

# Bacterial warfare is associated with virulence and antimicrobial resistance

Received: 10 March 2025

Connor Sharp <sup>1,2,3</sup>  & Kevin R. Foster <sup>1,2,4</sup> 

Accepted: 16 September 2025

Published online: 22 October 2025

 Check for updates

Bacteria have evolved a diverse array of mechanisms to inhibit and kill competitors. However, why some bacteria carry such weapons while others do not remains poorly understood. Here we explore this question using the genomics of the bacteriocins of *E. coli* as a model system, which have large well-annotated bioinformatic resources. While bacteriocins occur widely, we find that carriage is particularly associated with pathogenic extra-intestinal (ExPEC) strains. These pathogens commonly carry large plasmids encoding bacteriocins alongside virulence factors and antimicrobial resistance mechanisms. Across all strains, we find many orphan immunity proteins, which protect against bacteriocins and suggest that these bacterial weapons are important in nature. We also present evidence that bacteriocin toxins readily move between strains via plasmid transfer and even between plasmids via transposons. Finally, we show that several *E. coli* bacteriocins are widely shared with the pathogen *Salmonella enterica*, further cementing the link to virulence. Our work suggests that the bacteriocins of *E. coli* are important antibacterial weapons for dangerous antimicrobial-resistant strains.

Bacteria often exist in dense and diverse communities where competition for nutrients or even just space to grow is intense<sup>1–4</sup>. In response to this competitive world, bacteria have evolved an arsenal of competition systems, also known as bacterial weapons, to inhibit each other<sup>5–7</sup>. These weapons can give bacteria an advantage over neighbouring cells and are important in determining the composition of bacterial communities<sup>8,9</sup>. The ability to invade or persist within a community is also critical for many bacterial pathogens, including in the human gut where the dense community of bacteria in the microbiome can limit pathogen growth<sup>10–13</sup>. While the potential benefits of carrying antibacterial weapons is clear, there is wide variation in both the number and types of weapons bacteria possess<sup>7</sup> and we have a limited understanding of why some bacteria are heavily armed, while others are not<sup>14–16</sup>.

There is a need, therefore, to understand the role of bacterial weapons in the ecology of bacteria. Here we approach this challenge using large genomic databases that have recently become available for the model species *Escherichia coli*. In addition to the quality of the

genomic data for *E. coli*, the bacteriocins of *E. coli* are amongst the best-studied of all bacterial weapons and there is already a detailed understanding of many of their targets and molecular mechanisms<sup>17</sup>. But while bacteriocins were first described nearly 100 years ago<sup>18</sup>, we still understand little about their roles in natural systems. Currently, there are a range of conflicting arguments that bacteriocins are important for bacterial competition<sup>19–21</sup>, do not impact bacterial communities at all<sup>16,22,23</sup>, or even exist solely as selfish genetic elements, maintaining themselves through a toxin/immunity system at the cost of the host<sup>24,25</sup>. In addition to its genomic resources, *E. coli* is an ideal model system for investigating the role of bacterial weapons in the ecology of bacteria because there is a large amount of variability in both weapon carriage and pathogenicity within the species, which allows one to look for associations between weapon use and ecology while avoiding potentially confounding differences among species.

There are two main classes of *E. coli* bacteriocins—colicins and microcins—which are protein or smaller peptide toxins respectively, that target closely related strains and species. These bacteriocins are

<sup>1</sup>Department of Biology, University of Oxford, Oxford, UK. <sup>2</sup>Department of Biochemistry, University of Oxford, Oxford, UK. <sup>3</sup>School of Biological Sciences, University of Reading, Reading, UK. <sup>4</sup>Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ✉e-mail: [c.sharp@reading.ac.uk](mailto:c.sharp@reading.ac.uk); [kevin.foster@path.ox.ac.uk](mailto:kevin.foster@path.ox.ac.uk)

encoded either on plasmids or in the bacterial chromosome as a toxin and co-expressed immunity gene to prevent self-intoxication. Colicin-like proteins are also found in other Gammaproteobacteria species<sup>26</sup>, including in the important pathogens *Salmonella enterica*<sup>27</sup> and *Klebsiella pneumoniae*<sup>28</sup>. To enter a target cell, colicins and microcins hijack outer membrane proteins<sup>29</sup>, whose typical function is to import nutrients such as B<sub>12</sub>, nucleosides and iron<sup>17</sup>. Detecting bacteriocins on plasmids has, until recently been a limitation, due to the quality of data available<sup>30–32</sup>. However, the massive increase in plasmid sequencing and improvements in assembly means one can now perform thorough analysis of plasmid-borne traits. In particular, here we are able to study thousands of genomes and plasmids from two large databases in order to analyse bacteriocins across the *E. coli* species.

We find that *E. coli* bacteriocins are abundant in pathogenic strains and linked to known virulence factors such as siderophores, adhesins and enterotoxins, as well as large arrays of multiple antimicrobial resistance (AMR) genes. Importantly, we also find plasmids containing orphan immunity proteins, which protect against bacteriocin attack without producing toxins, and support the importance of bacteriocins for bacterial competition (as opposed to solely acting as a plasmid addiction system). Finally, our data suggest that horizontal gene transfer of bacteriocins occurs on multiple levels: between species, within species and even from plasmid-to-plasmid via the action of transposons. We conclude that bacteriocins are important for bacterial ecology, competition and virulence.

## Results

### Bacteriocins are widespread throughout *E. coli*

We curated a dataset of 2601 *E. coli* strains with complete genomes (1 contig for the chromosome and 1 for each plasmid) and scanned these genomes for bacteriocins: colicin and microcins, either on plasmids or in the chromosome (Table 1). This analysis identified 608 strains (23.4% of our *E. coli* genomes) that encode at least one bacteriocin. Bacteriocins are mostly encoded on plasmids, with over three times more strains encoding plasmid bacteriocins than chromosomally encoded bacteriocins (18.5% vs 6.1% of all strains). Focusing on the plasmids, there are 571 bacteriocin plasmids in the data set which encode 730 bacteriocin genes (plasmids can encode multiple different colicins and microcins) spread across 466 strains. Bacteriocins are also very widely distributed across *E. coli* strains and are present in all *E. coli* phylogroups: 75% of phylogroup G plasmids carried bacteriocins (21/28), 46.3% of B2 (173/374), 34.4% of B1 (166/482), 24% of C (22/93), and 12% of A (121/990) (Other or un-typeable phylogroups: 105/634). In addition, bacteriocins are present in >90% (38/44) of all sequence types (STs) which had more than 10 strains in the dataset (Fig. 1B). Similarly, we find bacteriocins in 121/183 of the predicted O-antigen types in our dataset, and in 13/13 of all O-antigen types present in 40 or more strains.

Previous estimates of bacteriocin carriage rates in *E. coli* vary considerably from 15 to 60%. However, sampling and method of identification also vary between estimates with some estimates based exclusively on human faecal samples or specific *E. coli* pathotypes<sup>33,34</sup>. Our analysis of thousands of *E. coli* genomes and plasmids suggests that differences in bacteriocin abundance among phylogroups may explain the different estimates of bacteriocin carriage among studies. For example, the high prevalence of phylogroup A in our dataset, which has low bacteriocin carriage (12%) may help to explain why our estimate is relatively low as compared to studies focussed on other phylogroups and pathogenic *E. coli*<sup>34–37</sup>. Another possible factor for our work is that plasmid sequences may have been lost from certain samples, although our quality control methods that selects for complete genomes should mean this is relatively unlikely. Loss of plasmid DNA during DNA extraction is a possibility that we cannot exclude because sequencing quality control cannot mitigate against this.

Whatever the case, our estimates of bacteriocin carriage are likely to be conservative.

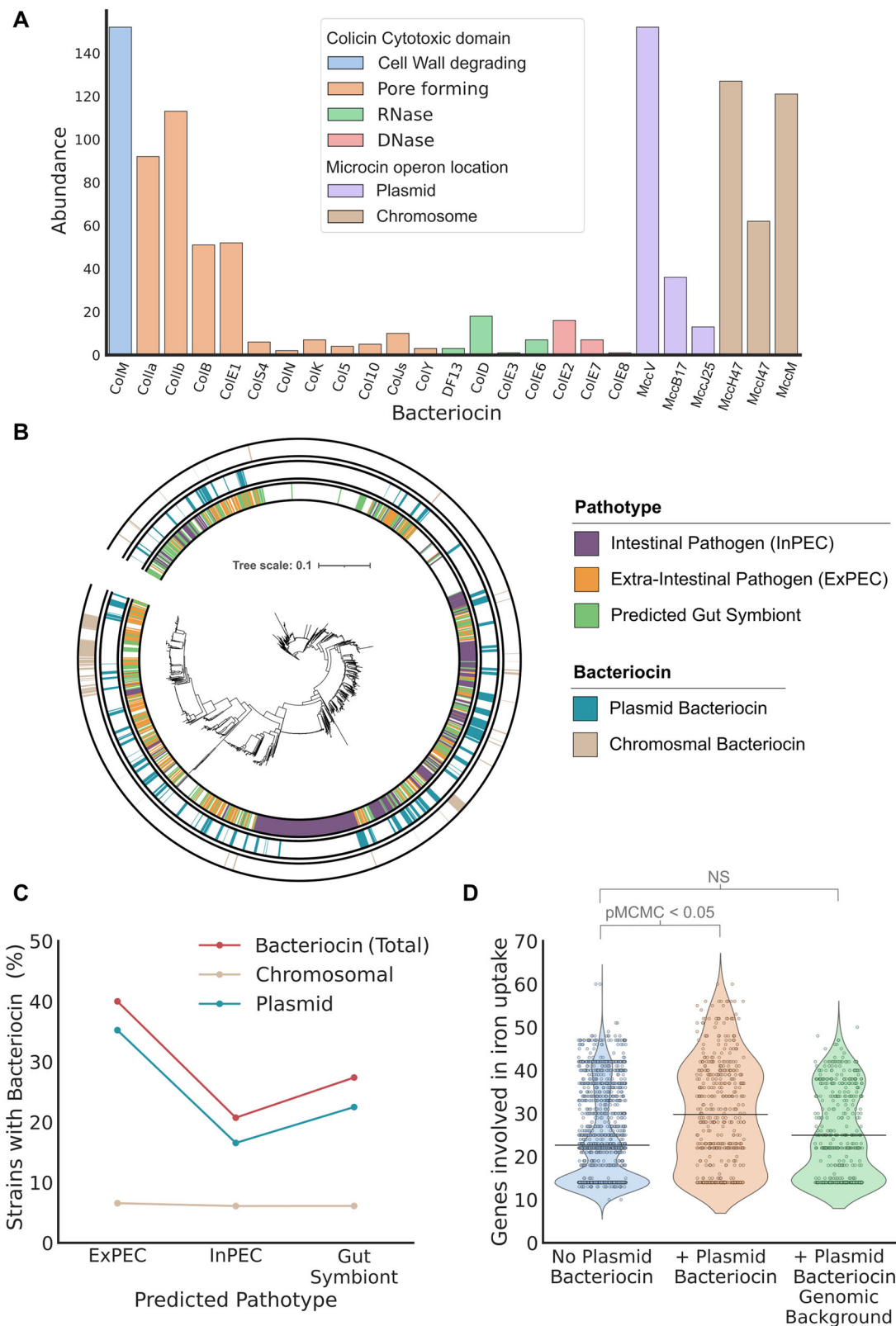
Some types of bacteriocin are much more prevalent than others in the data. Colicins M, Ia, Ib, E1 and Microcin V are the most abundant (Fig. 1A), which agrees with similar studies using cohorts of *E. coli* strains<sup>33,38</sup>. A key implication of this pattern is that pore-forming and peptidoglycan precursor degrading toxins, which both disrupt membrane integrity were more common than nucleases (DNases, tRNases and rRNases). Nuclease colicins, such as the E-type colicins, which are often used a model system for colicin translocation<sup>39</sup> and protein-protein interactions<sup>40</sup>, are relatively rare in our database, representing <10% of the identified colicins.

Colicins and microcins are typically studied separately, and each have their own literatures<sup>41,42</sup>. However, our analyses suggest that they are functionally linked, with a strongly significant association between microcin and colicin carriage at the strain level (binaryPGLMM:  $\beta = 2.07$ , SE = 0.188,  $p = 0.001$ ). A driver of this pattern is that most strains which encode a microcin on a plasmid also have a plasmid-borne colicin (63.5% of microcin carrying strains, Fisher exact test: OR 12.67,  $p = <0.001$ ). Taken from the plasmid perspective, this manifests as 48.3% of microcin plasmids also encoding a colicin, e.g. a plasmid may encode genes for the microcin MccV and colicin ColM. These

**Table 1 | Bacteriocins differ in their targeted receptor and the cytotoxic domain they use to kill sensitive cells**

Bacteriocin	Receptor	Cytotoxic Activity	Genetic Location
ColM	FhuA	Peptidoglycan precursor degrading	Plasmid
ColIa	Cir	Pore-formation	Plasmid
ColIb	Cir	Pore-formation	Plasmid
ColB	FepA	Pore-formation	Plasmid
ColE1	BtuB	Pore-formation	Plasmid
ColS4	OmpW	Pore-formation	Plasmid
ColN	OmpF	Pore-formation	Plasmid
ColK	Tsx	Pore-formation	Plasmid
Col5	Tsx	Pore-formation	Plasmid
ColJs	-	Pore-formation	Plasmid
ColY	-	Pore-formation	Plasmid
ColDF13	lutA	rRNase	Plasmid
ColD	FepA	rRNase	Plasmid
ColE3	BtuB	rRNase	Plasmid
ColE6	BtuB	rRNase	Plasmid
ColE2	BtuB	DNase	Plasmid
ColE7	BtuB	DNase	Plasmid
ColE8	BtuB	DNase	Plasmid
MccV (Colicin V) <sup>35</sup>	Cir	IM depolarisation	Chromosome
MccB17	OmpF	DNA Gyrase Inhibition	Plasmid
MccJ25	FhuA <sup>136</sup>	RNA polymerase inhibition <sup>137</sup>	Plasmid
MccH47 <sup>138</sup>	Cir, Fiu, FepA and IroN	ATP Synthase disruption	Chromosome
MccI47	Cir, Fiu, FepA and IroN <sup>139</sup>	-	Chromosome
MccM	Cir, Fiu, FepA and IroN <sup>139</sup>	inhibition of oxidative phosphorylation <sup>140</sup>	Chromosome

Bacteriocins can also either be plasmid encoded or chromosomal. '-' indicates an unknown receptor. Information is from<sup>17</sup>, unless otherwise stated.



**Fig. 1 | The bacteriocins of *E. coli* are diverse and associated with particular pathotypes.** **A** Abundance of different bacteriocins found throughout 2601 *E. coli* genomes. **B** Core-genome phylogeny of 2601 *E. coli* strains with predicted pathotype (inner ring), presence of bacteriocin plasmids (middle ring) and chromosomal bacteriocins (outer ring). **C** Bacteriocin plasmids are not evenly distributed between non-pathogenic and pathogenic strains: bacterial genotypes that cause extraintestinal infections are particularly likely to carry a bacteriocin. **D** Bacteria

with iron-uptake targeting bacteriocin plasmids are enriched in iron-uptake genes (orange) compared to strains without bacteriocin plasmids (blue) (MCMCgImm, posterior  $\mu = 6.07$ , 95% CI = [5.500, 6.669], pMCMC <  $3 \times 10^{-4}$ ). Removing genes encoded on these plasmids (green) reveals that that the enrichment is driven by the fact that bacteriocin plasmids carry iron-uptake genes rather than the plasmids associating with genomes that carry high numbers of iron uptake genes (MCMCgImm, posterior  $\mu = 0.4$ , 95% CI = [-0.1709, 0.9314]), pMCMC = 0.15).

patterns are striking given that only around 20% of *E. coli* bacterial strains have a bacteriocin at all. More generally, we find that strains with multiple weapon systems occur in much greater numbers than predicted if weapons were distributed independently (Poisson distribution KS test  $p < 0.005$ ), with 3.5% of strains (93 strains) encoding more than one weapon, compared to 0.83% (22 strains) predicted using a Poisson distribution (Supplementary Fig. 1). In sum, these analyses suggest that there is a subset of *E. coli* strains that are particularly specialised for bacterial warfare.

### Bacteriocin plasmids are enriched in extra intestinal *E. coli* infections

*E. coli* strains are often harmless symbionts that colonise the gut microbiome and can even protect against incoming bacterial pathogens<sup>10,13,43,44</sup>. By contrast, some *E. coli* strains cause severe intestinal and extra-intestinal disease—including: urinary tract infections, sepsis, pneumonia and meningitis – to the extent that *E. coli* as a species is a dominant cause of mortality<sup>45</sup>, and specifically the mortality associated with antibiotic resistant infections<sup>46</sup>. As a result, there is great interest in the drivers of *E. coli* virulence.

Many intestinal *E. coli* pathotypes are clinically defined by the presence of a small number of virulence factors<sup>47–49</sup>, while extra-intestinal infections are defined from the presence of *E. coli* in host niches outside the gut e.g. urinary tract, bloodstream or lung. We can, therefore, use the presence of certain virulence factors, alongside metadata on where a strain was isolated from, to categorise our strains of *E. coli* into Intestinal Pathogenic (InPEC), Extraintestinal Pathogenic (ExPEC), or gut associated and lacking virulence factors, which we use as a proxy for non-pathogenic *E. coli* gut symbionts (Table 2). This classification is reliant upon the quality of the available metadata and does not distinguish between hybrid strains capable of causing multiple pathologies. Nevertheless, across the large dataset we have analysed, we see clear and interesting differences between the strain categories. Our analysis also agrees with known trends in *E. coli* phylogroups and sequence types. We find that phylogroup B2 is the most common group in ExPECs, E and B1 in InPECs and A in predicted gut symbionts which agrees with PCR and genomic analysis of *E. coli* pathotypes<sup>50–52</sup> (Supplementary Fig. 2). Our methods also correctly classified common clinically important ExPEC lineages such as ST95, ST69 and ST131 and InPEC lineages such as ST11<sup>53</sup> and ST21<sup>54</sup>.

Our analysis reveals significant differences in the carriage of bacteriocins by the different pathotypes (Fig. 1C), with bacteriocins significantly enriched in the ExPEC pathotype compared to other strains (binaryPGLMM  $\beta = 0.60$ , SE = 0.20,  $p < 0.005$ ). Splitting bacteriocins into plasmid and chromosomal makes it clear that it is the plasmids that drive the association between bacteriocins and ExPEC strains (Plasmid weaponry: binaryPGLMM  $\beta = 0.93$ , SE = 0.22,  $p < 0.001$ , Chromosomal weaponry: binaryPGLMM  $\beta = -0.14$ , SE = 0.41,  $p = 0.74$ ) (Fig. 1C). Consistent with this pattern, ExPEC strains are dominated by globally disseminated pandemic lineages and bacteriocin plasmids are prevalent in a number of clinically-important ExPEC sequence types including: ST95 (52% of strains), ST58 (50%), ST69 (27%), ST131 (25.6%) and ST457 (28.6%).

**Table 2 | Virulence factors used to distinguish between different *E. coli* pathotypes**

Virulence factor Marker	Description	Pathotype
eae	Intimin attachment to epithelial cells	EPEC <sup>141</sup>
pINV	Invasion plasmid	EIEC <sup>142</sup>
LT	Heat labile enterotoxin	ETEC <sup>143</sup>
ST	Heat stable enterotoxin	ETEC <sup>144</sup>
Stx	Shiga toxin	EHEC <sup>145</sup>

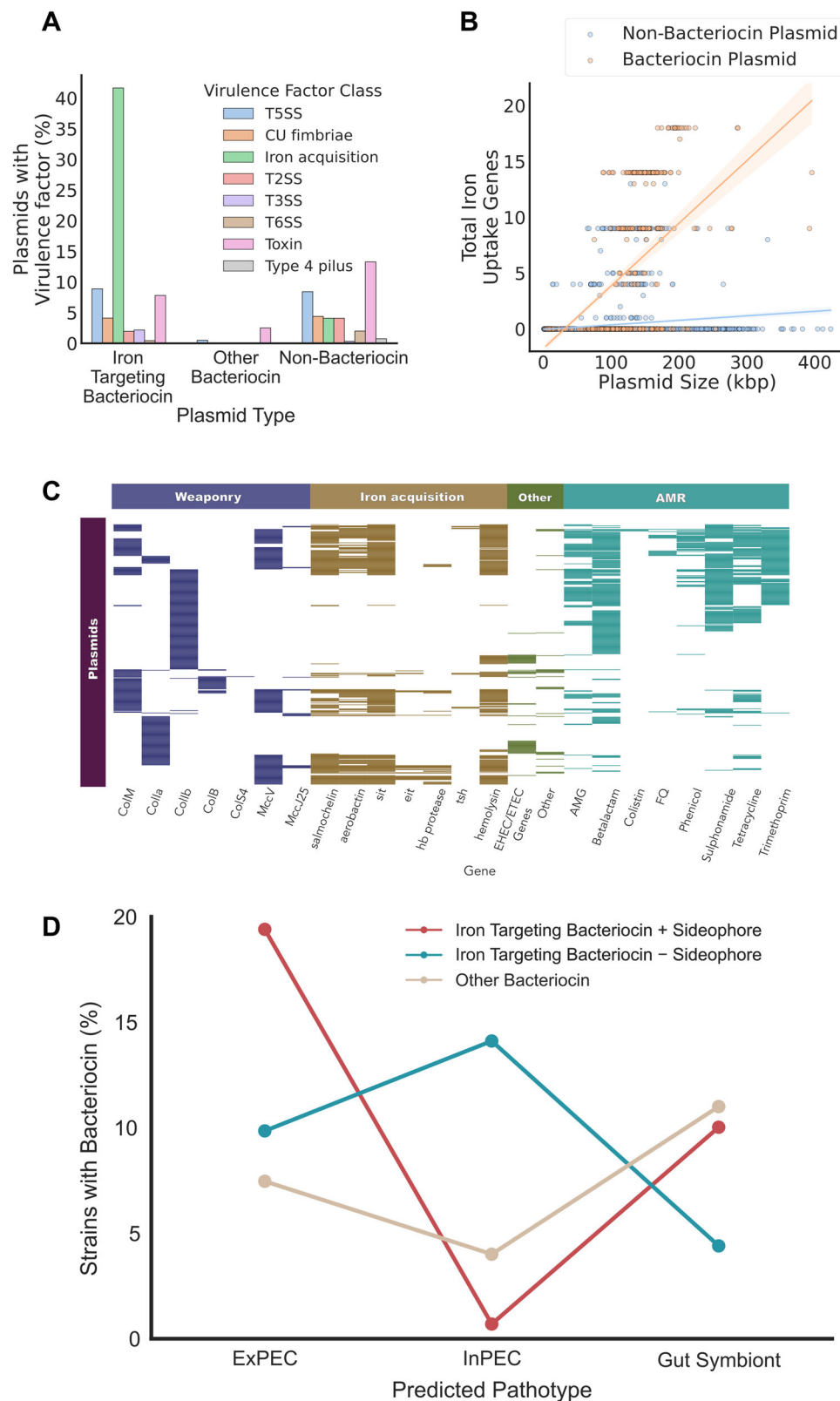
We next sought to identify the mechanisms that explain the link between bacteriocin plasmids and disease. Looking for associations across a range of *E. coli* virulence factors suggests that one class of virulence factors is key to the association with bacteriocins: mechanisms of iron acquisition (MCMCglmm, posterior  $\mu = 6.07$ , 95% CI = [5.500, 6.669], pMCMC < 0.01, other virulence factor classes were not significant) (Fig. 1D). This pattern could be explained by the bacteriocin plasmids encoding iron-uptake virulence factors or the plasmids occurring in strains that encode iron-uptake genes in their chromosome. However, we found no correlation between plasmid-borne bacteriocins and chromosomal iron-uptake genes (MCMCglmm, posterior  $\mu = 0.4$ , 95% CI = [-0.1709, 0.9314]), pMCMC = 0.15). Instead, the data suggest that plasmids can encode both bacteriocins and virulence factors and bring them into a cell together. We, therefore, decided to focus on the gene content of the plasmids to explore this pattern further.

### Bacteriocin plasmids are enriched in siderophores and other virulence factors

To further investigate the traits carried by bacteriocin plasmids, we turned to a large plasmid database. Scanning 5373 *E. coli* plasmids from the PLSDB<sup>55</sup> database reveals 660 plasmids encoding bacteriocins (12% of all *E. coli* plasmids in PLSDB: 568 colicin plasmids and 199 microcin plasmids). Consistent with the above analyses, these bacteriocin-carrying plasmids are enriched in genes related to siderophores and the acquisition of iron, even after adjusting for plasmid size (GLM,  $\beta = 0.06$ ,  $p < 2 \times 10^{-16}$ ) (Fig. 2A, B) However, the associations with siderophores only occur for certain bacteriocins, ColM, Colla, Collb, ColB, MccV and MccJ25, which all target iron-uptake receptors (CirA, FepA, FhuA), (Fig. 2A, C). Given that their target receptors are involved in the uptake of siderophores and iron, hereafter we call these bacteriocins ‘iron-targeting bacteriocins’ as an abbreviation of iron-uptake receptor targeting bacteriocins. Not only are iron-targeting bacteriocins associated with siderophores, the plasmids that carry siderophores often encode multiple iron-targeting weapons, including combinations of both microcins and colicins. For example, microcin V is associated with an iron-targeting colicin on 66% of MccV plasmids that also all carry siderophores.

The iron-targeting bacteriocin plasmids carry a range of siderophore systems: salmochelin and aerobactin, or ABC transporters such as Sit and Eit, all of which are important virulence factors involved in the acquisition and import of iron<sup>56–59</sup> (Fig. 2C). For example, salmochelin is a derivative of the siderophore enterobactin that is able to function in spite of the host releasing lipochalin-2 to suppress enterobactin functioning<sup>60</sup>, and aerobactin is a citrate-hydroxamate siderophore with extremely high affinity for iron<sup>58</sup>. Colicins are upregulated by DNA damage and we see this in our dataset, with LexA boxes upstream of nearly all of colicin genes (Supplementary Fig. 3). However, in addition to LexA, we find the majority of the iron-targeting bacteriocins associated with siderophores are upregulated via the Fur system, which responds to low-iron concentrations<sup>61</sup>. Fur regulated genes also include siderophores and their receptors, including the receptors targeted by iron-targeting bacteriocins: Cir<sup>62</sup>, FepA<sup>63</sup> and FhuA<sup>64</sup>.

In addition to siderophores, we find a range of other virulence factors encoded on the plasmids that contain iron-targeting bacteriocins (Fig. 2C, Table 3), including temperature sensitive hemagglutinin<sup>65,66</sup>, hemolysin<sup>67</sup> and haemoglobin-binding protease<sup>68</sup> which have all been implicated in the virulence of extra-intestinal *E. coli*. We also identify iron-targeting bacteriocin plasmids which lack siderophores but encode virulence factors such as adhesins, fimbriae, enterotoxins and proteases associated with many intestinal pathotypes including enterotoxigenic, shiga-toxic and enterohemorrhagic *E. coli*. These non-siderophore plasmids (Fig. 2C) can encode three different



**Fig. 2 | Bacteriocin plasmids carry a range of virulence factors and AMR genes and are enriched in iron-uptake genes.** **A** Bacteriocin plasmids are enriched in genes associated with iron-uptake genes as compared to other plasmids.

**B** Bacteriocin plasmids are more likely to encode iron-uptake genes than non-bacteriocin plasmids, even when accounting for plasmid length (GLM,  $\beta = 0.06$ ,  $p < 2 \times 10^{-16}$ ). Shaded areas indicate 95% CIs. **C** Iron-targeting bacteriocins occur alongside diverse iron acquisition systems and other loci involved in

extraintestinal, enterohaemorrhagic and enterotoxic pathotypes. Bacteriocin plasmids can also carry antibiotic resistance genes for a wide range of antibiotic classes. **D** Different *E. coli* pathotypes tend to carry different types of bacteriocin plasmid. *E. coli* strains that cause intestinal infections often carry plasmids with iron-targeting bacteriocin but without siderophores, while ExPEC strains are enriched for all types of iron-targeting bacteriocin plasmids.

**Table 3 | Virulence factors associated with bacteriocinogenic plasmids from Intestinal pathogens**

Virulence Factor	Description	Strain references
<b>Group 1</b>		
<i>espP</i>	Serine Protease, which cleaves coagulation factor V <sup>146</sup>	Outbreaks of haemorrhagic colitis in Montana in 1994 <sup>147</sup> and from Haemolytic Uraemic Syndrome (HUS) patients in Australia <sup>148</sup>
<i>epeA</i>	serine protease that degrades mucin <sup>51,149</sup>	
<i>Saa</i>	autoagglutinating adhesin <sup>150</sup>	
<i>subAB</i>	subtilase cytotoxin <sup>151</sup>	
<b>Group 2</b>		
ST	heat-stable enterotoxin <sup>152</sup>	Shiga toxin encoding <i>E. coli</i> ST443 found in cattle from south America <sup>153</sup> and cases of diarrhoea <sup>154</sup>
Csf	CS5 fimbriae <sup>155</sup>	
Css	CS6 fimbriae <sup>156</sup>	
<i>aat</i>	anti-aggregation protein or dispersin <sup>157</sup>	
<i>pic</i>	mucin degrading enzyme <sup>158</sup>	
<b>Group 3</b>		
<i>nleA</i>	immune disrupting effector of the T3SS <sup>159</sup>	<i>E. coli</i> O55:H7 strains, a progenitor strain to the highly pathogenic <i>E. coli</i> O155:H7, but itself responsible for major diarrhoea outbreaks and cases of HUS <sup>160</sup>
<i>stcE</i>	mucin protease <sup>161</sup>	
Etp	T2SS <sup>162</sup>	
<i>ibeC</i>	outer membrane protein linked to epithelial adherence and meningitis <sup>163</sup>	
<i>katP</i>	catalase-peroxidase <sup>164</sup>	

sets of virulence factors (Table 3) and many have been linked to outbreaks of severe intestinal disease and HUS.

We also find that iron-targeting bacteriocin plasmids encode large arrays of AMR genes: 62% of all iron-targeting bacteriocin plasmids contain AMR genes predicted to provide resistance to at least one class of antibiotic, with 36.7% encoding resistance to  $\geq 3$  antibiotic classes. In contrast, only 9% of non-iron targeting bacteriocin plasmids encode AMR genes. Mechanisms of resistance to 8 different antibiotic classes are seen in the data, including resistance to the ‘last resort’<sup>69</sup> antibiotic colistin, and one bacteriocin plasmid is predicted to have resistance to all 8 classes: pAR349, which encodes the ESBL *bla*<sub>CTX-M-14</sub> and colistin resistance gene *mcr-1* (Accession: NZ\_CP041997.1). The combination of bacteriocins and AMR genes was not restricted to a single replicon but found on 42 different replicons (predicted by PlasmidFinder). The most common bacteriocin/antibiotic combination replicons predicted were IncI1-ly, which encodes genes for Colla/Ib, followed by different IncF replicon subtypes. Bacteriocin carrying plasmids are not more likely than other plasmids to carry AMR genes once one accounts for their large size (GLM,  $\beta = -0.0001$ ,  $p = 0.34$ ) (Supplementary Fig. 4). Nevertheless, the effect is that, when a bacterial strain carries one of these plasmids, it becomes more likely to carry AMR genes than other strains, meaning that weaponry and AMR are significantly associated in *E. coli* (MCMCglmm, posterior  $\mu = 0.9138$ , 95% CI = [0.6896, 1.1461]), pMCMC = <0.01) (Supplementary Fig. 5).

### ***E. coli* bacteriocins associate with different pathotypes based on receptor target**

Our work suggests that large iron-targeting bacteriocin plasmids drive the key association between bacteriocin use, virulence and AMR in *E. coli*. These plasmids bring with them a complex of phenotypes. In particular, these plasmids often appear to help bacteria compete under low-iron conditions, such as during inflammation, when both iron-acquisition systems and iron-targeting weapons can be useful, as has been shown for Collb<sup>20</sup> and certain microcins<sup>14</sup>.

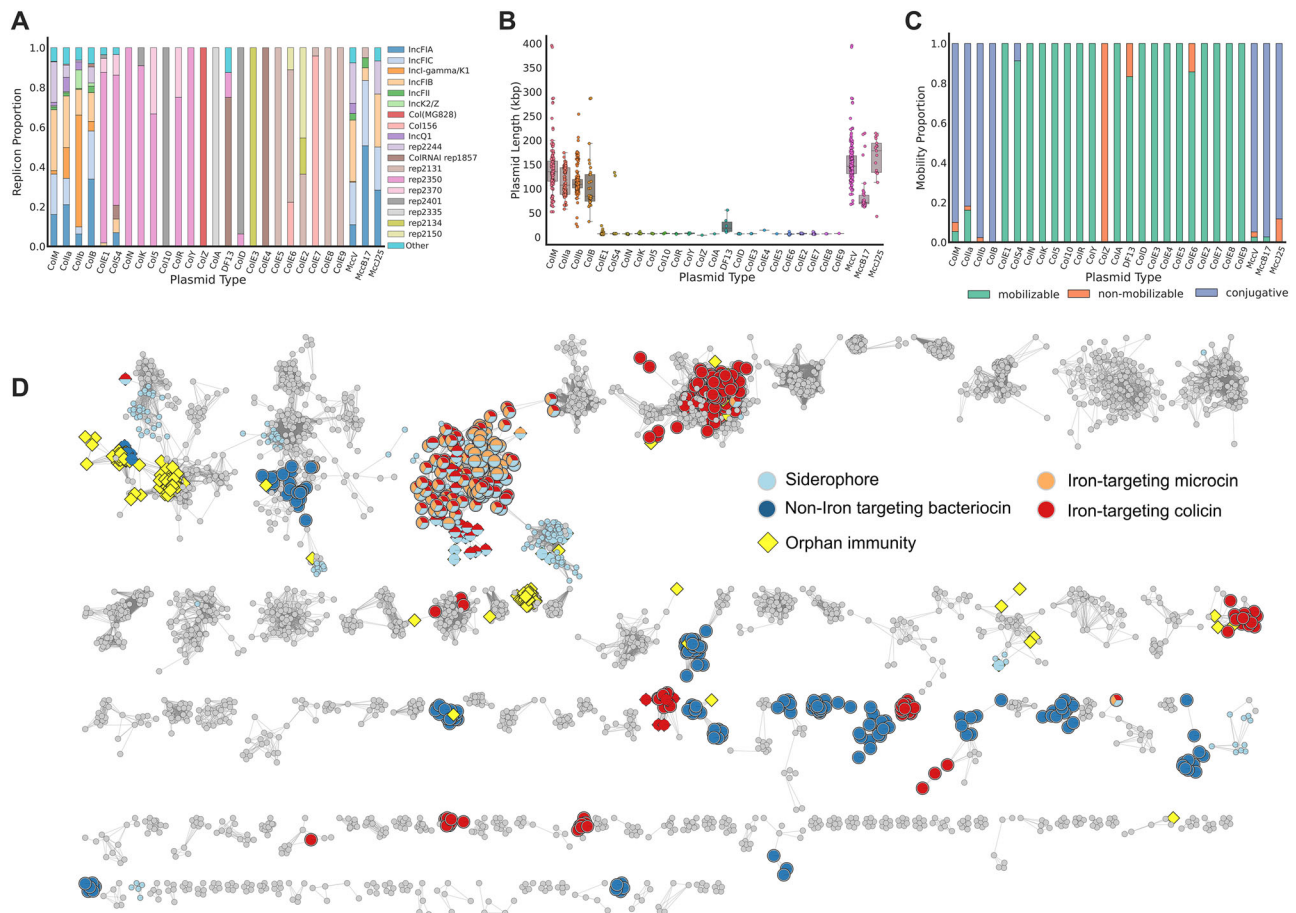
Consistent with the importance of these large iron-targeting bacteriocin plasmids, we do not observe any association between non-iron targeting bacteriocin plasmids and bacterial pathotypes in our dataset of *E. coli* genomes. By contrast, the iron-targeting bacteriocins are significantly more common in ExPEC vs non-pathogenic strains (Iron targeting bacteriocins: binaryPGLMM  $\beta = 1.02$ , SE = 0.23,  $p < 0.001$ ). However, there are also differences among pathogenic

strains in the bacteriocin plasmids that they carry. ExPEC are much more likely to carry plasmids with both iron-targeting bacteriocins and siderophores (binaryPGLMM  $\beta = 3.18$ , SE = 0.72,  $p < 0.001$ ) as compared to InPEC, which tend to carry iron-targeting bacteriocin plasmids lacking siderophores (binaryPGLMM  $\beta = 1.06$ , SE = 0.44,  $p < 0.05$ ). The specific genetic repertoire of bacteriocin plasmids is, therefore, associated with different *E. coli* pathologies, with siderophores associated with ExPEC strains while host-targeting toxins are abundant in InPEC strains (Fig. 2D). Given the importance of intestinal carriage for all *E. coli* strains, it is likely that these patterns reflect differences in the ecology of ExPEC and InPEC strains in the mammalian gut. However, the extra siderophores carried by ExPEC strains may also be important for their tendency to disseminate to other body sites<sup>70–72</sup>.

Our analyses fit with previous work on *E. coli* bacteriocins. For example, plasmids encoding the microcin MccV (known as pColV plasmids) have been previously identified in ExPEC strains and found to be maintained in lineages for long periods<sup>73,74</sup>. We also found pColV plasmids to be common in ExPEC strains, with 15.8% of predicted ExPEC strains encoding a MccV+ plasmid<sup>74</sup>. Previous research also has shown that ExPEC strains commonly possess plasmids with specific IncF replicons<sup>73–75</sup>. Analysing our ExPEC strains with pColV plasmids, we find the commonly reported IncF replicon sequence types including F18:A:B-, F18:A:B1, F24:A:B1, F24:A:B- and F2:A:B1. In addition, we found pColV plasmids commonly encode multiple bacteriocins, with 58% of pColV plasmids possessing additional bacteriocins such as ColM, Colla and Collb. Moreover, other ExPEC strains in the dataset encoded Colla, Collb and ColM bacteriocins on a range of MccV- IncF and IncI plasmids.

### **Colicin and microcin plasmids are diverse and mobilisable**

Our work suggests that the weaponry, virulence factors and AMR genes encoded on bacteriocin plasmids are important for the ecology and virulence of *E. coli*, yet we know little about these plasmids in the context of the full set of *E. coli* plasmids (its plasmidome). To study these plasmids, we can use PlasmidFinder<sup>76</sup> annotations provided in PLSDb to categorise plasmids by their replicons (regions of DNA associated with plasmid replication). This analysis reveals 93 different replicons across 5373 *E. coli* plasmids and, whilst bacteriocin-carrying plasmids represented only ~12% of our dataset, they carry over a third of all plasmid replicons found in *E. coli* (32 replicons). Even individual



**Fig. 3 | Bacteriocin plasmids are diverse in size, replication machinery and mobilisation, and orphan immunity genes occur widely.** **A** Bacteriocin genes occur alongside a wide range of replicon types. The 19 most abundant replicons out of a total of 31 are shown. **B** Bacteriocin plasmids vary in size and broadly divide into small and large plasmids. Each point represents a single plasmid. Plasmids with multiple colicins are represented in multiple columns, total colicins = 880. Lower and upper hinges indicate the 1st and 3rd quartiles. Central lines indicate the

median. Whiskers indicate the furthest point within 1.5× the interquartile range. **C** Bacteriocin plasmids are commonly mobilisable with larger plasmids encoding conjugation machinery and smaller plasmids carrying the genes to mobilise via a helper plasmid. **D** Clustering of 5353 *E. coli* plasmids from PLSDDB. Bacteriocin plasmids are diverse and spread throughout the *E. coli* plasmidome. Each circle represents an individual plasmid, edges indicate a Mash distance below the cut off (0.03). Only plasmid communities which contain  $\geq 5$  members are represented.

colicins were encoded on plasmids with a range of different replicon types. For example, colicin M is found with 16 different replicon types, whilst colicins Ia and Ib both associate with more than 10 different replicon types (Fig. 3A). More generally, bacteriocin plasmids are commonly mosaic plasmids in the sense that one can find multiple replicon types on a single bacteriocin plasmid (282/660 bacteriocin plasmids have multiple replicons, mean number of replicons per bacteriocin plasmid = 1.7). These patterns suggest frequent genetic exchange among the bacteriocin plasmids, including the transfer of the replication machinery itself. Carrying multiple replicons in this way may allow for a broader host range for plasmids, because replication machinery can differ in its ability to function in different bacterial strains and species<sup>77</sup>.

Previous analysis of colicin plasmids grouped them into type I small (~10 kb) multicopy plasmids and type II large (~40 kb) monocopy plasmids<sup>31</sup>. However, this distinction is based on small numbers of archetypal plasmids. We, therefore, looked in the large plasmid dataset to see if colicin plasmids lengths fell into these groups. This analysis reveals that, whilst there was a clear distinction between small type I and large type II bacteriocin plasmids, many far exceeded the 40 kb size estimate described in literature (Fig. 3B), with both colicin and microcin plasmids showing a large variation in size. For example, colicin M plasmids show a ~7.5-fold range in plasmid size (min = 52,297 bp, max = 395,758 bp).

Plasmid mobility - whether they can transfer to other cells via conjugation - has the potential to have a major impact during bacterial warfare because a weapon-encoding plasmid moving to a new strain could render it immune to attack<sup>75,78–81</sup>. Plasmid mobility can be predicted using the tool Mob-Suite<sup>82</sup>. Strikingly, we find that over 97% of colicin plasmids are predicted to be mobile, with 62.8% encoding their own conjugation machinery and a further 34.7% mobilisable by a helper plasmid (Fig. 3C). Similarly, 95% of microcins were mobilisable by conjugation. This analysis suggests that bacteriocin plasmids are highly mobile, even as compared to the average plasmid in the *E. coli* population where only 78% were predicted to be mobile.

### Bacteriocin plasmids and orphan immunity proteins occur throughout the plasmidome

The number of replicon types in bacteriocin-carrying plasmids suggests that they are phylogenetically diverse. Classifying diverse plasmids is difficult; recombination, insertions and co-integration means plasmid evolution cannot be explained with typical phylogenetic techniques. However, the use of approaches based on sequence similarity allows one to group plasmids into ‘communities’, which are clusters of related plasmids which share large percentages of DNA<sup>83</sup>. Using this approach, we are able to assign 73.7% of all *E. coli* plasmids into clusters with 5 or more members (Fig. 3D and Supplementary Fig. 6). Clusters are well defined and mixing between clusters is rare

(assortativity coefficient (a measure of edges shared between clusters, with 1 indicating no shared edges): 0.998). Consistent with our earlier replicon analyses, we find bacteriocin plasmids in 96/1148 plasmid clusters and they are well-dispersed across the *E. coli* plasmidome. Type I colicin plasmids tend to form clusters with plasmids encoding the same or closely related colicins. The larger type II plasmids exist in more diffuse clusters, which contain colicin, microcin and non-weaponry plasmids. We observed no mixing between iron-targeting bacteriocins and non-iron-targeting bacteriocins (assortativity coefficient 1), suggesting a limited shared evolutionary history. Interestingly, there are also many clusters where a single colicin plasmid occurs with non-colicinogenic plasmids implying a recent horizontal acquisition of a colicin gene from another plasmid.

Consistent with its functional importance, the association between bacteriocins and siderophores is not confined to a single plasmid cluster in our network analysis. Association of salmochelin and bacteriocins is observed in 14/19 clusters where salmochelin was present. More generally, salmochelin, aerobactin, Eit and Sit iron-uptake systems, are all significantly associated with clusters that encode iron-targeting bacteriocin plasmids (Fisher test: salmochelin: OR = 69.05  $p = 1.34 \times 10^{-15}$ , aerobactin: OR = 7.28  $p = 6.62 \times 10^{-4}$ , eit: OR = 16.71  $p = 1.1 \times 10^{-4}$ , sit: OR = 15.71,  $p = 1.46 \times 10^{-8}$ ).

Colicins, and most bacteriocins, are typically encoded as a toxin and downstream immunity protein. However, orphan immunity genes without a cognate toxin can protect a cell from incoming attacks without suffering the high cost of toxin production<sup>84</sup>. We find 243 plasmids that encode an orphan immunity gene, with more colicin immunity proteins than those for microcins (195 vs 53), which likely reflects the higher prevalence of colicins. Indeed, as is expected, the types of immunity orphan gene observed mirror the most common toxin domains carried by strains (Pore forming: 146, Microcin: 53, Peptidoglycan inhibiting: 46, tRNase: 2, DNase: 1). It is not known if orphan immunity genes occur due to natural selection against the high cost of toxin production (toxin loss) or because immunity genes are naturally selected to provide protection against toxin-wielding competitors (immunity gain). Distinguishing between toxin loss and immunity gain is important, because it can help decipher the role of bacteriocins in bacterial ecology. If bacteriocins evolve only because they can function as toxin-antitoxin systems that help to maintain plasmids, one expects to only find orphan immunity genes alongside recently degraded toxin genes within plasmid clusters containing bacteriocin plasmids (toxin loss). If bacteriocins are important agents in bacterial warfare, we would also expect cases where immunity genes are present in clusters without toxin genes, because picking up and maintaining immunity genes will confer protection from attacking bacteria (immunity gain).

136 of the orphan immunity plasmids that we observe occur in plasmid clusters which do not contain a bacteriocinogenic plasmid (across 24 distinct clusters). This pattern supports the immunity gain hypothesis and suggests that bacteriocin immunity, and therefore bacterial warfare via bacteriocins, is important for *E. coli*. One potential limitation of this analysis would be if we are missing the bacteriocinogenic plasmids that occur alongside orphan immunity plasmids due to insufficient sampling. However, there are many examples of clusters that contain multiple orphan immunity plasmids (range 2–58 orphan immunity plasmids per cluster) but no bacteriocinogenic plasmids, which is difficult to explain by sampling limitation (Fig. 3). Moreover, inspecting coding sequences from the four largest clusters of orphan immunity plasmids provides further evidence that the occurrence of orphans is not simply explained by them having a recently degraded toxin gene (Supplementary Fig. 7). The plasmids examined either completely lack a bacteriocin gene or they possess a heavily truncated bacteriocin gene, where this same truncation is found in multiple distinct plasmids in a cluster i.e. it has been maintained in plasmids of different sizes that were isolated

from different locations (Supplementary Fig. 8). These patterns, therefore, all support the idea that the orphan immunity genes are being independently maintained to protect against the toxins of other strains.

To further cement the conclusion that these bacteriocins are not simply plasmid addiction systems, we also analysed the plasmids for the presence of other toxin-antitoxin (TA) systems, which act inside the cell, and are known to be able to function as plasmid addiction systems, and are certainly not involved in bacterial warfare. We reasoned that, if the only function of bacteriocins is to stabilise plasmids, one expects fewer canonical toxin-antitoxin addiction systems on bacteriocin plasmids than other plasmids. Both presence and number of TA systems scales with plasmid size (GLM: Presence of TA system (binary) - Plasmid size (kbp),  $\beta = 0.0032$ ,  $p < 0.01$ ; TA system count (continuous) - Plasmid size (kbp),  $\beta = 0.0084$ ,  $p < 0.01$ ). However, the presence of a bacteriocin does not lower the probability of a TA system being on the same plasmid when controlling for plasmid size. In fact, non-bacteriocin plasmids have fewer canonical toxin-antitoxin systems than the bacteriocin plasmids, when accounting for plasmid size (Supplementary Fig. 9).

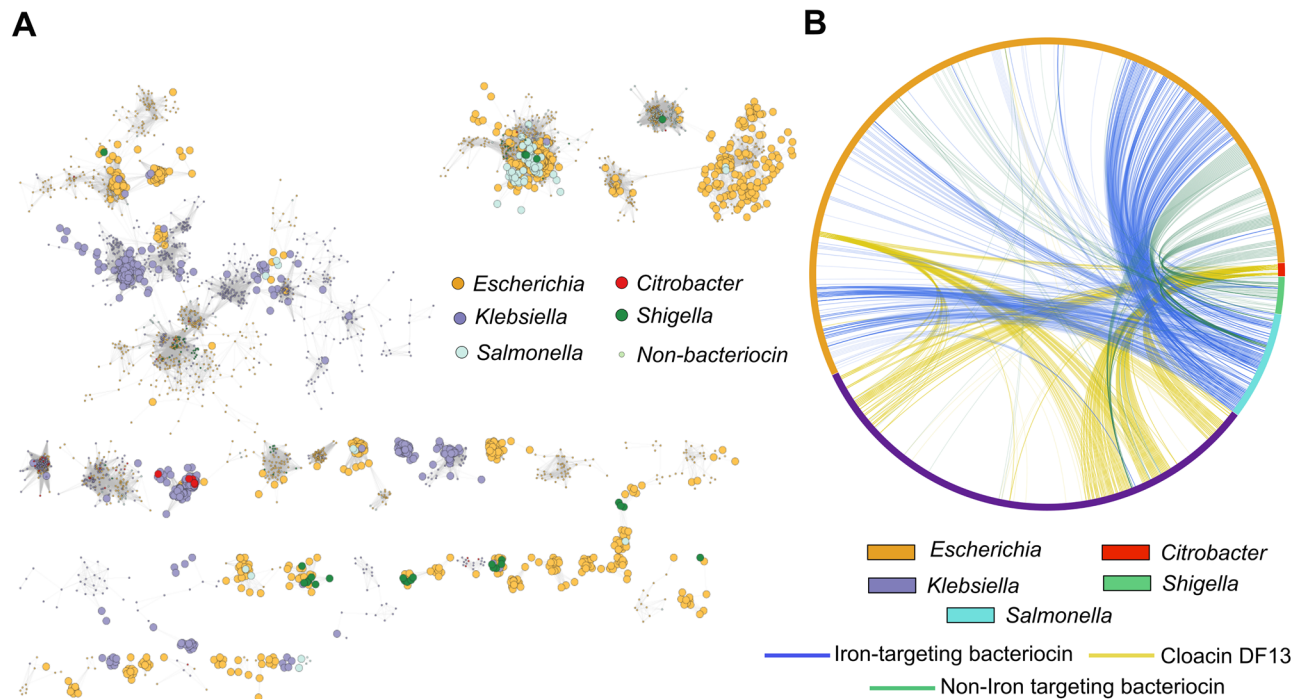
### Colicin and microcin genes are mobile among plasmids

Many bacteriocins are found in multiple clusters across our plasmid network, which suggests that either bacteriocin plasmids have evolved and diversified or bacteriocin genes are moving between plasmids. To better understand the processes at play, we built a phylogeny of one of the common colicins (Colla), which identifies multiple transitions between plasmid clusters (Fig. 4A). Comparing the plasmids associated with these transitions reveals examples where colicins sit in a large region that is shared between plasmids in different clusters, which is consistent with bacteriocin plasmids evolving and diversification. However, we also see cases where there is no significant region (>2kbp) of homology between plasmids (other than a shared colicin gene), which suggests that bacteriocins alone are moving between plasmids. Further consistent with this inter-plasmid mobility, many bacteriocin genes are flanked by two transposons which have the potential to transfer at the same time (i.e. composite transposons) and transport the DNA segment between them, including ColM, Collb, Collc, CloDF13, MccV, MJ25 and MccB17 (Fig. 4B). Capturing transposon movement between plasmids is difficult due to insufficient sampling depth and lack of longitudinal sampling. Despite this, we were able to identify events where a predicted composite transposon was shared to a high identity between two unrelated plasmids, which again suggests that the bacteriocins are moving between plasmids via composite transposons (Fig. 4C).

### Colicin plasmids are shared amongst Enterobacteriaceae

Unlike microcins, colicin-like proteins require no complex post-translational modifications or dedicated export system. This simplicity means that colicins have the potential to function in a wider range of species than microcins, and colicins are found throughout Enterobacteriaceae<sup>26,85</sup>. However, the degree to which colicin plasmids are shared between species and genera is not well understood. To investigate this, we used the same network-based approach as before to cluster the 11274 plasmids from 5 genera of Enterobacteriaceae (*Escherichia*, *Klebsiella*, *Salmonella*, *Citrobacter* and *Shigella*), in the plasmid (PLSDB) database (Fig. 5A). 8373 of these plasmids (74.3%) can be placed within one of 260 plasmid clusters that have five or more members (the assortativity coefficient of 0.998 again suggests well defined clusters). Many of these clusters contain plasmids from multiple genera (131/260), and edges between genera are common in the network (assortativity coefficient = 0.43), which is consistent with general ability of plasmids to move within Enterobacteriaceae. We then identified all colicins and related bacteriocins across *E. coli*, *Citrobacter*, *Shigella* (colicins), *K. pneumoniae* (klebicins) and *S. enterica*





**Fig. 5 | Colicin plasmids are shared among multiple genera within *Enterobacteriaceae*.** **A** Louvain clustering of all plasmids in the genera *Escherichia*, *Klebsiella*, *Salmonella*, *Citrobacter* and *Shigella*. For clarity, only plasmid clusters with contain at least 1 colicin plasmid are shown. Each circle represents an

individual plasmid and edges indicate a mash distance of < 0.03. Non-colicinogenic plasmids are shown as smaller circles. **B** All colicin plasmids identified in 5 different genera of Enterobacteriaceae are represented in a chord diagram. Connections show plasmids with  $\geq 99\%$  ANI shared between genera.

(salmocins) in the plasmid (PLSDB) database and mapped them to the clusters.

This analysis shows that 52 clusters contain at least one colicin plasmid and over 50% (28) of the colicin clusters contained plasmids from more than one genus (assortativity coefficient = 0.36). However, different colicin plasmids are distributed differently between the genera. The edges between *Escherichia*, *Klebsiella* and *Citrobacter* colicin plasmids are mostly between plasmids encoding Cloacin DF13. This bacteriocin hijacks the siderophore receptor LutA<sup>86</sup> of the aerobactin system, a siderophore system that is found in many *Enterobacteriaceae*, including *E. coli* and *Klebsiella* where it is an important virulence factor<sup>87</sup>. Edges between *Klebsiella* and other genera (*Salmonella* and *Shigella*) or involving non-Cloacin DF13 plasmids are relatively rare. Supplementary analyses suggest that the low rate of plasmid transfer with *Klebsiella* is not due to an inability to share plasmids between the species as closely related non-bacteriocin plasmids are found in *Salmonella* and *Klebsiella* strains (Supplementary Fig. 10). By contrast, edges between *Escherichia* and *Salmonella* colicinogenic plasmids are common, typically involving Colla and Collb plasmids and comprised 80% of all edges between species in our analysis (Fig. 5B and Supplementary Fig. 11). A dedicated phylogeny of Colla toxin and immunity genes from *E. coli* and *S. enterica* also suggests a high rate of transfer between these species (Supplementary Fig. 12). These patterns may reflect the niche in which *Salmonella* will often encounter *E. coli*, the inflamed mammalian gut where possession of an iron targeting Collb plasmid can provide a fitness advantage to blooming enterobacteria<sup>20,78</sup>.

## Discussion

Bacteria have evolved a diverse array of weapons that target competing strains, many of which have been extensively studied under laboratory conditions<sup>7,17,42</sup>. Here we have used the unrivalled genomic resources of *E. coli* to investigate the natural ecology and evolution of bacteriocins. The bacteriocins of *E. coli* occur both on the chromosome and on plasmids, and we find evidence that plasmid-encoded

bacteriocins can readily move between both strains and plasmids. Many of the plasmids that carry bacteriocins also carry AMR genes and virulence factors. This result suggests that problematic strains are commonly using weapons in order to establish and persist in communities.

In particular, we find that bacteriocins are associated with siderophores and other factors associated with growth under low-iron conditions. This pattern is driven by five bacteriocins: Colla, Collb, ColM, ColB and MccV, all weapons which themselves target iron uptake receptors (CirA, FepA and FhuA). Iron-limitation is common in hosts, especially during inflammation<sup>88</sup>. Under these conditions, bacteria upregulate the expression of iron uptake receptors, which increases their susceptibility to iron-uptake targeting weaponry<sup>20</sup>. The use of such weaponry, therefore, may give strains a fitness advantage under low iron conditions<sup>20</sup>, such as those found in the inflamed gut. From the gut, strains can then spread to extra-intestinal infection sights such as the urinary tract<sup>89–91</sup>. More generally, our findings suggest that iron-targeting bacteriocins, siderophores and other virulence factors can work in concert to help pathogens to both colonise diverse communities, such as the human microbiome and cause disease. Experimental tests are needed to confirm this hypothesis along with the potential for siderophores and iron-targeting bacteriocins to work in synergy to enable persistence of ExPEC strains in the microbiome<sup>92–94</sup>. However, consistent with these ideas, gut colonisation is an important step for the pathogenesis of extraintestinal *E. coli* in animal studies and extra-intestinal virulence factors can increase survival in the gut<sup>70,71,92–97</sup>. Disruption of gut colonisation represents a potential therapeutic strategy to prevent extra-intestinal *E. coli* infections<sup>98</sup>. Our work suggests that understanding the links between bacterial warfare and virulence will aid in this endeavour.

The function of bacterial weapons under natural conditions is surprisingly poorly understood. Though they are often potent killers under laboratory conditions, the in vivo activity of bacteriocins is less clear<sup>16,99–101</sup>. This pattern has led some to suggest that colicins function solely as selfish genetic elements, which prevent their loss from

bacterial hosts by intoxicating cells that lose them<sup>24,25</sup>. While colicins may indeed reduce plasmid loss from bacterial lineages<sup>24</sup>, our analyses do not support the view that this is their primary function. We found evidence that orphan immunity proteins are being maintained throughout the *E. coli* plasmidome, which is consistent with natural selection for protection against bacteriocin attack. In addition, we find that conventional toxin-antitoxin systems are abundant on bacteriocin plasmids, which again suggests that the bacteriocins are performing a separate function.

Our analyses, therefore, support the intuitive idea that bacteriocins function as weapons for bacteria to compete with other strains, albeit in a manner that can also benefit plasmids that carry them<sup>81</sup>. We also find that carriage of these weapons in *E. coli* is associated with particular ecologies, with a strong link between the common iron-receptor targeting colicins and virulence factors that help bacteria to cause disease. This link to virulence is further cemented by the observation that the bacteriocin plasmids of *E. coli* are commonly shared with the pathogen *S. enterica*. Overall, our work suggests that the bacteriocins of *E. coli* are both highly mobile and important for the ecology of dangerous strains.

## Methods

### *E. coli* genome dataset

*E. coli* genomes were accessed from the BV-BRC<sup>102</sup> database (formally known as PATRIC). Genomes were selected to ensure high quality complete genomes (Genome Status: Complete, Genome Quality: Good). Genomes of poor quality were removed by three methods. Firstly, any genome with an L50 > 1 (indicating the chromosome is split over multiple contigs). Secondly, the number of coding sequences versus the length of the genome was plotted. For an individual species this should be a linear relationship. We removed any genomes that deviated from the mean by >500 coding sequences<sup>103</sup> (Supplementary Fig. 13a). Finally, we used the panaroo-qc tool in the panaroo software to remove genomes with suspected contamination<sup>104</sup>. This tool uses mash to identify distances between the dataset and a database of refseq genomes. Multidimensional scaling allows visualisation of genomes which form outliers and are therefore possible contaminants (Supplementary Fig. 13b). The remaining 2601 genomes included samples isolated from 74 different countries, across six continents, with no single country representing more than 20% of the data. Genomes were sampled over 118 years between 1884–2022 and included genomes from 846 Bioproject Accessions and 104 publications. The dataset includes 438 different sequence types with 196 sequence types represented by more than 1 sequence (Supplementary Fig. 14). Genomes covered all known *E. coli* phylogroups (A: 990, B1: 482, B2: 374, C: 93, D: 234, E: 301, E/I: 1, F: 84, G:28, I: 4, Unknown: 10).

The pangenome of the 2601 *E. coli* genomes was calculated using Panaroo<sup>104</sup>. As colicins are present in populations at a low frequency we used the sensitive clean-mode parameter and set the core genome as genes present in ≥99% of genomes. Core genes were aligned with MAFFT and phylogeny calculated using RAxML-NG<sup>105</sup> with the Generalised time reversible model and gamma model of rate heterogeneity.

Virulence factors were identified using Abricate<sup>106</sup> and a database of curated *E. coli* virulence factors from VFDB<sup>107</sup> and literature<sup>108</sup>. AMR elements were identified using Resfinder<sup>109</sup>. Serotype was predicted using abricate and the EcOH dataset<sup>110</sup>. Sequence type was predicted using the *mst* software<sup>111,112</sup>. Phylogroup was predicted using the ClermonTyping tool<sup>113</sup>. All abricate searches required a coverage and identity >70%. Colicin and colicin immunity proteins were identified using BLASTp (Supplementary Tables 1 and 2) with open reading frames predicted using getORF<sup>114</sup>. An open reading frame was considered a colicin only with an *e* value < 5 × 10<sup>-5</sup> and coverage >80% of query length. Some colicins are very rare and though we they were included in our initial search, they were not identified even in our dataset of 2601 genomes (e.g. ColA, ColE4 and ColR, ColZ, ColE9).

Uropathogenic specific protein (Usp) shares close homology to colicins and was included in the search and hits to Usp were removed from analysis. Microcins were identified using a custom database of microcin genes and Abricate<sup>106</sup>. IncF RSTs were predicted using pMLST<sup>115</sup>.

### Pathotype prediction of *E. coli* genomes

*E. coli* exists as a non-pathogenic member of the microbiome but can cause intestinal (InPEC) and extraintestinal (ExPEC) infections. Different pathotypes within InPEC infections can be defined by a selection of virulence factors, whereas ExPEC infections are defined by isolation point (e.g. urinary tract, bloodstream). We used the presence of specific virulence factors often used as PCR markers<sup>48</sup> to define InPEC infections (Table 2) and scanned available metadata to identify *E. coli* isolated from non-intestinal body sites.

### Phylogenetic comparative analysis

To investigate the relationship between iron-targeting bacteriocins and different virulence factors, we used a Generalised Linear Mixed Model (MCMCglmm<sup>116</sup>) using the core genome phylogenetic covariance matrix to control for sampling. Chains were run with 100,000 iterations with a 10,000 burn-in and 30-step thinning. Uninformative flat priors were used (*V* = 1, *nu* = 0.02). Associations between binary traits such as *E. coli* pathotype and plasmids, were tested using binary Generalised Linear Mixed Models with different pathotypes as a binary response variable using the binaryPGLMM function in the Ape<sup>117</sup> package (e.g. Formula = Pathotype - Iron targeting Bacteriocin Plasmid, phy = Core gene phylogeny).

### Plasmid database

Plasmid sequences were accessed from PLSDB v2020\_11\_19, which contains plasmid sequences from NCBI and INSDC based on previously defined criteria<sup>118</sup>. The 5373 *E. coli* plasmids were annotated as *Escherichia coli* as the host species. Host species of each plasmid in the database was determined from NCBI metadata. Plasmid sequences <1000 bp were removed from analysis. Plasmid mobility was predicted using the mobtyper tool within Mob-Suite<sup>82</sup>.

### Identifying orphan immunity proteins

Bacteriocin immunity proteins are small proteins with homology to many antitoxins and enzymes. They are normally confirmed to be colicin immunity proteins by the presence of the cognate toxin, therefore making identifying orphan immunities difficult. To identify orphan immunity proteins, we first constructed a dataset of immunity protein genes ≤ 200 bp from a cognate toxin. We used this set of genes to scan against the PLSDB using blastp with strict parameters and only selected hits with a percentage identity >99%.

### Toxin-antitoxin system analysis

Hidden Markov Models for Toxin-Antitoxin systems were accessed from TAsmania<sup>119</sup> and PLSDB protein open reading frames were scanned using HMMer3<sup>120</sup> with a *e* value cutoff of < 5 × 10<sup>-5</sup>. A Toxin/Antitoxin system was defined as a toxin and antitoxin in contiguous open reading frames which had both passed the threshold. Associations between bacteriocins and Toxin/Antitoxin systems were performed using Generalised Linear Models and R.

### Plasmid network analysis

To form a plasmid network, ‘distances’ between plasmids were determined using Mash<sup>121</sup> with a sketch length of 5000, k-mer size of 13 and all other parameters set to default. To form a network, a threshold of 0.03 was applied to the mash distances with distances below this threshold forming the edges between plasmid nodes. Plasmid communities were estimated using the Louvain algorithm<sup>122</sup> to optimise network modularity<sup>83</sup>. The resulting network and communities were visualised using CytoScape<sup>123</sup>.

### Gene ontology enrichment analysis

Gene ontology enrichment analysis was performed on the *E. coli* genome datasets. First, the pangenome created using panaroo was annotated using interproscan (v5.57-90.0)<sup>124,125</sup> and the CDD, Pfam, Superfamily and TIGRFAM databases to identify Gene Ontology terms. GO abundances were normalised by plasmid size (kbp). GO term distribution was modelled as a poisson distribution. Significant differences in poisson means were calculated between bacteriocin and chromosomal and non-bacteriocin plasmids using R. *P* values were corrected using a Bonferroni correction.

### Plasmid backbone phylogeny

To generate a phylogeny of plasmid backbone genes for an individual plasmid cluster, we first calculated the pangenome of all plasmids within the cluster using panaroo, with the sensitive clean mode and the core gene cut off at 90% of plasmids. Core genes were then aligned using MAFFT<sup>126</sup> and phylogeny calculated using IQTree2<sup>127</sup> and ClonalFrameML<sup>128</sup> to account for recombination. The resulting phylogeny was visualised using iTOL<sup>129</sup>. Colicin phylogenies were calculated from start of the toxin to end of the immunity gene. Sequences were aligned using clustalo<sup>130</sup> and a phylogeny calculated using IQTree2.

### MEfinder

MobileElementFinder<sup>131</sup> was used to identify composite transposons on plasmids. Before scanning with MobileElementFinder, all plasmids were doubled (an extra copy of the sequence appended to the end) so composite transposons would be found across the plasmid break point. All PLSDb plasmids were then scanned for the presence of insertion sequences and transposons. Predicted transposons were then analysed for the presence of bacteriocins. Plasmids were visualised using gggenomes<sup>132</sup>.

### Prediction of Fur and LexA binding sites

Alignments of *E. coli* Fur and LexA binding sites were accessed from CollectTF<sup>133</sup>. Binding sites within plasmids were identified using FIMO<sup>134</sup> of the MEME<sup>135</sup> suite. Fur and LexA regulation were determined by the presence of a predicted binding site ( $p < 5 \times 10^{-5}$ ) within 200 bp of the bacteriocin gene start codon.

### Data availability

The genomic data used in this study is publicly available from BV-BRC using accessions listed in Supplementary Data 1. Plasmid data is available from PLSDb using accessions listed in Supplementary Data 2 and 3. Publicly available sequences used to identify bacteriocins from *E. coli* are listed in Supplementary Table 1. Publicly available sequences used to identify Colicin-like proteins from *Escherichia*, *Klebsiella*, *Salmonella*, *Citrobacter* and *Shigella*, are listed in Supplementary Table 2. Antimicrobial resistance genes and virulence factors were annotated using Abricate and the Resfinder (<http://genepi.food.dtu.dk/resfinder>) and *ecoli\_vf* datasets, ([https://github.com/phac-nml/ecoli\\_vf](https://github.com/phac-nml/ecoli_vf)). Hidden Markov Models for Toxin-Antitoxin Systems were accessed from TASmania (<https://shiny.bioinformatics.unibe.ch/apps/tasmania/>).

### References

- Hibbing, M. E. et al. Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8**, 15–25 (2010).
- Coyte, K. Z. & Rakoff-Nahoum, S. Understanding competition and cooperation within the mammalian gut microbiome. *Curr. Biol.* **29**, R538–R544 (2019).
- Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350**, 663–666 (2015).
- García-Bayona, L. & Comstock, L. E. Bacterial antagonism in host-associated microbial communities. *Science* **361**, eaat2456 (2018).
- Zhang, D. et al. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**, 18 (2012).
- Jamet, A. & Nassif, X. New players in the toxin field: polymorphic toxin systems in bacteria. *mBio* **6**, 1–8 (2015).
- Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The evolution and ecology of bacterial warfare. *Curr. Biol.* **29**, R521–R537 (2019).
- Alteri, C. J. et al. Subtle variation within conserved effector operon gene products contributes to T6SS-mediated killing and immunity. *PLOS Pathog.* **13**, e1006729 (2017).
- Speare, L. et al., Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proc. Natl. Acad. Sci. USA* **115**, E8528–E8537 (2018).
- Spragge, F. et al. Microbiome diversity protects against pathogens by nutrient blocking. *Science* **382**, ead3502 (2023).
- van der Waaij, D., Berghuis-de Vries, J. M. & Lekkerkerk-Van der Wees, J. E. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *Epidemiol. Infect.* **69**, 405–411 (1971).
- Bohnhoff, M., Drake, B. L. & Miller, C. P. Effect of streptomycin on susceptibility of intestinal tract to experimental *Salmonella* infection. *Proc. Soc. Exp. Biol. Med.* **86**, 132–137 (1954).
- Sorbara, M. T. & Pamer, E. G. Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. *Mucosal Immunol.* **12**, 1–9 (2019).
- Sassone-Corsi, M. et al. Microcins mediate competition among Enterobacteriaceae in the inflamed gut. *Nature* **540**, 280–283 (2016).
- Sana, T. G. et al. *Salmonella* Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. *Proc. Natl. Acad. Sci. USA* **113**, E5044–E5051 (2016).
- Samuels, A. N. et al. Deciphering the role of colicins during colonization of the mammalian gut by commensal *E. coli*. *Microorganisms* **8**, 664 (2020).
- Cascales, E. et al. Colicin biology. *Microbiol. Mol. Biol. Rev.* **71**, 158–229 (2007).
- Gratia, A. Sur un remarquable exemple d'antagonisme entre deux souches de colibacille. *CR Soc. Biol.* **93**, 1040–1041 (1925).
- Wang, L. et al. Colicins of *Escherichia coli* lead to resistance against the diarrhea-causing pathogen enterotoxigenic *E. coli* in pigs. *Microbiol. Spectr.* **10**, e01396–22 (2022).
- Nedialkova, L. P. et al. Inflammation fuels colicin Ib-dependent competition of *Salmonella* serovar Typhimurium and *E. coli* in enterobacterial blooms. *PLoS Pathog.* **10**, e1003844 (2014).
- Gillor, O., Giladi, I. & Riley, M. A. Persistence of colicinogenic in the mouse gastrointestinal tract. *BMC Microbiol.* **9**, 165 (2009).
- Feldgarden, M. & Riley, M. A. High levels of colicin resistance in *Escherichia coli*. *Evolution* **52**, 1270–1276 (1998).
- Kelstrup, J. & Gibbons, R. J. Inactivation of bacteriocins in the intestinal canal and oral cavity. *J. Bacteriol.* **99**, 888–890 (1969).
- Inglis, R. F. et al. The role of bacteriocins as selfish genetic elements. *Biol. Lett.* **9**, 20121173 (2013).
- Sudhakari, P. A. & Ramisetty, B. C. M. Modeling endonuclease colicin-like bacteriocin operons as 'genetic arms' in plasmid-genome conflicts. *Mol. Genet. Genom.* **297**, 763–777 (2022).
- Sharp, C. et al. Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using Hidden Markov Models. *PLOS Comput. Biol.* **13**, e1005652 (2017).
- Schneider, T. et al. Plant-made *Salmonella* bacteriocins salmocins for control of *Salmonella* pathogens. *Sci. Rep.* **8**, 4078 (2018).
- James, R. Molecular cloning and purification of klebicin B. *J. Gen. Microbiol.* **134**, 2525–2533 (1988).

29. Kleanthous, C. Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat. Rev. Microbiol.* **8**, 843–848 (2010).
30. Jeziorowski, A. & Gordon, D. M. Evolution of microcin V and colicin ia plasmids in *Escherichia coli*. *J. Bacteriol.* **189**, 7045–7052 (2007).
31. Hardy, K. G. et al. Two major groups of colicin factors: their evolutionary significance. *Mol. Gen. Genet.* **125**, 217–230 (1973).
32. Christenson, J. K. & Gordon, D. M. Evolution of colicin BM plasmids: the loss of the colicin B activity gene. *Microbiology* **155**, 1645–1655 (2009).
33. Gordon, D. M. & O'Brien, C. L. Bacteriocin diversity and the frequency of multiple bacteriocin production in *Escherichia coli*. *Microbiology* **152**, 3239–3244 (2006).
34. Cameron, A. et al. Bacteriocin occurrence and activity in *Escherichia coli* isolated from bovines and wastewater. *Toxins* **11**, 475 (2019).
35. Micenkova, L. et al. Human extraintestinal pathogenic *Escherichia coli* strains differ in prevalence of virulence factors, phylogroups, and bacteriocin determinants. *BMC Microbiol.* **16**, 218 (2016).
36. Micenkova, L. et al. Microcin determinants are associated with B2 phylogroup of human fecal *Escherichia coli* isolates. *Microbiologyopen* **5**, 490–498 (2016).
37. Smajs, D. et al. Bacteriocin synthesis in uropathogenic and commensal *Escherichia coli*: colicin E1 is a potential virulence factor. *BMC Microbiol.* **10**, 288 (2010).
38. Gordon, D. M., Riley, M. A. & Pinou, T. Temporal changes in the frequency of colicinogeny in *Escherichia coli* from house mice. *Microbiology* **144**, 2233–2240 (1998).
39. Housden, N. G. et al. Directional porin binding of intrinsically disordered protein sequences promotes colicin epitope display in the bacterial periplasm. *Biochemistry* **57**, 4374–4381 (2018).
40. Lensink, M. F. et al. Blind prediction of interfacial water positions in CAPRI. *Proteins* **82**, 620–632 (2014).
41. Calcuttawala, F. et al. Structural and functional insights into colicin: a new paradigm in drug discovery. *Arch. Microbiol.* **204**, 37 (2022).
42. Parker, J. K. & Davies, B. W. Microcins reveal natural mechanisms of bacterial manipulation to inform therapeutic development. *Microbiology* **168**, 001175 (2022).
43. Caballero-Flores, G., Pickard, J. M. & Núñez, G. Microbiota-mediated colonization resistance: mechanisms and regulation. *Nat. Rev. Microbiol.* **21**, 347–360 (2023).
44. Sequeira, R. P. et al. Commensal Bacteroidetes protect against colonization and transmission through IL-36 signalling. *Nat. Microbiol.* **5**, 304 (2020).
45. Ikuta, K. S. et al. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **400**, 2221–2248 (2022).
46. Murray, C. J. L. et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
47. Santos, A. C. et al. Diversity of hybrid- and hetero-pathogenic *Escherichia coli* and their potential implication in more severe diseases. *Front. Cell. Infect. Microbiol.* **10**, 339 (2020).
48. Robins-Browne, R. M. et al. Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Front. Cell. Infect. Microbiol.* **6**, 141 (2016).
49. Geurtsen, J. et al. Genomics and pathotypes of the many faces of *Escherichia coli*. *FEMS Microbiol. Rev.* **46**, fuac031 (2022).
50. Johnson, J. R. & Russo, T. A. Extraintestinal pathogenic *Escherichia coli*: the other bad *E. coli*. *J. Lab. Clin. Med.* **139**, 155–162 (2002).
51. Zhang, Y. & Lin, K. A phylogenomic analysis of *Escherichia coli*/Shigella group: implications of genomic features associated with pathogenicity and ecological adaptation. *BMC Evol. Biol.* **12**, 174 (2012).
52. Girardeau, J. P. et al. Association of virulence genotype with phylogenetic background in comparison to different seropathotypes of Shiga toxin-producing *Escherichia coli* isolates. *J. Clin. Microbiol.* **43**, 6098–6107 (2005).
53. Pugh, H. L. et al. *E. coli* ST11 (O157:H7) does not encode a functional AcrF efflux pump. *Microbiology* **169**, 001324 (2023).
54. Mellmann, A. et al. Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg. Infect. Dis.* **14**, 1287–1290 (2008).
55. Schmartz, G. P. et al. PLSDb: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res.* **50**, D273–D278 (2022).
56. Opal, S. M. et al. Aerobactin and -hemolysin as virulence determinants in *Escherichia coli* isolated from human blood, urine, and stool. *J. Infect. Dis.* **161**, 794–796 (1990).
57. Garénaux, A., Caza, M. & Dozois, C. M. The Ins and Outs of siderophore mediated iron uptake by extra-intestinal pathogenic *Escherichia coli*. *Vet. Microbiol.* **153**, 89–98 (2011).
58. Williams, P. H. & Carbonetti, N. H. Iron, siderophores, and the pursuit of virulence: independence of the aerobactin and enterochelin iron uptake systems in *Escherichia coli*. *Infect. Immun.* **51**, 942–947 (1986).
59. Gao, Q. et al. Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiol.* **12**, 143 (2012).
60. Fischbach, M. A. et al. The pathogen-associated *iroA* gene cluster mediates bacterial evasion of lipocalin 2. *Proc. Natl. Acad. Sci. USA* **103**, 16502–16507 (2006).
61. Althaus, E. W. et al. The ferric uptake regulation (Fur) repressor is a zinc metalloprotein. *Biochemistry* **38**, 6559–6569 (1999).
62. Griggs, D. W. & Konisky, J. Mechanism for iron-regulated transcription of the *Escherichia coli* *cir* gene: metal-dependent binding of fur protein to the promoters. *J. Bacteriol.* **171**, 1048–1054 (1989).
63. Hunt, M. D., Pettis, G. S. & McIntosh, M. A. Promoter and operator determinants for fur-mediated iron regulation in the bidirectional *fepA-fes* control region of the *Escherichia coli* enterobactin gene system. *J. Bacteriol.* **176**, 3944–3955 (1994).
64. Guan, L. et al. Development of a Fur-dependent and tightly regulated expression system in *Escherichia coli* for toxic protein synthesis. *BMC Biotechnol.* **13**, 25 (2013).
65. Heimer, S. R. et al. Autotransporter genes *pic* and *tsh* are associated with *Escherichia coli* strains that cause acute pyelonephritis and are expressed during urinary tract infection. *Infect. Immun.* **72**, 593–597 (2004).
66. Provence, D. L. & Curtiss, R. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. *Infect. Immun.* **62**, 1369–1380 (1994).
67. Ristow, L. C. & Welch, R. A. Hemolysin of uropathogenic *Escherichia coli*: a cloak or a dagger? *Biochim. Biophys. Acta* **1858**, 538–545 (2016).
68. Otto, B. R. et al. *Escherichia coli* hemoglobin protease autotransporter contributes to synergistic abscess formation and heme-dependent growth of *Bacteroides fragilis*. *Infect. Immun.* **70**, 5–10 (2002).
69. Andrade, F. F. et al. Colistin update on its mechanism of action and resistance, present and future challenges. *Microorganisms* **8**, 1716 (2020).
70. Yamamoto, S. et al. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by *Escherichia coli*. *J. Urol.* **157**, 1127–1129 (1997).

71. Nielsen, K. L. et al. Whole-genome comparison of urinary pathogenic *Escherichia coli* and faecal isolates of UTI patients and healthy controls. *Int. J. Med. Microbiol.* **307**, 497–507 (2017).
72. Mann, R. et al. Metabolic adaptations of uropathogenic *E. coli* in the urinary tract. *Front. Cell. Infect. Microbiol.* **7**, 241 (2017).
73. Arredondo-Alonso, S. et al. Plasmid-driven strategies for clone success in *Escherichia coli*. *Nat. Commun.* **16**, 2921 (2025).
74. Reid, C. J., Cummins, M. L. & Djordjevic, S. P. Major F plasmid clusters are linked with ColV and pUTI89-like marker genes in bloodstream isolates of *Escherichia coli*. *BMC Genom.* **26**, 57 (2025).
75. Lipworth, S. et al. The plasmidome associated with Gram-negative bloodstream infections: A large-scale observational study using complete plasmid assemblies. *Nat. Commun.* **15**, 1612 (2024).
76. Carattoli, A. et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).
77. Villa, L. et al. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *J. Antimicrob. Chemother.* **65**, 2518–2529 (2010).
78. Stecher, B. et al. Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc. Natl. Acad. Sci. USA* **109**, 1269–1274 (2012).
79. Muriana, P. M. & Klaenhammer, T. R. Conjugal transfer of plasmid-encoded determinants for bacteriocin production and immunity in *Lactobacillus acidophilus* 88. *Appl. Environ. Microbiol.* **53**, 553–560 (1987).
80. Wescombe, P. A. et al. Megaplasmids encode differing combinations of lantibiotics in *Streptococcus salivarius*. *Antonie Leeuwenhoek* **90**, 269–280 (2006).
81. Granato, E. T. et al. Horizontal gene transfer of molecular weapons can reshape bacterial competition. *PLoS Biol.* **23**, e3003095 (2025).
82. Robertson, J. & Nash, J. H. E. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.* **4**, e000206 (2018).
83. Matlock, W. et al. Genomic network analysis of environmental and livestock F-type plasmid populations. *ISME J.* **15**, 2322–2335 (2021).
84. Kennedy, N. W. & Comstock, L. E. Mechanisms of bacterial immunity, protection, and survival during interbacterial warfare. *Cell Host Microbe* **32**, 794–803 (2024).
85. Baquero, F. et al. Microcins in enterobacteriaceae: peptide antimicrobials in the eco-active intestinal chemosphere. *Front Microbiol* **10**, 2261 (2019).
86. Krone, W. J. A. et al. Plasmid-determined cloacin Df13-susceptibility in enterobacter-cloacae and *Klebsiella-edwardsii*—identification of the cloacin-Df13 aerobactin outer-membrane receptor proteins. *Antonie Leeuwenhoek* **51**, 203–218 (1985).
87. Russo, T. A. et al. Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous). *Infect. Immun.* **82**, 2356–2367 (2014).
88. Ganz, T. & Nemeth, E. Iron homeostasis in host defence and inflammation. *Nat. Rev. Immunol.* **15**, 500–510 (2015).
89. Robinson, A. E., Heffernan, J. R. & Henderson, J. P. The iron hand of uropathogenic *Escherichia coli*: the role of transition metal control in virulence. *Future Microbiol.* **13**, 745–756 (2018).
90. Chouikha, I. et al. Differential expression of iutA and ibeA in the early stages of infection by extra-intestinal pathogenic *E. coli*. *Microbes Infect.* **10**, 432–438 (2008).
91. Feldmann, F. et al. The salmochelin siderophore receptor IroN contributes to invasion of urothelial cells by extraintestinal pathogenic *Escherichia coli* in vitro. *Infect. Immun.* **75**, 3183–3187 (2007).
92. Le Gall, T. et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol. Biol. Evol.* **24**, 2373–2384 (2007).
93. Nowrouzian, F. L., Wold, A. E. & Adlerberth, I. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J. Infect. Dis.* **191**, 1078–1083 (2005).
94. Condamine, B. et al. Strain phylogroup and environmental constraints shape *Escherichia coli* dynamics and diversity over a 20-year human gut time series. *ISME J.* **19**, wrae245 (2025).
95. Magruder, M. et al. Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. *Nat. Commun.* **10**, 5521 (2019).
96. Nhu, N. T. K. et al. High-risk clones that cause neonatal meningitis and association with recrudescence infection. *Elife* **12**, RP91853 (2024).
97. Tchesnokova, V. L. et al. Pandemic uropathogenic fluoroquinolone-resistant *Escherichia coli* have enhanced ability to persist in the gut and cause bacteriuria in healthy women. *Clin. Infect. Dis.* **70**, 937–939 (2020).
98. Tariq, R. et al. Fecal microbiota transplantation for recurrent infection reduces recurrent urinary tract infection frequency. *Clin. Infect. Dis.* **65**, 1745–1747 (2017).
99. Schulz, S. et al. Broad and efficient control of major foodborne pathogenic strains of *Escherichia coli* by mixtures of plant-produced colicins. *Proc. Natl. Acad. Sci. USA* **112**, E5454–E5460 (2015).
100. Hrala, M. et al. *Escherichia coli* strains producing selected bacteriocins inhibit porcine enterotoxigenic *Escherichia coli* (ETEC) under both in vitro and in vivo conditions. *Appl. Environ. Microbiol.* **87**, e03121–20 (2021).
101. Cutler, S. A. et al. Dietary inclusion of colicin e1 is effective in preventing postweaning diarrhea caused by F18-positive *Escherichia coli* in pigs. *Antimicrob. Agents Chemother.* **51**, 3830–3835 (2007).
102. Olson, R. D. et al. Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res.* **51**, D678–D689 (2023).
103. Horesh, G. et al. A comprehensive and high-quality collection of *Escherichia coli* genomes and their genes. *Microb. Genom.* **7**, 000499 (2021).
104. Tonkin-Hill, G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
105. Kozlov, A. M. et al. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
106. Seemann, T. *Abricate*. Github, <https://github.com/tseemann/abricate>.
107. Chen, L. et al. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).
108. National Microbiology Laboratory, *Ecoli\_vf*. [https://github.com/phac-nml/ecoli\\_vf/](https://github.com/phac-nml/ecoli_vf/).
109. Zankari, E. et al. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).
110. Ingle, D. J. et al. In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O: H serotype combinations within and between pathogenic lineages. *Microb. Genom.* **2**, e000064 (2016).
111. Jolley, K. A. & Maiden, M. C. J. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinform.* **11**, 595 (2010).
112. Seemann, T. *mlst*. Github <https://github.com/tseemann/mlst>.
113. Beghain, J. et al. ClermonTyping: an easy-to-use and accurate method for genus strain phylotyping. *Microb. Genom.* **4**, e000192 (2018).
114. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).

115. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, 124 (2018).
116. Hadfield, J. D. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).
117. Paradis, E. & Schliep, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
118. Orlek, A. et al. Ordering the mob: Insights into replicon and MOB typing schemes from analysis of a curated dataset of publicly available plasmids. *Plasmid* **91**, 42–52 (2017).
119. Akarsu, H. et al. TASmania: a bacterial toxin-antitoxin systems database. *PLOS Comput. Biol.* **15**, e1006946 (2019).
120. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
121. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
122. Blondel, V. D. et al. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
123. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
124. Blum, M. et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).
125. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
126. Katoh, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
127. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 2461–2461 (2020).
128. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).
129. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
130. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
131. Johansson, M. H. K. et al. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* **76**, 101–109 (2021).
132. Hackl, T. & Ankenbrand, M. *gggenomes: a grammar of graphics for comparative genomics*.
133. Kiliç, S. et al. CollecTF: a database of experimentally validated transcription factor-binding sites in Bacteria. *Nucleic Acids Res.* **42**, D156–D160 (2014).
134. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
135. Bailey, T. L. et al. The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
136. Destoumieux-Garzon, D. et al. The iron-siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25: role of the microcin Val11-Pro16 beta-hairpin region in the recognition mechanism. *Biochem. J.* **389**, 869–876 (2005).
137. Semenova, E. et al. Structure-activity analysis of microcinJ25: distinct parts of the threaded lasso molecule are responsible for interaction with bacterial RNA polymerase. *J. Bacteriol.* **187**, 3859–3863 (2005).
138. Palmer, J. D. et al. Microcin H47: a class IIB microcin with potent activity against multidrug resistant enterobacteriaceae. *ACS Infect. Dis.* **6**, 672–679 (2020).
139. Vassiliadis, G. et al. Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. *Anti-microb. Agents Chemother.* **54**, 288–297 (2010).
140. Wang, K.-K. A. et al. Class IIB microcin MccM interferes with oxidative phosphorylation in *Escherichia coli*. *ACS Chem. Biol.* **19**, 1953–1962 (2024).
141. Donnenberg, M. S. et al. The role of the eae gene of enterohemorrhagic *Escherichia coli* in intimate attachment in vitro and in a porcine model. *J. Clin. Investig.* **92**, 1418–1424 (1993).
142. Hale, T. L. Genetic basis of virulence in *Shigella* species. *Microbiol. Rev.* **55**, 206–224 (1991).
143. Johnson, A. M. et al. Heat-labile enterotoxin promotes *Escherichia coli* adherence to intestinal epithelial cells. *J. Bacteriol.* **191**, 178–186 (2009).
144. Wang, H. et al. Heat-stable enterotoxins of enterotoxigenic *Escherichia coli* and their impact on host immunity. *Toxins* **11**, 24 (2019).
145. Hunt, J. M. Shiga toxin-producing *Escherichia coli* (STEC). *Clin. Lab. Med.* **30**, 21 (2010).
146. Brunder, W., Schmidt, H. & Karch, H. EspP, a novel extracellular serine protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol. Microbiol.* **24**, 767–778 (1997).
147. Gonzalez-Escalona, N. et al. Draft genome sequences of two O104:H21 *Escherichia coli* isolates causing hemorrhagic colitis during a 1994 Montana outbreak provide insight into their pathogenicity. *Genome Announc.* **1**, e00805–e00813 (2013).
148. Doughty, S. et al. Identification of a novel fimbrial gene cluster related to long polar fimbriae in locus of enterocyte effacement-negative strains of enterohemorrhagic *Escherichia coli*. *Infect. Immun.* **70**, 6761–6769 (2002).
149. Leyton, D. L. et al. Transfer region of pO113 from enterohemorrhagic *Escherichia coli*: similarity with R64 and identification of a novel plasmid-encoded autotransporter, EpeA. *Infect. Immun.* **71**, 6307–6319 (2003).
150. Paton, A. W. et al. Characterization of Saa, a novel auto-agglutinating adhesin produced by locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans. *Infect. Immun.* **69**, 6999–7009 (2001).
151. Paton, A. W. & Paton, J. C. *Escherichia coli* subtilase cytotoxin. *Toxins* **2**, 215–228 (2010).
152. Sack, D. A. et al. Diarrhoea associated with heat-stable enterotoxin-producing strains of *Escherichia coli*. *Lancet* **2**, 239–241 (1975).
153. Mussio, P. et al. Phenotypic and genotypic characterization of Shiga toxin-producing *Escherichia coli* strains recovered from bovine carcasses in Uruguay. *Front. Microbiol.* **14**, 1130170 (2023).
154. Smith, P. et al. High-quality whole-genome sequences for 21 enterotoxigenic *Escherichia coli* strains generated with PacBio sequencing. *Genome Announc.* **6**, e01311–e01317 (2018).
155. Clark, C. A., Heuzenroeder, M. W. & Manning, P. A. Colonization factor antigen CFA/IV (PCF8775) of human enterotoxigenic *Escherichia coli*: nucleotide sequence of the CS5 determinant. *Infect. Immun.* **60**, 1254–1257 (1992).
156. Helander, A., Hansson, G. C. & Svennerholm, A.-M. Binding of enterotoxigenic *Escherichia coli* to isolated enterocytes and intestinal mucus. *Microb. Pathogenesis* **23**, 335–346 (1997).
157. Nishi, J. et al. The export of coat protein from enteroaggregative *Escherichia coli* by a specific ATP-binding cassette transporter system. *J. Biol. Chem.* **278**, 45680–45689 (2003).

158. Flores-Sanchez, F. et al. Pic protein from enteroaggregative *E. coli* induces different mechanisms for its dual activity as a mucus secretagogue and a mucinase. *Front. Immunol.* **11**, 564953 (2020).
159. Yen, H., Sugimoto, N. & Tobe, T. Enteropathogenic *Escherichia coli* uses NleA to inhibit NLRP3 inflammasome activation. *PLoS Pathog.* **11**, e1005121 (2015).
160. McFarland, N. et al. Recurrent seasonal outbreak of an emerging serotype of Shiga toxin-producing *Escherichia coli* (STEC O55:H7 Stx2a) in the south west of England, July 2014 to September 2015. *Eurosurveillance* **22**, 30610 (2017).
161. Grys, T. E. et al. The StcE protease contributes to intimate adherence of enterohemorrhagic *Escherichia coli* O157:H7 to host cells. *Infect. Immun.* **73**, 1295–1303 (2005).
162. Ho, T. D. et al. Type 2 secretion promotes enterohemorrhagic *Escherichia coli* adherence and intestinal colonization. *Infect. Immun.* **76**, 1858–1865 (2008).
163. Che, X. et al. Involvement of IbeA in meningitic *Escherichia coli* K1-induced polymorphonuclear leukocyte transmigration across brain endothelial cells. *Brain Pathol.* **21**, 389–404 (2011).
164. Uhlich, G. A. KatP contributes to OxyR-regulated hydrogen peroxide resistance in *Escherichia coli* serotype O157: H7. *Microbiology* **155**, 3589–3598 (2009).

## Acknowledgements

Thank you to Elisa Granato, Olivier Cunrath, Erik Bakkeren and Sean Booth for feedback and comments. K.R.F. is funded by European Research Council Grant 787932 and Wellcome Trust Investigator award 209397/Z/17/Z.

## Author contributions

Conceptualisation, C.S. and K.R.F.; Methodology, C.S.; Investigation, C.S.; Writing—Original Draft, C.S. and K.R.F.; Writing—Review and Editing, C.S. and K.R.F.; Funding Acquisition, K.R.F.; Resources, C.S. and K.R.F.; Supervision, K.R.F.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-64363-5>.

**Correspondence** and requests for materials should be addressed to Connor Sharp or Kevin R. Foster.

**Peer review information** *Nature Communications* thanks Ruichao Li, David Smajs, Steven Djordjevic and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025