exchange connecting the human microbiome. Nature 480, 241–244 (2011).
5. Canosi U, Lüder G, Trautner TA. SPP1-mediated plasmid transduction. J Virol. 1982 44(2):431-6. doi: 10.1128/JVI.44.2.431-436.1982
6. Erdmann S, Tschitschko B, Zhong L, Raftery MJ, Cavicchioli R. A plasmid from an Antarctic haloarchaeon uses specialized membrane vesicles to disseminate and infect plasmid-free cells. Nat Microbiol. 2017 2(10):1446-1455. doi: 10.1038/s41564-017-0009-2
7. Zhang X, et al. Improvement in the efficiency of natural transformation of *Haemophilus parasuis* by shuttle-plasmid methylation. Plasmid. 2018 98:8-14. doi: 10.1016/j.plasmid.2018.07.001
8. Wein T, Dagan T. Plasmid evolution. Curr Biol. 2020 30(19):R1158-R1163. doi: 10.1016/j.cub.2020.07.003
9. Brito IL, et al. Mobile genes in the human microbiome are structured from global to individual scales. Nature. 535(7612):435-439. doi: 10.1038/nature18927
10. San Millan A, MacLean RC. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. Microbiol Spectr. 2017 5(5). doi: 10.1128/microbiolspec.MTBP-0016-2017
11. Hall JPJ, et al. Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation. PLoS Biol. 2021 19(10):e3001225. doi: 10.1371/journal.pbio.3001225
12. Nicolas-Chanoine MH, Bertrand X, Madec JY. *Escherichia coli* ST131, an intriguing clonal group. Clin Microbiol Rev 2014; 27:543–574
13. Tacconelli E, et al. 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. 2018 18(3):318-327. doi: 10.1016/S1473-3099(17)30753-3
14. Hopkins KL, 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. 2022 71(5). doi: 10.1099/jmm.0.001518
15. UK Health Security Agency. 2022. Quarterly laboratory surveillance of acquired carbapenemase-producing Gram-negative bacteria in England: October 2020 to March 2022. Health Protection Report 16(5).
16. Van der Bij AK, Peirano G, Pitondo-Silva A, Pitout JD. The presence of genes encoding for different virulence factors in clonally related *Escherichia coli* that produce CTX-Ms. Diagn Microbiol Infect Dis 2012; 72:297–302
17. Calhau V, Ribeiro G, Mendonça N, Da Silva GJ. Prevalent combination of virulence and plasmidic-encoded resistance in ST 131 *Escherichia coli* strains. Virulence 2013; 4:726–729
18. Poolman JT, Wacker M. Extraintestinal pathogenic *Escherichia coli*, a common human pathogen: challenges for vaccine development and progress in the field. J Infect Dis 2016; 213:6–13
19. Ben Zakour NL, et al. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. mBio 2016; 7:e00347
20. McNally A, et al. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PLoS Genet 2016; 12:1006280
21. Johnson JR, et al. Household Clustering of *Escherichia coli* Sequence Type 131 Clinical and Fecal Isolates According to Whole Genome Sequence Analysis. Open Forum Infect Dis 2016; 3:ofw129
22. Kallonen T, et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res 2017; 27:1437–1449

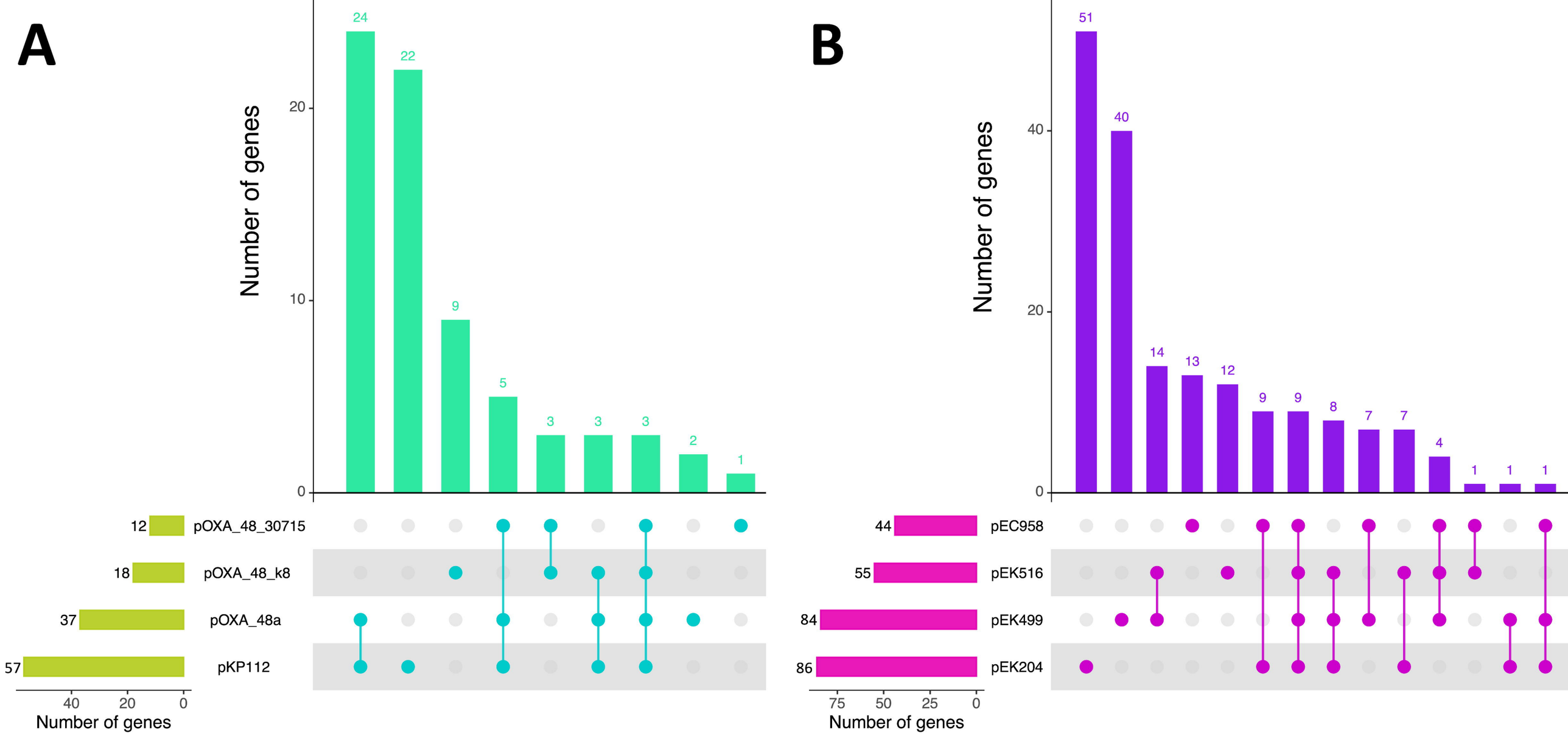
23. Goswami C, et al. Genetic analysis of invasive *Escherichia coli* in Scotland reveals determinants of healthcare-associated versus community-acquired infections. Microb Genom 2018; 4.
24. Decano AG, et al. Complete assembly of *Escherichia coli* ST131 genomes using long reads demonstrates antibiotic resistance gene variation within diverse plasmid and chromosomal contexts. mSphere 2019; 4:e00130–19
25. Ny S, Sandegren L, Salemi M, Giske CG. Genome and plasmid diversity of extended-spectrum β-lactamase-producing *Escherichia coli* ST131 – tracking phylogenetic trajectories with Bayesian inference. Sci Rep 2019; 9:10291
26. León-Sampedro R, et al. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. Nat Microbiol. 2021 6(5):606-616. doi: 10.1038/s41564-021-00879-y
27. Poirel L, Naas T, Nordmann P. 2010. Diversity, epidemiology, and genetics of class D beta-lactamases. Antimicrob Agents Chemother. 2010 54(1):24-38. doi: 10.1128/AAC.01512-08
28. Poirel L, Héritier C, Tolün V, Nordmann P. 2004. Emergence of oxacillinase-mediated resistance to imipenem in *Klebsiella pneumoniae*. Antimicrob Agents Chemother. 48(1):15-22. doi: 10.1128/AAC.48.1.15-22.2004
29. O'Brien DJ, et al. First isolation and outbreak of OXA-48-producing *Klebsiella pneumoniae* in an Irish hospital, March to June 2011. Euro Surveill. 2011 16(29):19921
30. Wrenn C, et al. Investigation of the first outbreak of OXA-48-producing *Klebsiella pneumoniae* in Ireland. J Hosp Infect. 2014 87(1):41-6. doi: 10.1016/j.jhin.2014.03.001
31. Morris D, et al. Dissemination of clonally related multidrug-resistant *Klebsiella pneumoniae* in Ireland. Epidemiol Infect. 2016 144(2):443-8. doi: 10.1017/S0950268815001041
32. Brehony C, et al. 2021. Molecular Epidemiology of an Extended Multiple-Species OXA-48 CPE Outbreak in a Hospital Ward in Ireland, 2018–2019. Antimicrobial Stewardship & Healthcare Epidemiology 1 (1). Cambridge University Press: e54. doi:10.1017/ash.2021.206
33. Johnston BD, et al. Global molecular epidemiology of carbapenem-resistant *Escherichia coli* (2002-2017). Eur J Clin Microbiol Infect Dis. 2021 doi: 10.1007/s10096-021-04310-6
34. Poirel L, Bonnin RA, Nordmann P. Genetic features of the widespread plasmid coding for the carbapenemase OXA-48. Antimicrob Agents Chemother. 2012 56(1):559-62. doi: 10.1128/AAC.05289-11
35. Downing T. Tackling Drug Resistant Infection Outbreaks of Global Pandemic *Escherichia coli* ST131 Using Evolutionary and Epidemiological Genomics. Microorganisms. 2015 3(2):236-67. doi: 10.3390/microorganisms3020236
36. Goswami C, et al. Origin, maintenance and spread of antibiotic resistance genes within plasmids and chromosomes of bloodstream isolates of *Escherichia coli*. Microb Genom. 2020 6(4):e000353. doi: 10.1099/mgen.0.000353
37. Woodford N, et al. Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516, encoding CTX-M enzymes in three major *Escherichia coli* lineages from the United Kingdom, all belonging to the international O25:H4-ST131 clone. Antimicrob Agents Chemother 2009; 53:4472–4482
38. Dhanji H, et al. Molecular epidemiology of fluoroquinolone-resistant ST131 *Escherichia coli* producing CTX-M extended-spectrum beta-lactamases in nursing homes in Belfast, UK. J Antimicrob Chemother. 2011 66(2):297-303. doi: 10.1093/jac/dkq463
39. Doumith M, Dhanji H, Ellington MJ, Hawkey P, Woodford N. Characterization of plasmids encoding extended-spectrum β-lactamases and their addiction systems circulating among *Escherichia coli* clinical isolates in the UK. J Antimicrob Chemother. 2012 67(4):878-85. doi: 10.1093/jac/dkr553
40. Decano AG, Downing T. An *Escherichia coli* ST131 pangenome atlas reveals population structure and evolution across 4,071 isolates. Sci Rep. 2019 9(1):17394. doi: 10.1038/s41598-019-54004-5
41. Dudley EG, et al. An IncI1 plasmid contributes to the adherence of the atypical enteroaggregative *Escherichia coli* strain C1096 to cultured cells and abiotic surfaces. Infect Immun 2006; 74:2102–211

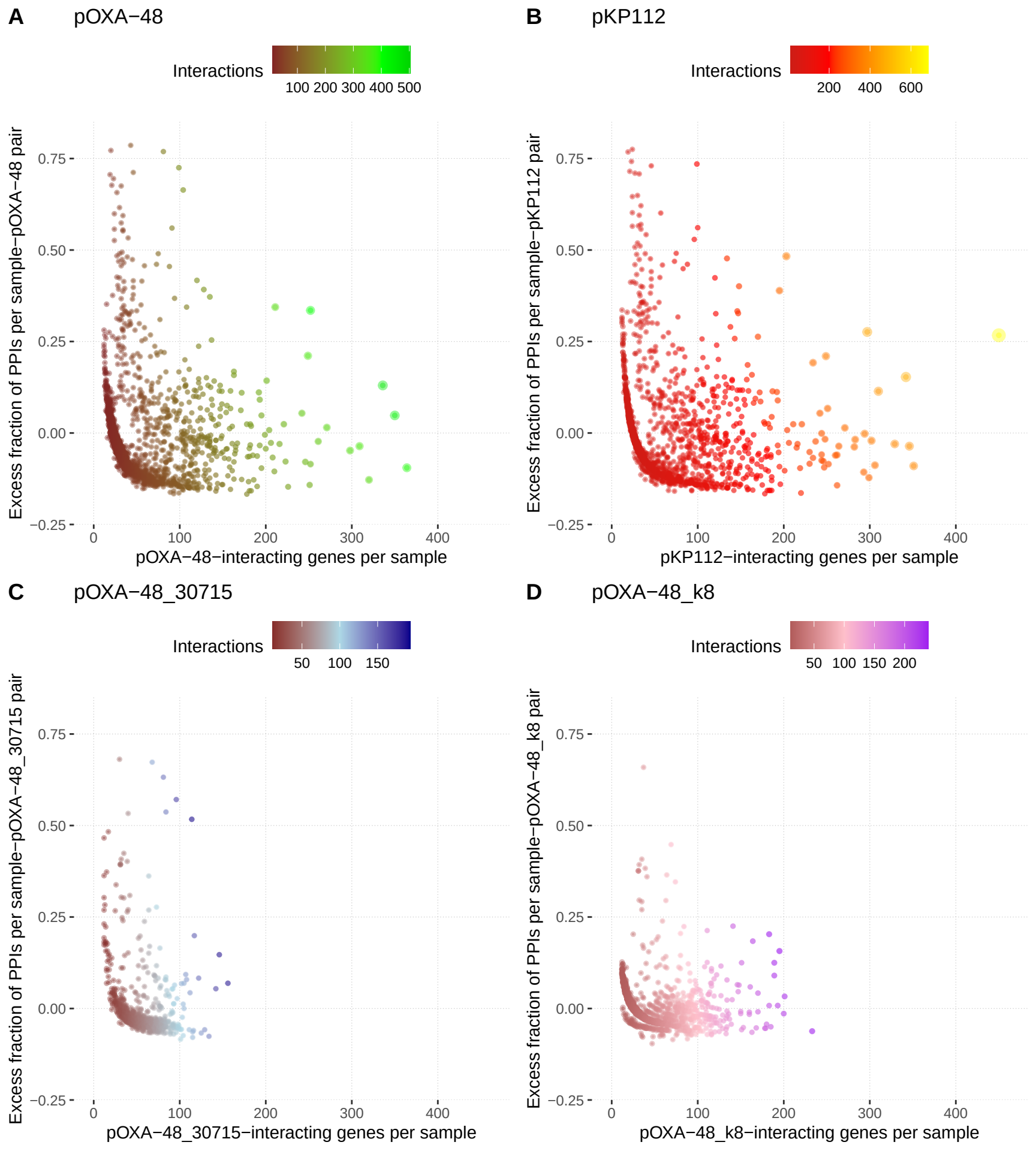
42. Sarkar S, Vagenas D, Schembri MA, Totsika M. Biofilm formation by multidrug resistant *Escherichia coli* ST131 is dependent on type 1 fimbriae and assay conditions. Pathog Dis 2016; 74:ftw013
43. Kuehn MJ, Heuser J, Normark S, Hultgren SJ. P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. Nature 1992; 356:252–255
44. Connell I, et al. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. Proc Natl Acad Sci U S A 1996; 93:9827–9832
45. Martinez JJ, Mulvey MA, Schilling JD, Pinkner JS, Hultgren SJ. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. Embo J 2000 19:2803–2812
46. Selvarangan R, et al. Interaction of Dr adhesin with collagen type IV is a critical step in *Escherichia coli* renal persistence. Infect Immun 2004 72:4827–4835
47. Nuccio SP, Baumler AJ. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. Microbiol Mol Biol Rev 2007 71(4):551-575
48. Werneburg GT & Thanassi DG (2018) Pili Assembled by the Chaperone/Usher Pathway in *Escherichia coli* and *Salmonella*. EcoSal Plus 8(1)
49. Wurpel DJ, Beatson SA, Totsika M, Petty NK & Schembri MA. Chaperone-usher fimbriae of Escherichia coli. PLoS One 2013 8(1):e52835
50. Craig L, Forest KT, Maier B. Type IV pili: dynamics, biophysics and functional consequences. Nat. Rev. Microbiol. 2019 17, 429–440
51. O'Toole GA, Kolter R. 1998. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. Mol Microbiol 30:295–304.
52. Comolli JC, et al. 1999. *Pseudomonas aeruginosa* gene products PilT and PilU are required for cytotoxicity in vitro and virulence in a mouse model of acute pneumonia. Infect Immun 67:3625–3630.
53. Schmidt AK, et al. A Filamentous Bacteriophage Protein Inhibits Type IV Pili To Prevent Superinfection of *Pseudomonas aeruginosa*. mBio. 2022 e0244121. doi: 10.1128/mbio.02441-21
54. García-Fernández A, et al. Multilocus sequence typing of IncI1 plasmids carrying extended-spectrum beta-lactamases in *Escherichia coli* and *Salmonella* of human and animal origin. J Antimicrob Chemother. 2008 61(6):1229-33. doi: 10.1093/jac/dkn131
55. Cottell JL, Webber MA, Piddock LJV. Persistence of transferable extended-spectrum-β-lactamase resistance in the absence of antibiotic pressure. Antimicrob. Agents Chemother. 2012 56(9): 4703-4706
56. Decano AG, et al. Plasmids shape the diverse accessory resistomes of *Escherichia coli* ST131. Access Microbiol. 2020 Nov 18;3(1):acmi000179. doi: 10.1099/acmi.0.000179
57. Szklarczyk D, et al. 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49(D1):D605-D612. doi: 10.1093/nar/gkaa1074
58. R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org
59. RStudio Team. 2022. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com
60. Ihaka R, Gentleman R. 1996. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics. 1996 (5)3:299-314
61. Winter DJ. 2017. rentrez: an R package for the NCBI eUtils API The R Journal 9(2):520-526
62. Szklarczyk D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research 2018 47.D1: D607-D613.
63. Wickham H. 2019. stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr
64. Wickham H, François R, Henry L, Müller K. 2022. dplyr: A Grammar of Data Manipulation. R package version 1.0.8. https://CRAN.R-project.org/package=dplyr
65. Wickham H. 2021. forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.1. https://CRAN.R-project.org/package=forcats
66. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

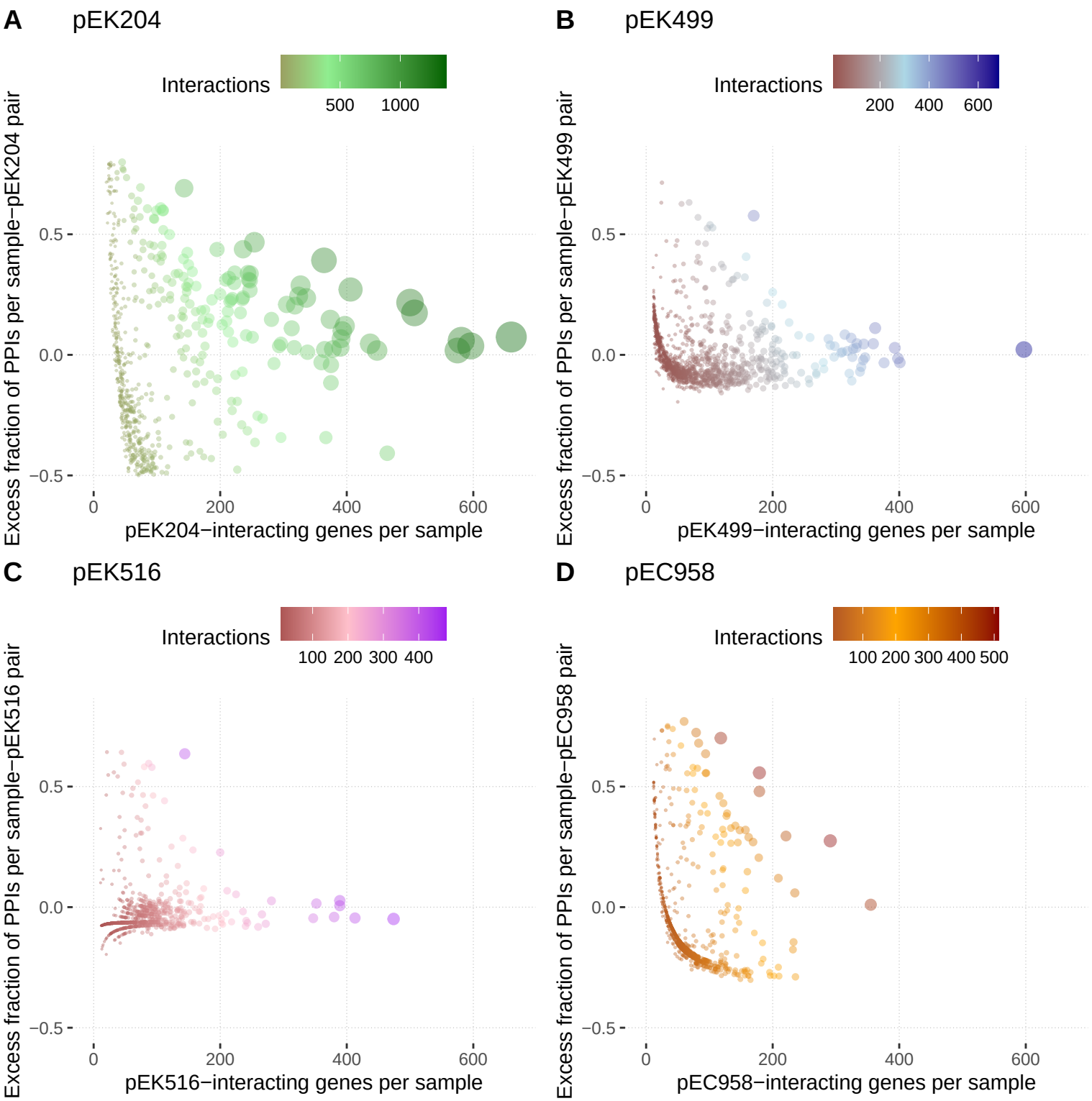
67. Slowikowski K. 2021. ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.9.1. https://CRAN.R-project.org/package=ggrepel
68. Wickham H, Hester J, Bryan J. 2022. readr: Read Rectangular Text Data. R package version 2.1.2. https://CRAN.R-project.org/package=readr
69. Wickham H, Bryan J. 2022. readxl: Read Excel Files. R package version 1.4.0. https://CRAN.R-project.org/package=readxl
70. Müller K, Wickham H. 2021. tibble: Simple Data Frames. R package version 3.1.6. https://CRAN.R-project.org/package=tibble
71. Wickham H, Girlich M. 2022. tidyr: Tidy Messy Data. R package version 1.2.0. https://CRAN.R-project.org/package=tidyr
72. Wickham H, et al. 2019. Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
73. Chen H. 2022. VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.7.3. https://CRAN.R-project.org/package=VennDiagram
74. Totsika M, 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):e26578. doi: 10.1371/journal.pone.0026578
75. Pagès H, Aboyoun P, Gentleman R, DebRoy S. 2021. Biostrings: Efficient manipulation of biological strings. R package version 2.60.2. https://bioconductor.org/packages/Biostrings
76. Becker G, Lawrence M. 2022. genbankr: Parsing GenBank files into semantically useful objects. R package version 1.24.0.
77. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017 33(18):2938-2940. doi: 10.1093/bioinformatics/btx364
78. Kim, S. 2015. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. Communications for Statistical Applications and Methods, 22(6), 665-674.
79. Downing T, Rahm A. 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. Alonso-del Valle, A. et al. Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities. Nat Commun. 2021 12(1):2653. doi: 10.1038/s41467-021-22849-y.
81. Tsuchimoto S, Ohtsubo H, Ohtsubo E. Two genes, pemK and pemI, responsible for stable maintenance of resistance plasmid R100. J Bacteriol. 1988 170(4):1461-6. doi: 10.1128/jb.170.4.1461-1466.1988
82. Bailone A, et al. PsiB polypeptide prevents activation of RecA protein in *Escherichia coli*. Mol Gen Genet. 1988 214(3):389-95. doi: 10.1007/BF00330471
83. Meyer RR, Laine PS. The single-stranded DNA-binding protein of *Escherichia coli*. Microbiol Rev. 1990 54(4):342-80. doi: 10.1128/mr.54.4.342-380.1990
84. Peix A, Ramírez-Bahena MH, Velázquez E. The current status on the taxonomy of *Pseudomonas* revisited: An update. Infect Genet Evol. 2018 57:106-116. doi: 10.1016/j.meegid.2017.10.026
85. Reimer LC, et al. BacDive in 2019: bacterial phenotypic data for High-throughput biodiversity analysis. Nucleic Acids Research; database issue 2019.
86. Yoshida T, Kim SR, Komano T. Twelve pil genes are required for biogenesis of the R64 thin pilus. J Bacteriol. 1999 181(7):2038-43. doi: 10.1128/JB.181.7.2038-2043.1999
87. Alm RA, Mattick JS. Identification of two genes with prepilin-like leader sequences involved in type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. J Bacteriol. 1996 178(13):3809-17. doi: 10.1128/jb.178.13.3809-3817.1996
88. Newbury A, et al. Fitness effects of plasmids shape the structure of bacteria-plasmid interaction networks. Proc Natl Acad Sci U S A. 2022 119(22):e2118361119. doi: 10.1073/pnas.2118361119
89. Ludden C, Decano AG, et al. Genomic surveillance of *Escherichia coli* ST131 identifies local expansion and serial replacement of subclones. Microb Genom. 2020 6(4):e000352. doi: 10.1099/mgen.0.000352
90. Foley SL, Kaldhone PR, Ricke SC, Han J. Incompatibility Group I1 (IncI1) Plasmids: Their Genetics, Biology, and Public Health Relevance. Microbiol Mol Biol Rev. 2021 85(2):e00031-20. doi: 10.1128/MMBR.00031-20

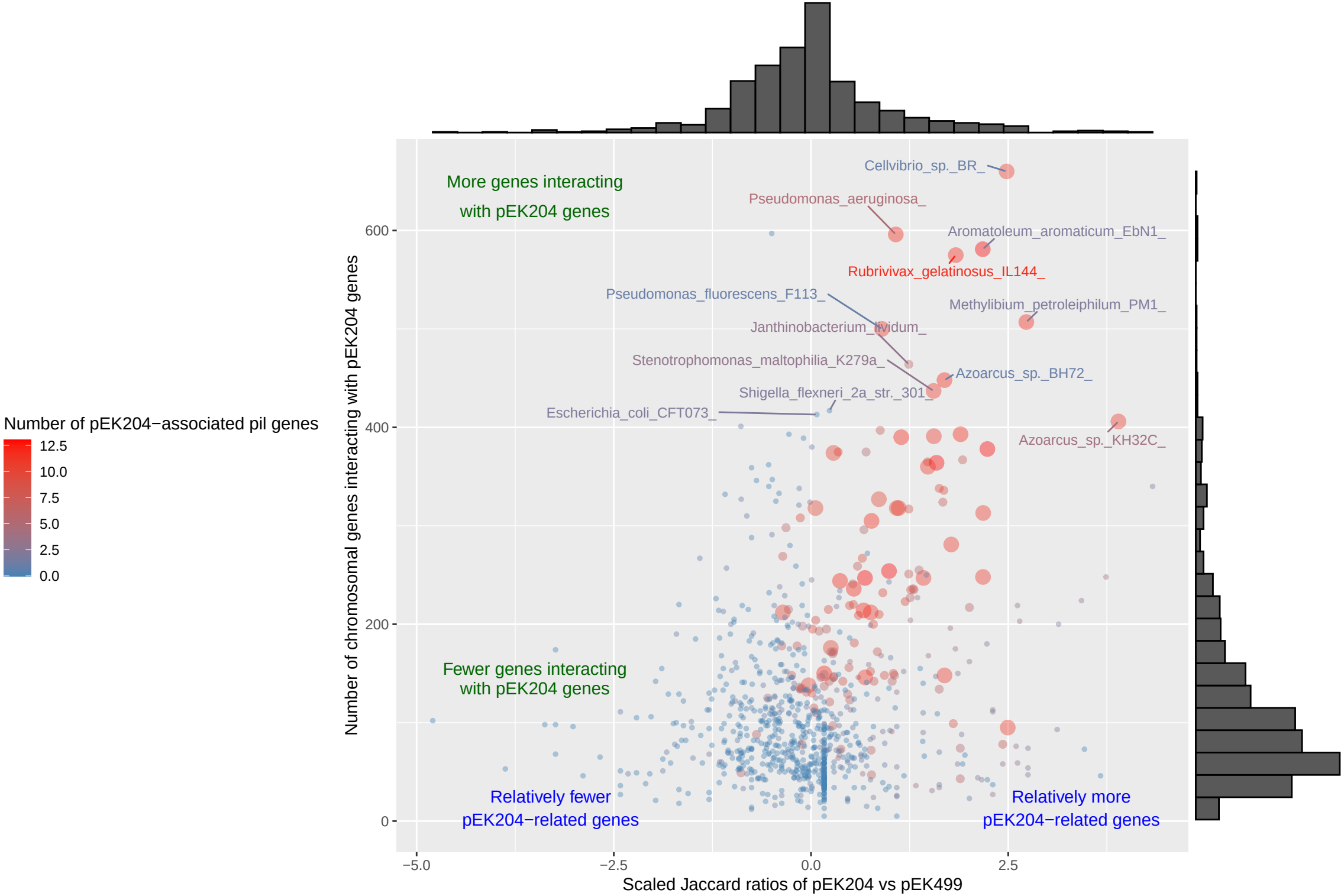
728 91. Carattoli A, Villa L, Fortini D, García-Fernández A. Contemporary IncI1 plasmids involved in the
729 transmission and spread of antimicrobial resistance in *Enterobacteriaceae*. Plasmid. 2021
730 118:102392. doi: 10.1016/j.plasmid.2018.12.001
731 92. Siryaporn, A. et al. Surface attachment induces *Pseudomonas aeruginosa* virulence. Proc. Natl
732 Acad. Sci. USA. 111, 16860–16865 (2014).
733 93. Persat A, Inclan YF, Engel JN, Stone HA, Gitai Z. Type IV pili mechanochemically regulate
734 virulence factors in *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A. 2015 112(24):7563-8.
735 doi: 10.1073/pnas.1502025112
736 94. Carter MQ, Chen J, Lory S. The *Pseudomonas aeruginosa* pathogenicity island PAPI-1 is
737 transferred via a novel type IV pilus. J Bacteriol. 2010 192(13):3249-58. doi: 10.1128/JB.00041-
738 10
739 95. Fonseca EL, et al. Emergence of XDR high-risk *Pseudomonas aeruginosa* ST309 in South
740 America: a global comparative genomic analysis BioRxiv 2021 doi:
741 https://doi.org/10.1101/2021.01.21.427610
742 96. Poulter RTM, Ho J, Handley T, Taiaroa G, Butler MI. Comparison between complete genomes of
743 an isolate of *Pseudomonas syringae* pv. actinidiae from Japan and a New Zealand isolate of the
744 pandemic lineage. Sci Rep. 2018 8(1):10915. doi: 10.1038/s41598-018-29261-5
745 97. Lorenz A, et al. Importance of flagella in acute and chronic *Pseudomonas aeruginosa* infections.
746 Environ Microbiol. 2019 21(3):883-897. doi: 10.1111/1462-2920.14468
747 98. Skurnik D, et al. A comprehensive analysis of in vitro and in vivo genetic fitness of *Pseudomonas*
748 *aeruginosa* using high-throughput sequencing of transposon libraries. PLoS Pathog. 2013
749 9(9):e1003582. doi: 10.1371/journal.ppat.1003582
750 99. Arlehamn CS, Evans TJ. *Pseudomonas aeruginosa* pilin activates the inflammasome. Cell
751 Microbiol. 2011 13(3):388-401. doi: 10.1111/j.1462-5822.2010.01541.x
752 100. Nagashima S, et al. Complete genome sequence of phototrophic betaproteobacterium *Rubrivivax*
753 *gelatinosus* IL144. J Bacteriol. 2012 194(13):3541-2. doi: 10.1128/JB.00511-12
754 101. Xia Y, Gally D, Forsman-Semb K, Uhlin BE. Regulatory cross-talk between adhesin operons in
755 *Escherichia coli*: inhibition of type 1 fimbriae expression by the PapB protein. Embo J 2000
756 19:1450–1457
757 102. Holden NJ, Uhlin BE, Gally DL. PapB paralogues and their effect on the phase variation of type 1
758 fimbriae in *Escherichia coli*. Mol Microbiol 2001 42:319–330
759 103. Holden NJ, Gally DL. Switches, cross-talk and memory in *Escherichia coli* adherence. J Med
760 Microbiol 2004; 53:585–593
761 104. Zhou H, et al. Functions predict horizontal gene transfer and the emergence of antibiotic
762 resistance. Sci Adv. 2021 7(43):eabj5056. doi: 10.1126/sciadv.abj5056


Protein-protein interaction data
27 Enterobacterales samples from StringDB
Check for PPIs with pOXA-48
24 Enterobacterales PPI data: Escherichia / Shigella (n=11), Klebsiella (n=2) & Serratia (n=11)
(iii)
2
For each sample, the:
- number of sample proteins interacting with plasmid proteins
- number of PPIs per plasmid
- Jaccard index of interacting proteins vs combined total of chromosomal & plasmid proteins
3
6
PPI data for pEK204, pEK499, pEK516 & pEC958
1
Get pEK204-related pil genes & PPIs
1,060 samples had pil-related PPIs
5
3,359 samples had no pil-related PPIs
4
4,363 samples with valid PPI data
(v)
Quality control
4,445 potential samples from StringDB
Protein-protein interaction data
(iv)
For each sample, the:
- number of sample proteins interacting with pOXA-48 proteins
- number of PPIs per pOXA-48 isoforms
- Jaccard index of interacting proteins vs combined total of chromosomal & pOXA-48 proteins
(vii)
(ii)
4 pOXA-48 isoforms
Check gene content
(i)
31 plasmids matching "pOXA-48"
(vi)
2,396 samples with pOXA-48 -related PPI data
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
600
400
200
0
−5.0
−2.5
0.0
2.5
Scaled Jaccard ratios of pEK204 vs pEK499