



## Ecological and genetic determinants of plasmid distribution in *Escherichia coli*.

Journal:	<i>Environmental Microbiology and Environmental Microbiology Reports</i>
Manuscript ID	EMI-2016-0603.R1
Manuscript Type:	EMI - Research article
Journal:	Environmental Microbiology
Date Submitted by the Author:	n/a
Complete List of Authors:	Medaney, Frances; Royal Holloway University of London School of Biological Sciences Ellis, Richard; Animal and Plant Health Agency Raymond, Ben; University of Exeter, College of Life and Environmental Sciences
Keywords:	antibiotic resistance, mobile elements, population structure, plasmid profile, replicon typing

SCHOLARONE™  
Manuscripts

Ecological and genetic determinants of plasmid distribution in *Escherichia*

2 *coli*.

4 Frances Medaney<sup>1</sup>, Richard J. Ellis<sup>2</sup>, Ben Raymond<sup>1,3\*</sup>.

<sup>1</sup>School of Biological Science, Royal Holloway University of London, Egham, Surrey,

6 TW20 0EX, UK. <sup>2</sup>Specialist Scientific Support Department, Animal and Plant Health

Agency, APHA Weybridge, Addlestone, Surrey, KT15 3NB, UK. <sup>3</sup>University of Exeter,

8 Penryn campus, Penryn, Cornwall, TR10 9FE, UK. \*corresponding author,

b.raymond@exeter.ac.uk; tel +441326259085.

10

12 *Running title:*

The fundamental ecology of *E. coli* plasmids

14

16

*Originality Significance statement:*

2 Large conjugative plasmids are clearly very important as carriers of virulence and  
antimicrobial resistance genes. Nevertheless the vast majority of research on plasmids  
4 has focussed on genetics, while study of plasmid ecology and evolutionary biology is in  
its infancy. Plasmid ecology and evolution are fundamentally important as they can  
6 explain how and why plasmids move through bacterial communities and why they  
might be retained or lost by particular lineages.

8       Theoretical and experimental evolution studies on plasmid evolutionary ecology  
are relatively common but we have little basis for knowing whether the ecological  
10 assumptions these studies make are realistic. The motivation for this study was the lack  
data on the fundamental field ecology of large plasmids. Previous studies of plasmid  
12 “ecology” have typically been either fishing for novel resistance plasmids, or surveying  
plasmids in populations recently exposed to antibiotics, are under-replicated or take no  
14 account of bacterial population structure. High throughput next generation sequencing  
approaches have also not been particularly helpful as they tend to lose vital data on  
16 bacteria plasmid associations.

The fundamental field ecology approach we have taken is original as it  
18 incorporates a fully replicated hierarchical design that characterizes plasmid profiles  
within a bacterial population genetic structure; this allows us to make inferences about  
20 how both ecological and genetic factors can shape plasmid profiles. Novel findings  
include the effect of bacterial diversity on the diversity of plasmids found in a  
22 community; the fact that different plasmid replicons have different host ranges within  
species; the fact that the presence of large replicons can increase or decrease the  
24 likelihood of carriage of other replicons; the effect of site on the probability of carriage  
and the substantial and unexplained effect of bacterial genetic background on the

probability that isolates will have no large plasmids. As an original basic study of  
2 plasmid field ecology this work generates a large number of new hypotheses about how  
plasmids and bacteria might interact with each other

4

### Summary

6 Bacterial plasmids are important carriers of virulence and antibiotic resistance genes.  
Nevertheless, we know little of the determinants of plasmid distribution in bacterial  
8 populations. Here we explored the factors affecting the diversity and distribution of the  
large plasmids of *Escherichia coli* in populations of cattle grazing on semi-natural  
10 grassland, a set of populations with low frequencies of antibiotic resistance genes.  
Critically, we also characterized the population genetic structure of bacterial hosts. This  
12 revealed structured *E. coli* populations with high diversity between sites and individuals  
but low diversity within cattle hosts. Plasmid profiles, however, varied considerably  
14 within the same *E. coli* genotype. Both ecological and genetic factors affected plasmid  
distribution: plasmid profiles were affected by site, *E. coli* diversity, *E. coli* genotype and  
16 the presence of other large plasmids. Notably 3/26 *E. coli* serotypes accounted for half  
the observed plasmid-free isolates indicating that within species variation can  
18 substantially affect carriage of the major conjugative plasmids. The observed population  
structure suggest that most of the opportunities for within species plasmid transfer  
20 occur between different individuals of the same genotype and support recent  
experimental work indicating that plasmid-host coevolution, and epistatic interactions  
22 on fitness costs are likely to be important in determining occupancy.

24 Keywords: antibiotic resistance, mobile elements, population structure, plasmid profile,  
replicon typing

## Introduction

2           The persistence, replication and spread of bacterial plasmids is a major focus of  
research because so many plasmids carry virulence and antibiotic resistance genes  
4 (Carattoli, 2011; Dröge *et al.*, 2000; Hopkins *et al.*, 2006; Johnson *et al.*, 2007; Marcadé  
*et al.*, 2009; Pallecchi *et al.*, 2007). Despite their clinical importance, the ecology and  
6 evolutionary biology of plasmids has received relatively little attention. There are a  
range of contrasting views of the relationship between plasmids and their bacterial  
8 hosts. We can think of plasmids as genetic parasites, costly junk DNA that offset the  
costs they impose on their hosts by conjugating and infecting new hosts (Lili *et al.*,  
10 2007; Modi and Adams, 1991; Smith, 2010). Plasmids can also be seen as carriers of  
intermittently selected traits (Eberhard, 1990; Lilley and Bailey, 1997b), or indeed as  
12 “domesticated” accessory genomes that may end up losing conjugational ability  
entirely, such as the plasmid pXO2 that encodes encapsulation genes for *Bacillus*  
14 *anthracis* (Hu *et al.*, 2009).

Understanding how plasmids can persist in bacterial populations over the long  
16 term is an evolutionary dilemma referred to as the “plasmid paradox”. The persistence  
of plasmids and their association with periodically beneficial traits is a complex and  
18 developing theoretical area. Theory suggests that if plasmids are on average beneficial  
to their hosts, because of the traits they carry, then selection should favour bacteria that  
20 transfer these beneficial traits to the chromosome, as they can then avoid the costs of  
plasmid carriage (Bergstrom *et al.*, 2000). Conversely lack of selection for non-essential  
22 plasmid encoded traits can lead to their loss from the plasmid itself (Dahlberg and Chao,  
2003). In the view of some, the costs incurred by plasmids, and segregational loss  
24 during vertical transfer, means that it is difficult for plasmids that do not encode  
beneficial traits to persist (Bergstrom *et al.*, 2000).

Emerging evidence is helping to elucidate why plasmids are common and tend to be associated with particular classes of traits. Studies of conjugation rates in biofilms, diverse communities or semi-natural settings have shown that these rates can be much higher than previously thought (Dionisio *et al.*, 2002; Luo *et al.*, 2005; Sorensen *et al.*, 2005; Vilas-Bôas *et al.*, 2008). In addition, the ability of bacterial chromosomes to rapidly adapt and ameliorate or negate the fitness costs of plasmid carriage seems to be a common pattern in ensuring plasmid persistence (Dahlberg and Chao, 2003; Dionisio *et al.*, 2005; Harrison *et al.*, 2015; San Millan *et al.*, 2014). While these varied viewpoints may reflect real variation in bacterial plasmid associations it is worthwhile asking how applicable these views are to natural populations of bacteria and hosts. For example, what proportion of plasmids carry intermittently selected traits? Are bacterial lineages associating with plasmids for long enough to lead to substantial selection pressure on costs of carriage? If bacteria can adapt to ameliorate the fitness costs of plasmid carriage what then limits the prevalence and host range of plasmids? Exploring their fundamental ecology can help provide basic information for evaluating current theory but also help us develop novel theory to explain the variation that we see.

Our current understanding of plasmid ecology has been biased by the common means of collection: most plasmids are collected using selective agents that isolate bacteria carrying particular resistance genes, typically for antibiotics, heavy metals or other pollutants (Dierikx *et al.*, 2012; Lilley and Bailey, 1997a). Plasmid multi locus sequence typing (pMLST) has been used to identify the prevalence of resistance plasmids in livestock, and has demonstrated the presence of identical extended spectrum beta-lactamase (ESBL) genes and plasmids in livestock and humans, suggesting the transfer of bacteria and plasmids through the food chain (Dierikx *et al.*, 2010; Dierikx *et al.*, 2012; Hordijk *et al.*, 2013; Leverstein-van Hall *et al.*, 2011).

Another common approach is to recover plasmids from environment that may recently  
2 have been subject to strong selection for antibiotic resistance (Dierikx *et al.*, 2010;  
Dierikx *et al.*, 2012; Dröge *et al.*, 2000). However, these studies are not usually designed  
4 to assess bacterial population structure and so sample sizes can be small, and herd and  
host animal replication low, (Dierikx *et al.*, 2010; Hordijk *et al.*, 2013). In addition,  
6 bacterial populations may have low diversity if they have recently been subject to strong  
selective sweeps, and we are left knowing little about the distribution and ecology of  
8 plasmids that do not carry these intermittently selected traits. Exceptions, i.e. plasmid  
studies not focussed on resistance genes, tend to be from the older literature, when  
10 studies were limited to low replication and examined very broad species / plasmid  
associations (Datta and Hughes, 1983) or are modern deep-sequencing approaches  
12 targetting small circular molecules, in which bacterial host-association information is  
typically lost (Kav *et al.*, 2012).

14 Thus, despite years of study of plasmids and their genetics we have few data on  
how bacterial hosts and ecological factors determine distribution. Importantly, this may  
16 mean that we do not know what will happen to plasmids after selection for resistance  
has ceased. The aims of this study were to explore the fundamental ecology of  
18 *Escherichia coli* plasmids in a natural context, without strong selection, and to explore  
how variation in bacterial host populations, particularly genetic diversity and  
20 abundance, could affect the diversity and distribution of plasmid carriage. We selected a  
study system based on isolating culturable strains from fresh bovine faecal matter  
22 produced by cattle grazing in unimproved semi-natural grassland, on either common  
land or land managed for conservation. We anticipated that these outdoor reared  
24 animals would maintain diverse enteric populations that had not been subject to strong  
selection for antibiotic resistance. This study system permitted a hierarchical sampling

strategy (plasmids within bacteria within pats within fields) that could be replicated  
2 across multiple sites with independent cattle populations.

#### 4 **Results**

Diversity and antibiotic resistance in the *E. coli* populations.

6 We characterized the bacterial population in order to explore levels of antibiotic  
resistance; to confirm that our bacteria were comprised of distinct populations and to  
8 assess the distribution of bacterial diversity between sites and vertebrate hosts. In total  
24483 colonies were isolated from 84 cowpats across 9 sites. Of these, 527 isolates were  
10 genotyped and screened for plasmid replicons and antibiotic susceptibility. Populations  
were diverse, strongly structured and characterized by a low frequency of resistance  
12 isolates. Of the 527 isolates just 14 (less than 3%) were resistant to any of the  
antibiotics tested: 9 isolates were ampicillin resistant, 4 were chloramphenicol resistant,  
14 and just 1 was resistant to cefotaxime. No multi-resistant isolates were found. Resistant  
isolates were found in 5 of the 9 sites, with four sites resistance-free.

16 Bacterial populations were highly aggregated with the majority of the population  
concentrated in relatively few samples; just 7/84 pats contributed more than 50% of the  
18 total colony counts. Within each pat, there was relatively low within species diversity,  
the vast majority of pats contained one or two H-antigen genotypes (Fig 1A). However,  
20 there was opportunity for plasmid transfer within genotypes within bovine hosts. We  
examined the range in numbers of replicons between isolates for pats that had at least 4  
22 isolates of the same serotype. Most commonly plasmid carriage was not at fixation, in  
37 pats there was heterogeneous plasmid carriage within bacterial genotypes, while  
24 only 20 pats showed uniform plasmid profiles (Fig 1B).



Different sites contained distinctive aggregations of *E. coli* genotypes (Fig 1C) confirming that these were good biological replicates. Hierarchical *F*-statistics were calculated to assess the contribution of sites, and pats within sites to the *E. coli* population structure: Both pat within site ( $F_{\text{pat/site}} = 0.457, p < 1 \times 10^{-4}$ ), and site ( $F_{\text{site/total}} = 0.197, p < 1 \times 10^{-4}$ ) were found to have strong effects on the population structure of *E. coli* (10,000 permutations). In other words, 20% of the observed genotype variation can be explained by site, and 46% of this variation can be explained between pats within site, corresponding to great and very great genetic differentiation respectively (Holsinger and Weir, 2009; Wright, 1978).

*E. coli* abundance does not predict plasmid diversity

We explored whether larger *E. coli* populations could support a greater variety of plasmid replicons. In order to test this, regression was conducted with total *E. coli* abundance against plasmid replicon diversity (Shannon diversity). When the data was pooled by site, *E. coli* abundance was not a good predictor of plasmid replicon diversity ( $F_{1,7} = 2.81, P = 0.14$ ; Figure 2A). Abundance was also not a good predictor of plasmid diversity within pat, based on an analysis of faecal samples with good sampling of *E. coli* ( $N = 7$  or more isolates) ( $F_{1,53} = 0.35, P = 0.56$ , Figure 2B).

Replicons are not evenly distributed across *E. coli* subgroups

The majority of *E. coli* isolates contained at least one plasmid replicon; just 26 % (136 isolates) had no detectable plasmids. Of the 18 possible replicons picked up by the screen 8 were detected. There was one clear dominant type, IncFIB, which was found in 352 isolates (67%). Of the other replicon types identified, IncFIA and IncP were common (> 100 isolates); IncFIC and IncI1 were moderately common (> 30 isolates); and IncB/O,

IncK/B and IncY were rare (< 10 isolates). The full plasmid distribution and H-antigen  
2 data sets are available on Dryad (link to be confirmed after acceptance of MS).

Figure 3A highlights the non-random distribution of replicons across *E. coli*  
4 subgroups. Flagellin sequences were used to allocate isolates to the four major flagellin  
gene phylogenetic groups (Ec1a, Ec1b, Ec2, Se2) (Wang *et al.*, 2003). Significantly more  
6 replicons than would be expected by chance were found for: IncY in Ec1a; IncP in Ec2;  
I1, IncFIA and IncFIC in Se2. The replicon distribution for subgroup Se2 is particularly  
8 striking, as this is the smallest subgroup (n = 72). Three replicon types are over-  
represented in Se2, including IncFIC. Significantly reduced carriage of replicons than  
10 expected by chance were found for: IncI1, IncFIA and IncFIC replicons in serotype group  
Ec1a; for IncY and IncP in group Ec1b and for IncP in Se2.

12 Given that particular groups of *E. coli* serotypes were enriched for different  
plasmid replicons we might expect that increased within species diversity would lead to  
14 increased plasmid diversity. At the site level, H serotype diversity is a reasonably good  
predictor of plasmid diversity ( $F_{1,7} = 6.23$ ,  $R^2 = 0.41$ ,  $P = 0.041$ , Figure 3B). This  
16 relationship appeared to be independent of the effects of *E. coli* abundance since *E. coli*  
abundance and H serotype diversity were uncorrelated (Figure 3C,  $F_{1,7} = 1.09$ ,  $P = 0.33$ ).  
18 Variation in H serotype diversity, and overall site effects both significantly affected  
plasmid diversity at the pat level (Figure 3D,  $F_{1,51} = 13.8$ ,  $P = 0.0005$ , and  $F_{8,52} = 4.02$ ,  $P =$   
20 0.0009) and together explained 54% of the variation in plasmid diversity.

Plasmid replicons were also not randomly distributed with respect to each other.  
22 We examined whether the three most common replicons, IncP, IncF1A, and IncF1B  
associated independently. We found that IncP was strongly associated with IncF1B  
24 replicons much more commonly than we would expect by chance (Fig 4A, Fisher's exact  
test  $P < 0.001$ ), IncP was also found in isolates that tended to lack IncF1A replicons (Fig

4B, Fisher's exact test  $P < 0.001$ ), while IncF1A was typically found in association with  
2 IncF1B (Fig 4C, Fisher's exact test  $P < 0.001$ ).

The number of replicons contained within each isolate was also affected by  
4 genetic and ecological factors. H serotype had a significant effect on the number of  
replicons found in each isolate (Figure 5A, *glm* with Poisson errors  $df = 28$ ,  $\chi^2 = 156$ ,  $P <$   
6  $0.0001$ ). However, this variation was largely due to the influence of three 'plasmid shy'  
serotypes (H4, H27, H28), these three serotypes accounted for more than half of the  
8 plasmid-free isolates (Figure 5). The variation in replicon number could be neatly  
described by dividing bacterial serotypes into two categories: 'plasmid shy' serotypes  
10 (H4, H27, H28) and 'plasmid normal' (all other genotypes) without any significant loss of  
explanatory power ( $df = 27$ ,  $\chi^2 = 37.5$ ,  $P = 0.086$ ); in other words there was no significant  
12 variation in replicon counts among the plasmid normal strains. The varying prevalence  
of these plasmid shy strains could explain some of the variation in replicon number per  
14 isolate that we saw between sites (*glm* with Poisson errors  $df = 8$ ,  $\chi^2 = 115$ ,  $P < 0.0001$ ),  
in particular note the high prevalence of serotypes H4 and H28 in sites OC and RH  
16 respectively (Figure 1). Nevertheless, site still had an effect on plasmid occupancy when  
we considered plasmid normal serotypes only (*glm* with Poisson errors  $df = 8$ ,  $\chi^2 = 58$ ,  $P$   
18  $< 0.0001$ ) especially in Site OC which had on average nearly one fewer replicon per  
isolate than most other sites (Figure 5B). The number of replicons per isolate fitted a  
20 Poisson distribution fairly well. However, we saw fewer isolates with zero replicons  
than we would expect by chance and an excess of isolates with two replicons in  
22 particular (Pearson's goodness of fit test,  $df = 3$ ,  $\chi^2 = 16.8$ ,  $P < 0.001$ , Figure 5C). This  
pattern was even more marked when we considered the plasmid normal strains only  
24 (Pearson's goodness of fit test,  $df = 3$ ,  $\chi^2 = 40.0$ ,  $P < 0.0001$ ).

## Discussion

2 Host population structure is recognized as a key determinant of genetic diversity in  
parasites (Van den Broeck *et al.*, 2014). As plasmids can be considered to act as  
4 parasites in some contexts, it follows that the population structure of host cells may also  
have a large influence of plasmid genetic diversity. Bacterial population structure also  
6 has many implications for the maintenance of investment in virulence (Buckling and  
Brockhurst, 2008; van Leeuwen *et al.*, 2015) and our understanding of how plasmids  
8 are maintained in bacterial communities. It is therefore vital to consider the host  
population structure when examining plasmid populations. In this study, many pats  
10 appear to contain *E. coli* populations with very limited diversity, commonly all isolates  
from a given pat had the same H-antigen type. Time series studies of cattle and human  
12 hosts have shown that hosts tend to be colonized by a single dominant *E. coli* subtype  
and a range of less frequent transient subtypes over time (Anderson *et al.*, 2006;  
14 Caugant *et al.*, 1981; Jenkins *et al.*, 2003). When high levels of antimicrobial resistance  
have been found in *E. coli* in cattle single genotypes can dominate entire farms (Sawant  
16 *et al.*, 2007), a pattern that we did not see, potentially because herds reared for  
conservation management are not subject to strong selective sweeps based on antibiotic  
18 exposure. The data in this study fit well with the rarity of mixed infections shown by  
many bacteria (Balmer and Tanner, 2011), a phenomenon that can be explained by  
20 bottlenecks in host colonization (van Leeuwen *et al.*, 2015).

22 The data on the structure of the host *E. coli* population allow us to make some  
inferences about how plasmids persist in this type of community. We found little within  
24 species bacterial diversity within hosts, but considerable variation in plasmid carriage  
within bacterial genotypes within faecal sample (Figure 1, Figure 5). Since *E. coli* within

one pat (animal) are mostly of one H-antigen type, plasmids must persist via vertical  
2 transmission, or if they are horizontally transferred, the majority of conspecific potential  
recipients are likely to be closely related plasmid-free bacteria of the same genotype. A  
4 low diversity *E. coli* population within an animal therefore has implications for the  
evolutionary ecology of plasmid persistence; repeated encounters of bacteria and  
6 plasmid genotypes should favour ongoing coevolution between the two. However,  
transient or rare *E. coli* subtypes, or other Gram negative species could introduce new  
8 plasmids to the system via conjugation (Caugant *et al.*, 1981).

10       One of the most interesting findings in this study is that bacterial host abundance  
has little effect on plasmid replicon diversity, whereas bacterial host diversity has a  
12 significant effect. This patterns fitted well with the observation that plasmid replicons  
were associated with specific host types (Figure 3), as has been seen in between species  
14 surveys of plasmid carriage (Carattoli, 2009). The strong associations of plasmid  
replicons with particular bacterial genotypes indicates that coevolution of plasmids and  
16 hosts through altered conjugation rate, loss of genes or changes in gene expression is  
ecologically realistic (Harrison and Brockhurst, 2012; Slater *et al.*, 2008). Experimental  
18 evolution has revealed that the fitness costs associated with the carriage of recently  
acquired plasmids can be ameliorated very quickly, either by mutations on plasmids or  
20 the host chromosome (Bouma and Lenski, 1988; Dahlberg and Chao, 2003; Dionisio *et al.*,  
*et al.*, 2005; Harrison *et al.*, 2015; San Millan *et al.*, 2014). However, while plasmids can  
22 have low fitness costs in many contexts (Vogwill and Maclean, 2015), these costs can  
vary considerably in different host bacteria species (Hall *et al.*, 2015; Riccobono *et al.*,  
24 2015). The variation in fitness cost may be due to different lengths of coevolutionary  
history between bacteria and plasmids (Vogwill and Maclean, 2015), but it is not clear

whether this can account for the host associations shown here, within *E. coli*, or across  
2 the Enterobacteriaceae (Carattoli, 2009).

Conceivably the opportunities for reducing the costs of carriage are not equal in  
4 all genetic contexts. Epistatic interactions between plasmids, that can have significant  
impacts on fitness costs, may play a role here. Previous work has shown that small  
6 plasmids can have substantial effects on the costs of carrying large plasmids, and meta-  
analysis of sequence data show that large and small plasmids are commonly found  
8 together (San Millan *et al.*, 2013). Here, we found evidence for a frequent occurrence of  
two replicons per bacterial host in ecological communities. While some of these  
10 associations may arise because large IncF -type plasmids will often carry both IncF1A  
and F1B replicons (Villa *et al.*, 2010), the association between IncF1B and IncP replicons  
12 was novel. Conversely, antagonistic interactions between IncF1A and IncP, which are  
known to interfere in other's conjugation potential (Miller *et al.*, 1985) may explain the  
14 rare co-occurrence of these two replicon types.

Potentially different bacterial lineages can follow different evolutionary  
16 trajectories shaped by initial plasmid complements. In this context, the occurrence of  
serotypes with consistently low rates of carriage, the plasmid shy strains, is particularly  
18 intriguing, especially given the high prevalence of two of them (H4, H28). Investigating  
the mechanism behind these low rates of carriage would be worthy of future study. The  
20 genetic background of *E. coli* plasmid recipients are known to have a considerable,  
unexplained, effect on conjugation rates (Dionisio *et al.*, 2002). Other possible  
22 mechanisms include genetic backgrounds that have higher segregation rates, or which  
harbour cryptic plasmids which confer partial incompatibility. Site also had a strong  
24 influence on rates of plasmid carriage, and this was true even when we excluded the  
effect of the prevalence of plasmid shy strains. After excluding genotype effects, the site

with the lowest rate of plasmid occupancy (site OC) also had the highest prevalence of  
2 plasmid shy bacteria. This is particularly intriguing. In the same way that highly  
efficient plasmid donors in diverse communities can accelerate plasmid spread (Dionisio  
4 *et al.*, 2002), we can speculate that the high prevalence of strains that are 'plasmid sinks'  
might also have an effect on plasmid carriage at the level of the population.

6  
It has been proposed that farm animals may be important carriers of, and reservoirs for,  
8 ESBL genes (Carattoli, 2008). However, data on the prevalence of ESBL producing  
bacteria in food or food-animals is highly variable, ranging from 0.2% to 40% (Liebana  
10 *et al.*, 2013). This difference is highlighted by the contrast between data in this study  
and a previous study of faecal samples from dairy cattle; < 3 % of *E. coli* were resistant  
12 to ampicillin, chloramphenicol or cefotaxime in this study, whereas 48 % and 20 % of *E.*  
*coli* were resistant to ampicillin and chloramphenicol respectively in dairy herds  
14 (Sawant *et al.*, 2007). It is of course possible that the *E. coli* sampled in this study carry  
resistance to antibiotics other than those tested. However, the lack of ESBL resistance is  
16 noteworthy as it may indicate that resistance in livestock is not as prevalent as is  
supposed. Five of the nine sites in the study were small grazing herds kept on nature  
18 reserves. It is highly likely that these cattle have had very low exposure to antibiotics,  
and the low levels of resistance found suggest that this means of cattle rearing may  
20 reduce the prevalence of resistance. The spread of antibiotic resistant organisms  
through the food chain is considered to be a significant risk to public health, but if  
22 resistance in livestock can be managed effectively, this risk may be minimised (Liebana  
*et al.*, 2013). Another basic issue relating to evolution of antibiotic resistance was the  
24 observation of the common variation in plasmid profiles within bacterial genotypes  
within mammalian hosts. The explanation for this variation and lack of fixation in

plasmid carriage is still not clear. Similarly, we understand little about the striking  
2 levels of polymorphism in antibiotic resistance seen in many bacteria (Colijn *et al.*,  
2010). Given the common association between plasmid carriage and antibiotic  
4 resistance understanding the basic polymorphism in plasmids might be a useful step.

6 This study presents some fundamental ecology of enteric *E. coli* and plasmids  
from cattle. It has found surprisingly low levels of resistance, giving hope that  
8 alternative means of cattle rearing may lead to lower prevalence of antibiotic resistance.  
It has also demonstrated the link between host *E. coli* diversity and plasmid diversity,  
10 and found that *E. coli* populations are highly structured. Further sequencing of plasmids  
would allow us to build a fuller picture of the plasmid population, sequencing of  
12 bacterial genomes may allow us to identify genetic determinants of plasmid carriage and  
deep sequencing may also uncover small plasmids that were missed by replicon typing,  
14 and enable more accurate analysis of plasmid mobility.

## 16 **Experimental procedures**

### Sample collection & isolation of *E. coli*

18 Bovine faecal samples (cowpats) were collected from 9 sites across Surrey and  
Berkshire: four farms or areas of common land (sites A, D, R & W) and five nature  
20 reserves with grazing cattle managed by Surrey Wildlife Trust (sites CC, OC, RH, TM and  
WC). These were chosen because the cattle are extensively grazed on semi-natural  
22 grassland and therefore would not have general exposure to antibiotics as part of  
routine diets. Sampling was conducted in a hierarchical manner: 10 cowpats were  
24 sampled from each site and 10 *E. coli* were screened from each pat (unless numbers  
were too low, in which case all available pats/ *E. coli* were screened). Cowpats were



sampled at regular intervals along walking transects in each field, with single pats being  
2 taken randomly from any clustered deposits. *E. coli* survival in cowpats in the field is  
robust for up to 30 days, but fresh pats were selected when possible (Van Kessel *et al.*,  
4 2007). Additional details of sites, site management and mean *E. coli* density is given in  
Supplementary table S1.

6 Approximately 2 g faecal matter was collected from each pat using sterile  
collecting tubes with spoon lids (Nunc™, Thermo Scientific, UK). Samples were returned  
8 to the laboratory where they were stored at 10°C until processing. All samples were  
processed with 24 hours of collection by homogenisation in 5 ml sterile 0.85% w/v NaCl  
10 followed by centrifugation at 500 g for 2 minutes to pellet the solid material. The  
supernatants were diluted 100-fold with sterile 0.85% w/v NaCl, and 100 µl was plated  
12 onto HiCrome™ Coliform Agar containing 5 mg l<sup>-1</sup> novobiocin (Sigma, UK). Total coliform  
counts were recorded after overnight incubation at 37°C. Ten colonies per cowpat  
14 sample were re-streaked onto HiCrome™ plates to obtain clonal colonies. Overnight  
cultures in LB broth were inoculated from these clonal colonies for testing with Kovac's  
16 Reagent (Sigma). 10 µl Kovac's Reagent was added to 150 µl overnight culture; a cherry  
red colour confirms presence of *E. coli*. Isolated colonies were stored long term in 80%  
18 glycerol at -80°C.

#### 20 DNA extraction & plasmid replicon typing

Overnight cultures were set up in 2 ml LB broth in 24 well plates. These were  
22 centrifuged at 5500 g for 5 minutes and the pellet resuspended in 0.5 ml molecular  
grade H<sub>2</sub>O. This suspension was boiled for 10 minutes at 100°C and then centrifuged at  
24 4500 g for 5 minutes. Supernatants were used as templates for replicon typing PCR and  
flagellin sequencing; best results were obtained when supernatants were used

immediately. Replicon typing PCR was used to identify large plasmids in isolated *E. coli*.

2 This method comprises 3 multiplex-PCR panels with 18 primer pairs recognising the  
major plasmid incompatibility groups of the *Enterobacteriaceae* (Carattoli *et al.*, 2005;  
4 Johnson *et al.*, 2007). Note that this panel is targeted at large conjugative plasmids and  
is not therefore expected to cover smaller less well-defined plasmids. PCR reagent  
6 concentrations were as follows: 0.4mM dNTPs, 0.625 units Taq (Qiagen) and 2.4  $\mu$ M of  
each primer in the multiplex panel. 0.5  $\mu$ l boiled lysate template was added to each 24.5  
8  $\mu$ l reaction mix. Cycling conditions were: initial denaturation, 95°C, 5 minutes, followed  
by 30 cycles of denaturation (95°C, 30 seconds), annealing (60°C, 30 seconds) and  
10 extension (72°C, 90 seconds) and a final extension, step of 72°C for 5 minutes.

## 12 *E. coli* genotyping

*E. coli* isolates were genotyped by sequencing the variable region of the flagellin  
14 gene, *fliC*. This method is an adaptation of previous genotyping methods using whole  
gene sequencing (Wang *et al.*, 2006) and PCR-RFLP (restriction fragment length  
16 polymorphism) of flagellin genes (Machado *et al.*, 2000). The *fliC* gene was amplified  
using the PCR-RFLP primers *fliC* 1 (5-3' CAAGTCATTAATAACAGCC) and *fliC* 2 (5-3'  
18 GACATRTTRGAVACTTCSGT) (Machado *et al.*, 2000). Cycling conditions were as follows:  
initial denaturation, 95°C, 2 minutes, followed by 30 cycles of denaturation (95°C, 40  
20 seconds), annealing (50°C, 30 seconds) and extension (72°C, 120 seconds) and a final  
extension step of 72°C for 3 minutes. PCR products were cleaned up using Exo1 (Fisher  
22 Scientific) and TSAP (Promega UK, Southampton, UK) both at final concentration 0.045  
units  $\mu$ l<sup>-1</sup>. Clean up mixes were incubated at 37°C for 30 minutes, followed by 80°C for  
24 15 minutes. Sequencing reactions were conducted using BigDye® terminator v3.1 cycle  
sequencing kit (Life Technologies Ltd. Paisley, UK) using the *fliC* 2 primer only.

Sequencing reactions and clean-ups followed the BigDye recommendations and  
2 sequencing runs were conducted by the Department of Zoology, University of Oxford,  
UK. Trimmed sequences were searched against the whole flagellin gene sequences of  
4 Wang *et al.*, (2003) using Geneious (Biomatters 2012). Sequences were assigned a  
flagellin type (H-antigen type) when they had at least 200 bp of high identity (>95%)  
6 with one *fliC* sequence. Where several matches were made the match with highest  
identity was selected. This method allowed 97% sequences to be assigned a flagellin  
8 type. *E. coli* H-antigen types were placed into one of four subgroups defined by Wang *et al.*  
(2003)(Supplementary table 2). These are based on H-antigen groupings of variable  
10 region of the flagellin gene (V). Wang *et al.* (2003) failed to categorise the H16 group  
based on its V region sequence, but placed it in the Ec1b group based on the conserved  
12 C1 and C2 regions.

#### 14 Antibiotic susceptibility screening

The susceptibility of *E. coli* isolates to ampicillin, cefotaxime and chloramphenicol was  
16 assessed by disc diffusion assays. MIC breakpoints and interpretation of zone diameters  
were taken from the British Society for Antimicrobial Chemotherapy Guidelines version  
18 11.1 (May 2012). Susceptibility testing methods followed Andrews (2005), using six  
antibiotic discs per plate placed at uniform intervals using sterile forceps. Zones of  
20 inhibition were measured after overnight incubation at 37°C, using a disc diameter  
template. Two doses were used for each antibiotic (2 and 10 µg ml<sup>-1</sup> ampicillin; 5 and 30  
22 µg ml<sup>-1</sup> cefotaxime; 10 and 30 µg ml<sup>-1</sup> chloramphenicol) and strains were determined to  
be resistant if they scored positively with both discs.

24

Data analysis

All statistical analyses were conducted in *R* v3.1.2 (R Core Team 2013). Plasmid and H-antigen diversity were calculated using Simpson's and the Shannon diversity index using the *Vegan* package in *R* (Oksanen *et al.*, 2013). Linear regressions were carried out to assess the affect of *E. coli* abundance on both plasmid and H-antigen diversity, and generalized linear models with Poisson errors used to investigate the number of replicons per bacterial host. Plasmid replicon distribution across subgroups and with respect to each was assessed Pearson's Chi squared goodness of fit tests and with Fisher's exact tests. Hierarchical *F*-statistics were calculated using *HierFstat* (de Meeus and Goudet, 2007).

10

## 12 **Acknowledgments**

FM was supported by a joint Royal Holloway and Animal Health and Veterinary Laboratories Agency studentship, BR by a NERC Advanced Research fellowship. Thanks to Tatiana Dimitriu for comments and to Surrey Wildlife Trust for permitting access to several sites.

## 18 **ReferencesUncategorized References**

- Anderson, M.A., Whitlock, J.E., and Harwood, V.J. (2006). Diversity and distribution of *Escherichia coli* genotypes and antibiotic resistance phenotypes in feces of humans, cattle, and horses. *Appl Environ Microbiol* **72**: 6914-6922.
- Andrews, J.M. (2005). BSAC standardized disc susceptibility testing method (version 4). *J Antimicrob Chemother* **56**: 60-76.
- Balmer, O., and Tanner, M. (2011). Prevalence and implications of multiple-strain infections. *Lancet Infect Dis* **11**: 868-878.

- 1 Bergstrom, C.T., Lipsitch, M., and Levin, B.R. (2000). Natural selection, infectious transfer  
2 and the existence conditions for bacterial plasmids. *Genetics* **155**: 1505-1519.
- 3 Bouma, J.E., and Lenski, R.E. (1988). Evolution of a bacteria plasmid association. *Nature*  
4 **335**: 351-352.
- 5 Buckling, A., and Brockhurst, M.A. (2008). Kin selection and the evolution of virulence.  
6 *Heredity* **100**: 484-488.
- 7 Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K.L., and Threlfall, E.J. (2005).  
8 Identification of plasmids by PCR-based replicon typing. *J Microbiol Methods* **63**:  
219-228.
- 9 Carattoli, A. (2008). Animal reservoirs for extended spectrum  $\beta$ -lactamase producers.  
10 *Clin Microbiol Infect* **14**: 117-123.
- 11 Carattoli, A. (2009). Resistance Plasmid Families in Enterobacteriaceae. *Antimicrob*  
12 *Agents Chemother* **53**: 2227-2238.
- 13 Carattoli, A. (2011). Plasmids in Gram negatives: molecular typing of resistance  
14 plasmids. *Int J Med Microbiol* **301**: 654-658.
- 15 Caugant, D.A., Levin, B.R., and Selander, R.K. (1981). Genetic diversity and temporal  
16 variation in the *E. coli* population of a human host. *Genetics* **98**: 467-490.
- 17 Colijn, C., Cohen, T., Fraser, C., Hanage, W., Goldstein, E., Givon-Lavi, N. *et al.*, (2010).  
18 What is the mechanism for persistent coexistence of drug-susceptible and drug-  
19 resistant strains of *Streptococcus pneumoniae*? *J R Soc Interface* **7**: 905-919.
- 20 Dahlberg, C., and Chao, L. (2003). Amelioration of the cost of conjugative plasmid  
21 carriage in *Escherichia coli* K12. *Genetics* **165**: 1641-1649.
- 22 Datta, N., and Hughes, V.M. (1983). Plasmids of the same Inc groups in Enterobacteria  
23 before and after the medical use of antibiotics. *Nature* **306**: 616-617.
- 24

- de Meeus, T., and Goudet, J. (2007). A step-by-step tutorial to use HierFstat to analyse  
2 populations hierchically structured at multiple levels. *Inf Genet Evol* **7**: 731-735.
- Dierikx, C., van Essen-Zandbergen, A., Veldman, K., Smith, H., and Mevius, D. (2010).  
4 Increased detection of extended spectrum beta-lactamase producing *Salmonella*  
*enterica* and *Escherichia coli* isolates from poultry. *Vet Microbiol* **145**: 273-278.
- Dierikx, C.M., van Duijkeren, E., Schoormans, A.H.W., van Essen-Zandbergen, A., Veldman,  
6 K., Kant, A. *et al.*, (2012). Occurrence and characteristics of extended-spectrum- $\beta$ -  
8 lactamase- and AmpC-producing clinical isolates derived from companion animals  
and horses. *J Antimicrob Chemother* **67**: 1368-1374.
- 10 Dionisio, F., Matic, I., Radman, M., Rodrigues, O.R., and Taddei, F. (2002). Plasmids spread  
very fast in heterogeneous bacterial communities. *Genetics* **162**: 1525-1532.
- 12 Dionisio, F., Conceicao, I.C., Marques, A.C.R., Fernandes, L., and Gordo, I. (2005). The  
evolution of a conjugative plasmid and its ability to increase bacterial fitness **1**: 250-  
14 252.
- Dröge, M., Puhler, A., and Selbitschka, W. (2000). Phenotypic and molecular  
16 characterization of conjugative antibiotic resistance plasmids isolated from bacterial  
communities of activated sludge. *Mol Gen Genet* **263**: 471-482.
- 18 Eberhard, W.G. (1990). Evolution in bacterial plasmids and levels of selection. *Q Rev Biol*  
**65**: 3-22.
- 20 Hall, J.P.J., Harrison, E., Lilley, A.K., Paterson, S., Spiers, A.J., and Brockhurst, M.A. (2015).  
Environmentally co-occurring mercury resistance plasmids are genetically and  
22 phenotypically diverse and confer variable context-dependent fitness effects.  
*Environ Microbiol* **adv. online pub.**
- 24 Harrison, E., and Brockhurst, M.A. (2012). Plasmid-mediated horizontal gene transfer is  
a coevolutionary process **20**: 262-267.

- Harrison, E., Guymmer, D., Spiers, A.J., Paterson, S., and Brockhurst, M.A. (2015). Parallel  
2 compensatory evolution stabilizes plasmids across the parasitism-mutualism  
continuum. *Curr Biol* **25**: 2034-2039.
- 4 Holsinger, K., and Weir, B.S. (2009). Genetics in geographically structured populations:  
defining, estimating and interpreting  $F_{ST}$ . *Nat Rev Genet* **10**: 639-650.
- 6 Hopkins, K.L., Liebana, E., Villa, L., Batchelor, M., Threlfall, E.J., and Carattoli, A. (2006).  
Replicon typing of plasmids carrying CTX-M or CMY beta-lactamases circulating  
8 among *Salmonella* and *Escherichia coli* isolates. *Antimicrob Agents Chemother* **50**:  
3203-3206.
- 10 Hordijk, J., Wagenaar, J.A., Kant, A., van Essen-Zandbergen, A., Dierikx, C., Veldman, K. *et*  
*al.*, (2013). Cross-sectional study on prevalence and molecular characteristics of  
12 plasmid mediated ESBL/AmpC-producing *Escherichia coli* isolated from veal calves  
at slaughter. *PLoS ONE* **8**: e65681.
- 14 Hu, X., Van der Auwera, G., Timmerly, S., Zhu, L., and Mahillon, J. (2009). Distribution,  
diversity, and potential mobility of extrachromosomal elements related to the  
16 *Bacillus anthracis* pXO1 and pXO2 virulence plasmids. *Appl Env Microbiol* **75**: 3016-  
3028.
- 18 Jenkins, M.B., Hartel, P.G., Olexa, T.J., and Stuedemann, J.A. (2003). Putative temporal  
variability of ribotypes from yearling steers. *J Environ Qual* **32**: 305-309.
- 20 Johnson, T.J., Wannemuehler, Y.M., Johnson, S.J., Logue, C.M., White, D.G., Doetkott, C., and  
Nolan, L.K. (2007). Plasmid replicon typing of commensal and pathogenic  
22 *Escherichia coli* isolates. *Appl Environ Microbiol* **73**: 1976-1983.
- Kav, A.B., Sasson, G., Jami, E., Doron-Faigenboim, A., Benhar, I., and Mizrahi, I. (2012).  
24 Insights into the bovine rumen plasmidome. *Proc Natl Acad Sci USA* **109**: 5452-  
5457.

- Leverstein-van Hall, M.A., Dierikx, C.M., Cohen Stuart, J., Voets, G.M., van den Munckhof, M.P., van Essen-Zandbergen, A. *et al.*, (2011). Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. *Clin Microbiol Infect* **17**: 873-880.
- Liebana, E., Carattoli, A., Coque, T.M., and Threlfall, J. (2013). Public health risks of enterobacterial isolates producing extended-spectrum -lactamases or AmpC -lactamases in food and food-producing animals. *Clin Inf Dis* **56**: 1030-1037.
- Lili, L.N., Britton, N.F., and Feil, E.J. (2007). The persistence of parasitic plasmids. *Genetics* **177**: 399-405.
- Lilley, A.K., and Bailey, M.J. (1997a). The acquisition of indigenous plasmids by a genetically marked pseudomonad population colonizing the sugar beet phytosphere is related to local environmental conditions. *Appl Environ Microbiol* **63**: 1577-1583.
- Lilley, A.K., and Bailey, M.J. (1997b). Impact of plasmid pQBR103 acquisition and carriage on the phytosphere fitness of *Pseudomonas fluorescens* SBW25: burden and benefit. *Appl Environ Microbiol* **63**: 1584-1587.
- Luo, H.L., Wan, K., and Wang, H.H. (2005). High-frequency conjugation system facilitates biofilm formation and pAM beta 1 transmission by *Lactococcus lactis*. *Appl Environ Microbiol* **71**: 2970-2978.
- Machado, J., Grimont, F., and Grimont, P.A.D. (2000). Identification of *Escherichia coli* flagellar types by restriction of the amplified *fliC* gene. *Res Microbiol* **151**: 535-546.
- Marcadé, G., Deschamps, C., Boyd, A., Gautier, V., Picard, B., Branger, C. *et al.*, (2009). Replicon typing of plasmids in *Escherichia coli* producing extended-spectrum beta-lactamases. *J Antimicrob Chemother* **63**: 67-71.



- Miller, J.F., Lanka, E., and Malamy, M.H. (1985). F-Factor Inhibition of Conjugal Transfer  
2 of Broad-Host-Range Plasmid-Rp4 - Requirement for the Protein Product of Pif  
Operon Regulatory Gene Pifc. *J Bacteriol* **163**: 1067-1073.
- 4 Modi, R.I., and Adams, J. (1991). Coevolution in bacterial-plasmid populations. *Evolution*  
**45**: 656-667.
- 6 Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R. *et al.*, (2013).  
Vegan: community ecology package. Available at [http://CRAN.R-](http://CRAN.R-project.org/package=vegan)  
8 [project.org/package=vegan](http://CRAN.R-project.org/package=vegan).
- Pallecchi, L., Lucchetti, C., Bartoloni, A., Bartalesi, F., Mantella, A., Gamboa, H. *et al.*,  
10 (2007). Population structure and resistance genes in antibiotic-resistant bacteria  
from a remote community with minimal antibiotic exposure. *Antimicrob Agents*  
12 *Chemother* **51**: 1179-1184.
- Riccobono, E., Di Pilato, V., Di Maggio, T., Revollo, C., Bartoloni, A., Pallecchi, L., and  
14 Rossolini, G.M. (2015). Characterization of IncI1 sequence type 71 epidemic plasmid  
lineage responsible for the recent dissemination of CTX-M-65 extended-spectrum  $\beta$ -  
16 lactamase in the Bolivian Chaco region. *Antimicrob Agents Chemother* **59**: 5340-  
5347.
- 18 San Millan, A., Heilbron, K., and Maclean, R.C. (2013). Positive epistasis between co-  
infecting plasmids promotes plasmid survival in bacterial populations. *ISME J* **8**:  
20 601-612.
- San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z.V., McLean, A.R., Cooper, B.S., and  
22 MacLean, R.C. (2014). Positive selection and compensatory adaptation interact to  
stabilize non-transmissible plasmids. *Nat Commun* **5**: 5208.

- Sawant, A.A., Hegde, N.V., Straley, B.A., Donaldson, S.C., Love, B.C., Knabel, S.J., and Jayarao, B.M. (2007). Antimicrobial-resistant enteric bacteria from dairy cattle. *Appl Environ Microbiol* **73**: 156-163.
- Slater, F.R., Bailey, M.J., Tett, A.J., and Turner, S.L. (2008). Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Micro Ecol* **66**: 3-13.
- Smith, J. (2010). Superinfection drives virulence evolution in experimental populations of bacteria and plasmids. *Evolution* **65**: 831-841.
- Sorensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., and Wuertz, S. (2005). Studying plasmid horizontal transfer in situ: A critical review. *Nat Rev Microbiol* **3**: 700-710.
- Van den Broeck, F., Meurs, L., Raeymaekers, J.A.M., Boon, N., Dieye, T.N., Volckaert, F.A.M. *et al.*, (2014). Inbreeding within human *Schistosoma mansoni*: do host-specific factors shape the genetic composition of parasite populations? *Heredity* **113**: 32-41.
- Van Kessel, J.S., Pachepsky, Y.A., Shelton, D.R., and Karns, J.S. (2007). Survival of *Escherichia coli* in cowpats in pasture and in laboratory conditions. *J Appl Microbiol* **103**: 1122-1127.
- van Leeuwen, E., Neill, S.O.A., Matthews, A., and Raymond, B. (2015). Making pathogens sociable: The emergence of high relatedness through limited host invasibility. *ISME J* **9**: 2328.
- Vilas-Bôas, G., Vilas-Boas, L.A., Lereclus, D., and Arantes, O.M.N. (2008). *Bacillus thuringiensis* conjugation under environmental conditions. *FEMS Microbiol Ecol* **24**: 369-374.
- Villa, L., García-Fernández, A., Fortini, D., and Carattoli, A. (2010). Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *J Antimicrob Chemother* **65**: 2518-2529.

Vogwill, T., and Maclean, R.C. (2015). The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol Appl* **8**: 284-295.

Wang, L., Rothemund, D., Curd, H., and Reeves, P.R. (2003). Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. *J Bacteriol* **185**: 2936-2943.

Wang, Y., Leung, P.C., Qian, P.-Y., and Gu, J.-D. (2006). Antibiotic resistance and plasmid profile of environmental isolates of *Vibrio* species from Mai Po Nature Reserve, Hong Kong. *Ecotoxicology* **15**: 371-378.

Wright, S. (1978). Variability within and among natural populations. *Evolution and the genetics of populations: a treatise in four volumes*. Chicago: Chicago University Press pp 246-326.

**Supporting Information.** Additional Supporting Information may be found in the online version of the article at the publisher's website. **Table S1 ....**

**Figure 1.** The structure of the *E. coli* population in faecal samples from 9 sites of extensively grazing cattle. A. The frequency distribution of bacterial serotype diversity within samples (= pat). The histogram shows sample counts on the y axis and bins of Simpson's diversity on the x-axis. The inverse Simpson's diversity index was used to illustrate this distribution as this takes a value of 1 when samples are clonal; most samples were dominated by one serotype. B. A histogram of the variation in number of distinct plasmid profiles within bacterial serotype within each faecal sample, with isolate counts on the y axis and bins of profile richness on the x-axis. In only a minority of samples did isolates of the same serotype have identical plasmid profiles. These data used  $N = 57$  pats for which we had at least four isolates of the same serotype. C. The variation in genetic diversity of *E. coli* and distribution of serotype across sites, each serotype has a unique colour, frequency refers to the counts of genotype bacteria at each site. Note that some sites are under-sampled due to low *E. coli* abundance.

14

**Figure 2.** Plasmid replicon diversity does not correlate with the abundance of *E. coli* at A. the level of each site study site or B. at the level of the cowpat across all sites. Data in B are  $\log_{10}$  *E. coli* per pat / sample.

18

**Figure 3.** Plasmid replicons are not randomly distributed across *E. coli* serotype groups. A. Mosaic plot of replicon / serotype group association; the sizes of the rectangles correspond to total isolate counts each class; classes with significantly higher than expected counts are coloured in blue, lower than expected counts coloured in red, with the residuals of Pearson's goodness of fit tests in the legend,  $N = 527$  isolates. Note that categorical labels use the abbreviations FA, FB, FC, P etc to indicate IncF1A, IncF1B, IncF1C and Inc P etc replicons. B. Sites with high *E. coli* serotype diversity tend to have

high plasmid replicon diversity. C. *E. coli* serotype diversity is not significantly  
 2 correlated with *E. coli* abundance. D. Replicon diversity is positively correlated with *E.*  
*coli* serotype diversity within sites. Serotype and replicon diversity data are Shannon  
 4 diversity indices for each pat with good sampling, while lines represent the fitted model  
 with a common slope and different intercepts for each site.

6

**Figure 4.** Plasmid replicons are not randomly distributed with respect to each other in  
 8 *E. coli* isolates. Mosaic plots (see **Fig. 3** legend for explanation) show the association  
 between the three most common replicon incompatibility groups in the 527 isolates. A.  
 10 IncP replicons are strongly associated with IncF1B replicons across isolates. B. Inc P  
 replicons are found less often in isolates carrying IncF1A replicons than expected by  
 12 chance. C. IncF1A and IncF1B replicons are strongly associated across bacterial  
 isolates.

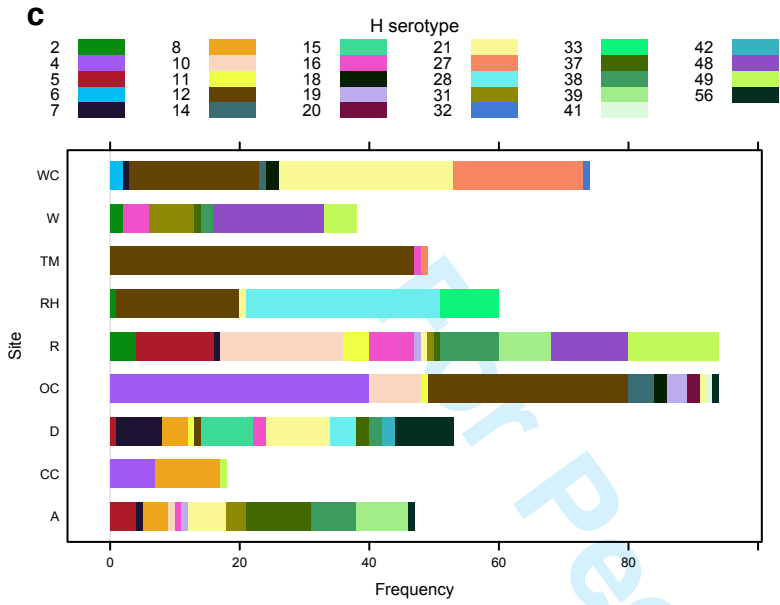
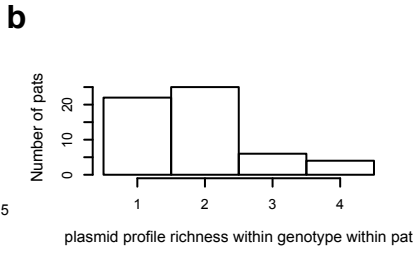
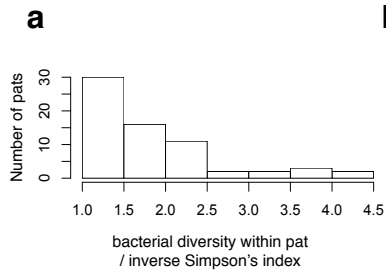
14

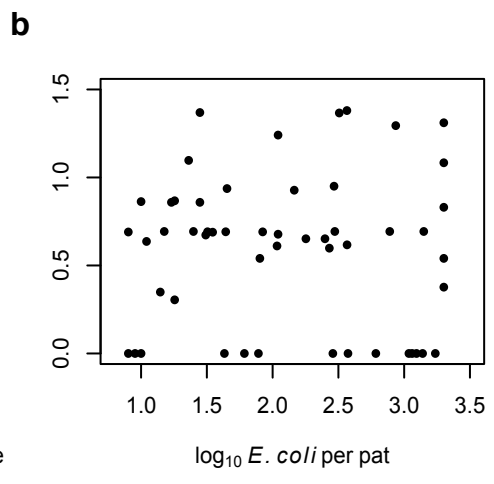
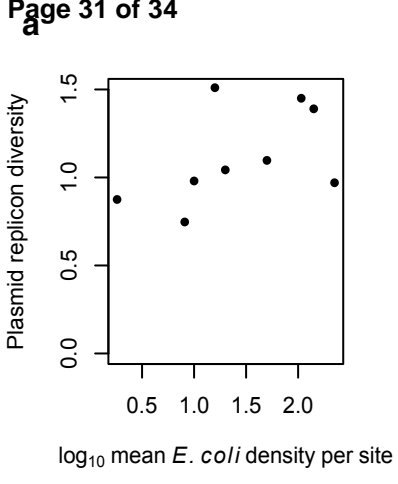
**Figure 5.** A. The variation in numbers of plasmid replicons per bacterial isolate  
 16 according to *E. coli* H serotype. Frequency refers to counts of isolates with particular  
 replicon numbers. Three serotypes (indicated by asterisks) have significantly fewer  
 18 replicons than the bulk of the genotypes, these are the 'plasmid shy' strains; \*\* and \*\*\*  
 indicate significantly fewer replicons per isolate in *post hoc* comparison at the  $P < 0.01$   
 20 and  $P < 0.001$  respectively. B. Site level variation in numbers of replicons per isolate in  
 plasmid normal serotypes, *i.e.* excluding plasmid shy strains. Data are means  $\pm$  1SE. C.  
 22 A histogram of the distribution of the number of replicons per isolate; expected  
 frequencies were generated from a Poisson distribution with a mean of 1.25 replicons  
 24 per host, frequencies are isolate counts (on the y axis), while the x axis represents bins

of number of distinct replicons. We found a significant mismatch between expected and  
2 observed frequencies (Pearson's goodness of fit test:  $\chi^2 = 16.8$ ,  $df = 3$ ,  $P < 0.001$ ).

4

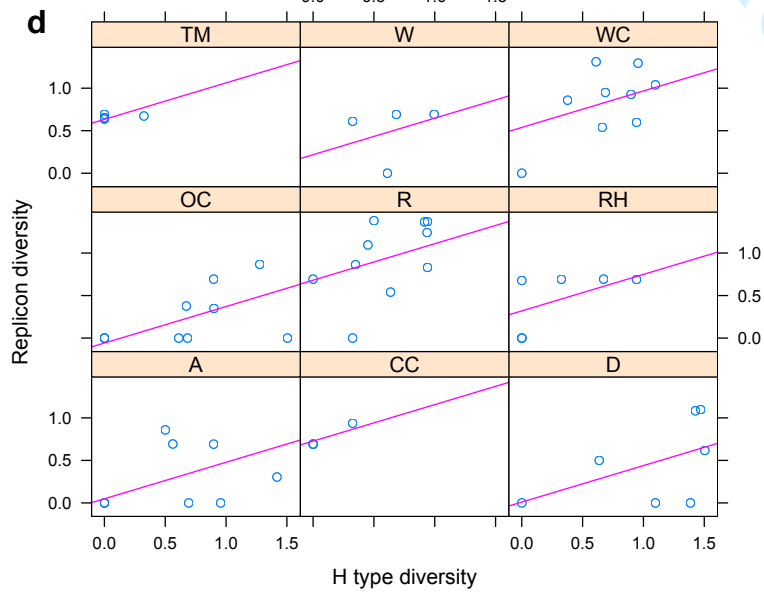
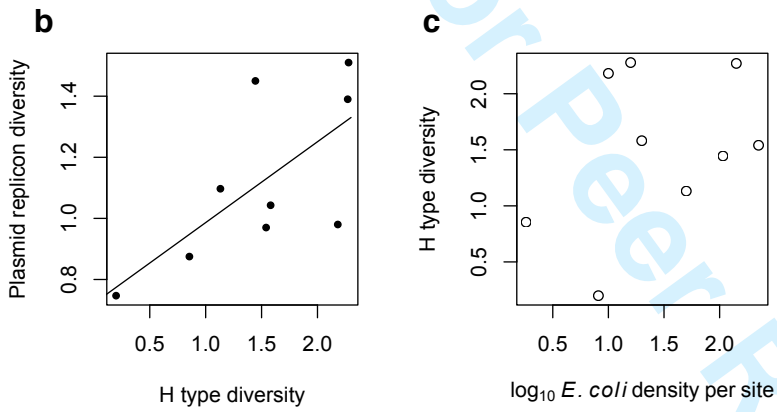
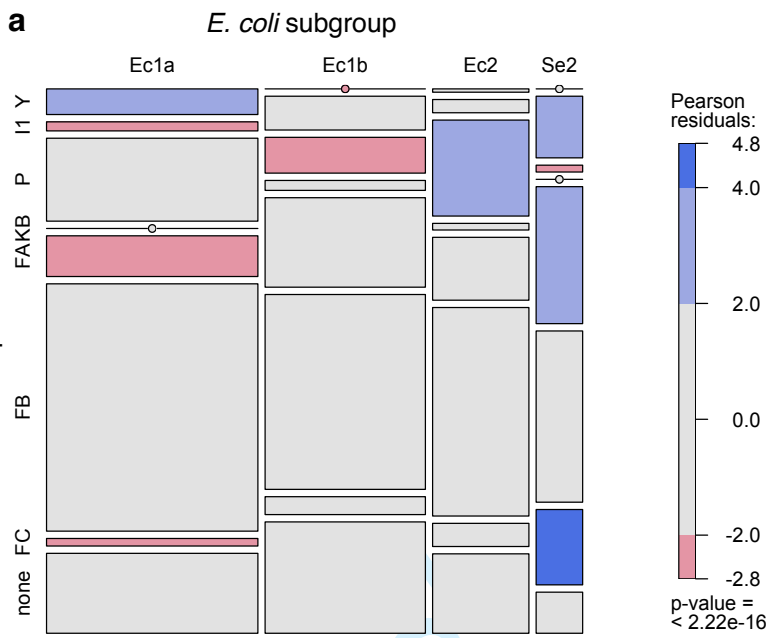
For Peer Review Only

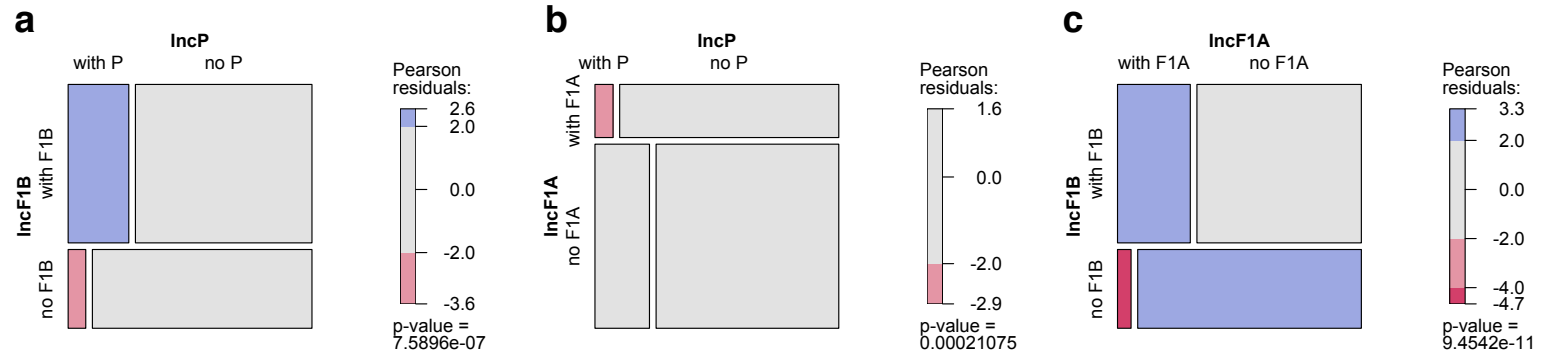




For Peer Review Only







For Peer Review Only

