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

Frances Medaney, Richard J. Ellis and Ben Raymond

1 School of Biological Science, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK.
2 Specialist Scientific Support Department, Animal and Plant Health Agency, APHA Weybridge, Addlestone, Surrey, KT15 3NB, UK.
3 University of Exeter, Penryn Campus, Penryn, Cornwall, TR10 9FE, UK.

Introduction

The persistence, replication and spread of bacterial plasmids is a major focus of research because so many plasmids carry virulence and antibiotic resistance genes (Dröge *et al.*, 2000; Hopkins *et al.*, 2006; Johnson *et al.*, 2007; Pallecchi *et al.*, 2007; Marcadé *et al.*, 2009; Carattoli, 2011). Despite their clinical importance, the ecology and evolutionary biology of plasmids has received relatively little attention. There are a range of contrasting views of the relationship between plasmids and their bacterial 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 (Modi and Adams, 1991; Lili *et al.*, 2007; Smith, 2010). Plasmids can also be seen as carriers of intermittently selected traits (Eberhard, 1990; Lilley and Bailey, 1997b), or indeed as ‘domesticated’ accessory genomes that may end up loosing conjugational ability entirely, such as the plasmid pXO2 that encodes encapsulation genes for *Bacillus anthracis* (Hu *et al.*, 2009).

Understanding how plasmids can persist in bacterial populations over the long 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 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 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 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 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., 2011; 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; Leverstein-van Hall et al., 2011; Dierikx et al., 2012; Hordijk et al., 2013). Another common approach is to recover plasmids from environment that may recently have been subject to strong selection for antibiotic resistance (Dröge et al., 2000; Dierikx et al., 2010, 2012). However, these studies are not usually designed 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, 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 plasmids that do not carry these intermittently selected traits. Exceptions, that is, plasmid studies not focussed on resistance genes, tend to be from the older literature, when studies were limited by low replication and examined very broad species/plasmid associations (Datta and Hughes, 1983) or are modern deep-sequencing approaches targeting small circular molecules, in which bacterial host-association information is typically lost (Kav et al., 2012).

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 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 Escherichia coli plasmids in a natural context, without strong selection, and to explore how variation in bacterial host populations, particularly genetic diversity and abundance, could affect the diversity and distribution of plasmid carriage. We used a study system based on isolating culturable strains from fresh bovine faecal matter. We sampled cattle grazing on unimproved semi-natural grassland, on either common land or land managed for conservation. We anticipated that these outdoor reared animals would maintain diverse enteric populations that had not been subject to strong selection for antibiotic resistance. This design permitted a hierarchical sampling strategy (plasmids within bacteria within pats within fields) that could be replicated across multiple sites with independent cattle populations.

Results

Diversity and antibiotic resistance in the E. coli populations

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 assess the distribution of bacterial diversity between sites and vertebrate hosts. In total 24 483 colonies were isolated from 84 cowpats across 9 sites. Of these, 527 isolates were genotyped and screened for plasmid replicons and antibiotic susceptibility. Populations were diverse, strongly structured and characterized by a low frequency of resistance isolates. Of the 527 isolates just 14 (< 3%) were resistant to any of the antibiotics tested: 9 isolates were ampicillin resistant, 4 were chloramphenicol resistant, and just 1 was resistant to cefotaxime. No multiresistant isolates were found. Resistant isolates were found in 5 of the 9 sites, with 4 sites resistance-free.

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 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, 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 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 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 (Wright, 1978; Holsinger and Weir, 2009).

**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 (\(F_{1,7} = 2.81, p = 0.14; \text{Fig. 2A}\)). 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, Fig. 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 data sets are available on Dryad (https://datadryad.org/resource/doi:10.5061/dryad.5f1f8).

Figure 3A highlights the non-random distribution of replicons across *E. coli* 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 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 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 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.

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 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, Fig. 3B). This relationship appeared to be independent of the effects of *E. coli* abundance since *E. coli* abundance and *H* serotype diversity were uncorrelated (Fig. 3C, *F*₁,₇ = 1.09, *p* = 0.33). Variation in *H* serotype diversity, and overall site effects both significantly affected plasmid diversity at the pat level (Fig. 3D, *F*₁,₅₁ = 13.8, *p* = 0.0005, and *F*₈,₅₂ = 4.02, *p* = 0.0009) and together explained 54% of the variation in plasmid diversity.

Plasmid replicons were also not randomly distributed with respect to each other. We examined whether the three most common replicons, IncP, IncFIA and IncFIB associated independently. We found that IncP was strongly associated with IncF1B 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
The number of replicons contained within each isolate was also affected by genetic and ecological factors. H serotype had a significant effect on the number of replicons found in each isolate (Fig. 5A, glm with Poisson errors $df = 28$, $\chi^2 = 156$, $p < 0.0001$). However, this variation was largely due to the influence of three ‘plasmid shy’ serotypes (H4, H27, H28), three of which are accounted for more than half of the plasmid-free isolates (Fig. 5). The variation in replicon number could be neatly described by dividing bacterial serotypes into two categories: ‘plasmid shy’ serotypes (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 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 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 respectively (Fig. 1). Nevertheless, site still had an effect on plasmid occupancy when we considered plasmid normal serotypes only (glm with Poisson errors $df = 8$, $\chi^2 = 58$, $p < 0.0001$) especially in Site OC which had on average nearly one fewer replicon per isolate than most other sites (Fig. 5B). The number of replicons per isolate fitted a 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 particular (Pearson’s goodness of fit test, $df = 3$, $\chi^2 = 16.8$, $p < 0.001$, Fig. 5C). This pattern was even more marked when we considered the plasmid normal strains only (Pearson’s goodness of fit test, $df = 3$, $\chi^2 = 40.0$, $p < 0.0001$).

**Discussion**

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 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 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 are maintained in bacterial communities. It is therefore vital to consider the host population structure when examining plasmid populations. In this study, many paths appear to contain E. coli populations with very limited diversity, commonly all isolates from a given path had the same H-antigen type. Time series studies of cattle and human 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 (Caugant et al., 1981; Jenkins et al., 2003; Anderson et al., 2006). When high levels of antimicrobial resistance have been found in E. coli in cattle, single genotypes can dominate entire farms (Sawant et al., 2007). We did not see this pattern potentially because herds reared for conservation management are not subject to strong selective sweeps based on antibiotic 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 bottlenecks in host colonization (Van Leeuwen et al., 2015).

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 species bacterial diversity within hosts, but considerable variation in plasmid carriage within bacterial genotypes within faecal sample (Figs. 1 and 5). Since E. coli within one pat (animal) are mostly of one H-antigen type, plasmids must persist via vertical 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 low diversity E. coli...
population within an animal therefore has implications for
the evolutionary ecology of plasmid persistence; repeated
encounters of bacteria and plasmid genotypes should
favour ongoing coevolution between the two. However,
transient or rare \textit{E. coli} subtypes, or other Gram negative
species could introduce new plasmids to the system via
conjugation (Caugant \textit{et al.}, 1981).

One of the most interesting findings in this study is that
bacterial host abundance has little effect on plasmid repli-
con diversity, whereas bacterial host diversity has a
significant effect. This patterns fitted well with the observa-
tion that plasmid replicons were associated with specific
host types (Fig. 3), as has been seen in between species
surveys of plasmid carriage (Carattoli, 2009). The strong
associations of plasmid replicons with particular bacterial
genotypes indicates that coevolution of plasmids and hosts
through altered conjugation rate, loss of genes or changes
in gene expression is ecologically realistic (Slater \textit{et al.},
2008; Harrison and Brockhurst, 2012). Experimental evolu-
tion 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 the host chromosome (Bouma and Lenski, 1988; Dahlberg and Chao, 2003; Dionisio et al., 2005; San Millan et al., 2014; Harrison et al., 2015). However, while plasmids can 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., 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 the Enterobacteriaceae (Carattoli, 2009).

Conceivably the opportunities for reducing the costs of carriage are not equal in 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 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 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 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 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 rare co-occurrence of these two replicon types.

Potentially different bacterial lineages can follow different evolutionary 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 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 genetic background of E. coli plasmid recipients are known to have a considerable, unexplained, effect on conjugation rates (Dionisio et al., 2002). Other possible mechanisms include genetic backgrounds that have higher segregation rates, or which harbour cryptic plasmids which confer partial incompatibility. Site also had a strong 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 plasmid shy bacteria. This is particularly intriguing. In the same way that highly efficient plasmid donors in diverse communities can accelerate plasmid spread (Dionisio 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.

It has been proposed that farm animals may be important carriers of, and reservoirs for, 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 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 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 (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 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 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 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 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 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 levels of polymorphism in antibiotic resistance seen in many bacteria (Colijn et al., 2010). Given the common association between plasmid carriage and antibiotic resistance understanding the basic polymorphism in plasmids might be a useful step.

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 alternative means of cattle rearing may lead to lower prevalence of antibiotic resistance. It has also demonstrated a link between host E. coli diversity and plasmid diversity, 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 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, and enable more accurate analysis of plasmid mobility.

Experimental procedures

Sample collection and isolation of E. coli

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 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 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 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 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., 2007). Additional details of sites, site management and mean E. coli density is given in Supporting Information Table S1.

Approximately 2 g faecal matter was collected from each pat using sterile collecting tubes with spoon lids (Nunc™, Thermo Scientific, UK). Samples were returned 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 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 onto HiCrome™ Coliform Agar containing 5 mg l⁻¹ novobiocin (Sigma, UK). Total coliform counts were recorded after overnight incubation at 37°C. Ten colonies per cowpat 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 Reagent (Sigma). About 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% glycerol at −80°C.

DNA extraction and plasmid replicon typing

Overnight cultures were set up in 2 ml LB broth in 24 well plates. These were centrifuged at 5500g for 5 minutes and the pellet resuspended in 0.5 ml molecular grade H2O. This suspension was boiled for 10 minutes at 100°C and then centrifuged at 4500g 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. This method comprises 3 multiplex-PCR panels with 18 primer pairs recognising the major plasmid incompatibility groups of the Enterobacteriaceae (Carattoli et al., 2005; Johnson et al., 2007). Note that this panel is targeted at large conjugative plasmids and is therefore expected to cover smaller less-well-defined plasmids. PCR reagent concentrations were as follows: 0.4 mM dNTPs, 0.625 units Taq (Qiagen) and 2.4 μM of each primer in the multiplex panel. About 0.5 μl boiled lysisate template was added to each 24.5 μl 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 extension (72°C, 90 seconds) and a final extension, step of 72°C for 5 minutes.

E. coli genotyping

E. coli isolates were genotyped by sequencing the variable region of the flagellin 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 polymorphism) of flagellin genes (Machado et al., 2000). The flic gene was amplified using the PCR-RFLP primers flic 1 (5’-3’ CAAGCTATTACGCAACACG) and flic 2 (5’-3’ GACATRTRGAVACTTCSCTG) (Machado et al., 2000). Cycling conditions were as follows: initial denaturation, 95°C, 2 minutes, followed by 30 cycles of denaturation (95°C, 40 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 Scientific) and TSAP (Promega UK, Southampton, UK) both at final concentration 0.045 units μl⁻¹. Clean up mixes were incubated at 37°C for 30 minutes, followed by 80°C for 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 sequencing runs were conducted by the Department of Zoology, University of Oxford, UK. Trimmed sequences were searched against the whole flagellin gene sequences of 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%) 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 type. E. coli H-antigen types were placed into one of four subgroups defined by Wang et al. (2003) (Supporting Information Table 2). These are based on H-antigen groupings of variable 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 C1 and C2 regions.

Antibiotic susceptibility screening

The susceptibility of E. coli isolates to ampicillin, cefotaxime and chloramphenicol was assessed by disc diffusion assays. MIC breakpoints and interpretation of zone diameters were taken from the British Society for Antimicrobial Chemotherapy Guidelines version 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 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⁻¹ ampicillin; 5 and 30 μg ml⁻¹ cefotaxime; 10 and 30 μg ml⁻¹ chloramphenicol) and strains were determined to be resistant if they scored positively with both discs.

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 effect of E. coli.
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).

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.

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.

Bergstrom, C.T., Lipsitch, M., and Levin, B.R. (2000) Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics 155: 1505–1519.

Bouma, J.E., and Lenski, R.E. (1988) Evolution of a bacteria plasmid association. Nature 335: 351–352.

Buckling, A., and Brockhurst, M.A. (2008) Kin selection and the evolution of virulence. Heredity 100: 484–488.

Carattoli, A. (2008) Animal reservoirs for extended spectrum β-lactamase producers. Clin Microbiol Infect 14: 117–123.

Carattoli, A. (2009) Resistance plasmid families in enterobacteria. Antimicrob Agents Chemother 53: 2227–2238.

Carattoli, A. (2011) Plasmids in Gram negatives: Molecular characterization of conjugative antibiotic resistance plasmids isolated from bacterial communities. Mol Gen Genet 283: 471–482.

Dahlberg, C., and Chao, L. (2003) Amelioration of the cost of conjugal plasmid carriage in Escherichia coli K12. Genetics 165: 1641–1649.

Datta, N., and Hughes, V.M. (1983) Plasmids of the same Inc groups in Enterobacteria before and after the medical use of antibiotics. Nature 306: 616–617.

De Meeus, T., and Goudet, J. (2007) A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. Int Genet Evol 7: 731–735.

Dierikx, C., van Essen-Zandbergen, A., Veldman, K., Smith, H., and Mevius, D. (2010) 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 Duikeren, E., Schoormans, A.H.W., van Essen-Zandbergen, A., Veldman, K., Kant, A., et al. (2012) Occurrence and characteristics of extended-spectrum-β-lactamase- and AmpC-producing clinical isolates derived from companion animals and horses. J Antimicrob Chemother 67: 1368–1374.

Dionisio, F., Matic, I., Radman, M., Rodrigues, O.R., and Taddie, F. (2002) Plasmids spread very fast in heterogeneous bacterial communities. Genetics 162: 1525–1532.

Dionisio, F., Conceicao, I.C., Marques, A.C.R., Fernandes, L., and Gordo, I. (2005) The evolution of a conjugal plasmid and its ability to increase bacterial fitness. Biol Lett 1: 250–252.

Dröge, M., Puhlner, A., and Selbitshka, W. (2000) Phenotypic and molecular characterization of conjugative antibiotic resistance plasmids isolated from bacterial communities of activated sludge. Mol Gen Genet 263: 471–482.

Eberhard, W.G. (1990) Evolution in bacterial plasmids and levels of selection. Q Rev Biol 65: 3–22.

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 phenotypically diverse and confer variable context-dependent fitness effects. Environ Microbiol 17: 5008–5022.

Harrison, E., and Brockhurst, M.A. (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends Microbiol 20: 262–267.

Harrison, E., Guymer, D., Spiers, A.J., Paterson, S., and Brockhurst, M.A. (2015) Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. Curr Biol 25: 2034–2039.

Holsinger, K., and Weir, B.S. (2009) Genetics in geographically stuctured populations: Defining, estimating and interpreting FST. Nat Rev Genet 10: 639–650.

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 β-lactamases circulating among Salmonella enterica isolates. Antimicrob Agents Chemother 50: 3203–3206.

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 plasmid mediated ESBL/AmpC-producing Escherichia coli isolated from veal calves at slaughter. PLoS ONE 8: e65681.

Hu, X., Van der Auwera, G., Timmer, S., Zhu, L., and Mahillon, J. (2009) Distribution, diversity, and potential mobility of extrachromosomal elements related to the Bacillus anthracis pXO1 and pXO2 virulence plasmids. Appl Environ Microbiol 75: 3016–3028.

Jenkins, M.B., Hartel, P.G., Olexa, T.J., and Studemand, J.A. (2003) Putative temporal variability of ribotypes from yearling steers. J Environ Qual 32: 305–309.

Johnson, T.J., Wannemuehler, Y.M., Johnson, S.J., Logue, C.M., White, D.G., Doekkott, C., and Nolan, L.K. (2007)
Plasmid replicon typing of commensal and pathogenic Escherichia coli isolates. *Appl Environ Microbiol* **73**: 1976–1983.

Kav, A.B., Sasson, G., Jamé, E., Doron-Faigenboim, A., Benhar, I., and Mizrahi, I. (2012) 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 β-lactamases or AmpC β-lactamases in food-producing animals. *Clin Infect 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., Grimon, F., and Grimon, 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., Langa, E., and Malamy, M.H. (1985) F-factor inhibition of conjugal transfer of broad-host-range plasmid-p47 – requirement for the protein product of pif operon regulatory gene PifC. *J Bacteriol* **163**: 1067–1073.

Modi, R.I., and Adams, J. (1991) Coevolution in bacterial-plasmid populations. *Evolution* **45**: 656–667.

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-project.org/package=vegan.

Pallecchi, L., Lucchetti, C., Bartoloni, A., Bartalesi, F., Mantella, A., Gamboa, H., et al. (2007) Population structure and resistance genes in antibiotic-resistant bacteria from a remote community with minimal antibiotic exposure. *Antimicrob Agents Chemother* **51**: 1179–1184.

Riccobono, E., Di Pilato, V., Di Maggio, T., Revollo, C., Bartoloni, A., Pallecchi, L., and Rossolini, G.M. (2015) Characterization of IncI1 sequence type 71 epidemic plasmid lineage responsible for the recent dissemination of CTX-M-65 extended-spectrum β-lactamase in the Bolivian Chaco region. *Antimicrob Agents Chemother* **59**: 5340–5347.

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**: 601–612.

San Millan, A., Peña-Miller, R., Toll-Riera, M., Halbert, Z.V., McLean, A.R., Cooper, B.S., and 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?. *Hereditas* **113**: 32–41.

Van Kessel, J.S., Pachepsky, Y.A., Shelfon, D.R., and Karsn, 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 invisibility. *Isme J* **9**: 2328.

Villas-Boas, G., Villas-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., Garcia-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., Rothermund, 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.