Submission Type: Article, Discoveries

Article Title: Coevolution with bacteriophages drives genome-wide host evolution and constrains the acquisition of abiotic-beneficial mutations

Authors Pauline D. Scanlan[1,2,*], Alex R. Hall[1,3], Gordon Blackshields[4], Ville-P. Friman[1,5] Michael R. Davis Jr[6], Joanna B. Goldberg[6,7] & Angus Buckling[1,8].

Affiliations

[1] Department of Zoology, South Parks Road, University of Oxford, Oxford OX1 3PS, United Kingdom

[2] Present Address: Teagasc Food Research Centre, Moorepark, Fermoy, Co Cork, Ireland. Email: Pauline.scanlan@teagasc.ie or paulinescanlan@yahoo.co.uk

[3] Present Address: Institute of Integrative Biology, ETH Zürich, 8092 Zürich, Switzerland. Email: alex.hall@env.ethz.ch

[4] Central Pathology Laboratory, Department of Histopathology and Morbid Anatomy, School of Medicine, Trinity College Dublin, Ireland. Email: gblackshields@gmail.com

[5] Present Address: Imperial College London, Department of Life Sciences, Silwood Park Campus, Buckhurst Road, Ascot Berkshire, SL5 7PY, UK. Email: vifriman@gmail.com

[6] Department of Microbiology, Immunology and Cancer Biology, University of Virginia Health System, Charlottesville, Virginia, USA. Email: michaelrdavisjr@gmail.com

[7] Joanna B. Goldberg. Present Address: Department of Paediatrics, Division of Pulmonology, Allergy/Immunology, Cystic Fibrosis and Sleep, Emory University School of Medicine, Rollins Research Center, 510 Clifton Road NE, Suite3009 Atlanta GA 30322, USA. Email: joanna.goldberg@emory.edu

[8] Angus Buckling. Present address: University of Exeter, Penryn Campus, Cornwall, TR10 9EZ, UK. Email: A.J.Buckling@exeter.ac.uk

***Corresponding author:** Pauline Scanlan, Address: Teagasc Food Research Centre, Moorepark, Fermoy, County Cork, Ireland. Email: pauline.scanlan@teagasc.ie.

Abstract

Studies of antagonistic coevolution between hosts and parasites typically focus on resistance and infectivity traits. However, coevolution could also have genome-wide effects for the host due to pleiotropy, epistasis or selection for evolvability. Here we investigate these effects in the bacterium *Pseudomonas fluorescens* SBW25 during ~400 generations of evolution in the presence or absence of bacteriophage (coevolution or evolution treatments respectively). Coevolution resulted in variable phage resistance, lower competitive fitness in the absence of phages, and greater genome-wide divergence both from the ancestor and between replicates, in part due to the evolution of increased mutation rates. Hosts from coevolution and evolution treatments had different suites of mutations. A high proportion of mutations observed in coevolved hosts were associated with a known phage target binding site, the Lipopolysaccharide (LPS), and correlated with altered LPS length and phage resistance. Mutations in evolved bacteria were correlated with higher fitness in the absence of phages. However, the benefits of these growth-promoting mutations were completely lost when these bacteria were subsequently coevolved with phages, indicating that they were not beneficial in the presence of resistance mutations (consistent with negative epistasis). Our results show that in addition to affecting genome-wide evolution in loci not obviously linked to parasite resistance, coevolution can also constrain the acquisition of mutations beneficial for growth in the abiotic environment.

Introduction

Parasites are abundant in all natural ecosystems and play a central role in the evolution and ecology of their hosts (Woolhouse et al. 2002). Moreover, antagonistic coevolution (defined as the reciprocal evolution of host resistance and parasite infectivity) is recognised as a key driver of genetic divergence between populations of coevolving hosts and parasites, and by extension, antagonistic coevolution is considered a crucial factor in the origin and maintenance of biodiversity (Buckling and Rainey 2002; Thompson 2005). Host-parasite coevolution between bacteria and their lytic viral parasites, bacteriophages, is studied extensively to understand the cause and consequence of coevolution, as coevolution can be observed over a matter of days and weeks (Bohannan and Lenski 2000). However, in addition to their utility for research, using bacteria and bacteriophage to study coevolution also has important implications for our understanding of the ecology and evolution of microbial populations, the role of phage in the evolution of bacterial virulence and the use of phage in clinical settings (Pirnay et al. 2011; Scanlan and Buckling 2012; Hosseinidoust et al. 2013; Koskella and Brockhurst 2014).

Coevolution with bacteriophages has many important phenotypic consequences for bacteria, including costs of resistance (Bohannan and Lenski 2000), effects on diversity and niche competition in spatially structured environments (Buckling and Rainey 2002; Brockhurst et al. 2004), the social behavior of hosts (Morgan et al. 2012) and evolution of mutation rates (Pal et al. 2007). However, little is known about the genomic consequences of bacterial coevolution with phages. While there have been a number of studies using whole genome sequencing (WGS) of coevolved bacteria (Kashiwagi and Yomo 2011; Marston et al. 2012; Le et al. 2014), interpreting the underlying genomic changes is problematic as populations are also adapting to the abiotic environment. An understanding of the genomic consequences

Alex 20/12/14 12:24

Comment [1]: Does this mean reduced growth rate due to phage, or due to resistance? "Reduced population size" or "costs of resistance" would be better.

of coevolution therefore requires a systematic comparison of evolved and coevolved bacterial populations, which is the aim of the current study. Such an experimental approach has been successful in studying the effects of coevolution on virus populations and has revealed that coevolution accelerates molecular evolution in traits linked to infectivity (Paterson et al. 2010; Kashiwagi and Yomo 2011), however a similar comparison has yet to be conducted on bacteria.

Nonetheless, previous theoretical and empirical studies allow us to make general predictions on the potential genomic effects of coevolution with phage. First, we expect greater genomic divergence of hosts among replicates and from the ancestor in coevolving populations. This is because coevolution creates a continually changing selection pressure by generating parasites with novel infectivity alleles and can also select for mechanisms that generate genetic variation i.e. elevated mutation rates and recombination (Hamilton et al. 1990; Pal et al. 2007). This prediction is further supported by work on coevolving/evolving viruses and phenotypic studies of bacteria (Buckling and Rainey 2002; Brockhurst et al. 2004). Second, we expect different suites of mutations to be acquired in evolved and coevolved bacteria. Somewhat trivially, we only expect resistance mutations to bacteriophage to be acquired in coevolving populations but a less obvious prediction is that mutations acquired in evolved populations will not be acquired in coevolved populations, despite coevolved hosts experiencing a superset of the selective pressures (i.e. not just abiotic but abiotic plus biotic) experienced by evolved bacteria. This is because coevolution may slow down the acquisition of some beneficial mutations as a result of competition between mutations (clonal interference) (Felsenstein 1974; Gerrish and Lenski 1998) and lower population sizes caused by coevolving parasites (Zhang and Buckling 2011) as well as epistatic interactions between resistance and other mutations (Lenski 1988; Weinreich et al. 2005; Buckling et al. 2006).

To address these predictions we characterised the genetics of resistance evolution to phage during antagonistic coevolution, detailed the effects of coevolution with phage on genome-wide evolution of bacterial hosts and investigated the relative costs of adaption to the abiotic environment in the presence and absence of coevolving parasites (hereafter referred to as coevolution and evolution treatments). To do so we carried out phenotypic and genomic analyses of the bacterium *Pseudomonas fluorescens* SBW25 that had been evolved in isolation or coevolved with dsDNA lytic phage (Phi2) for up to 400 generations in nutrient media. These organisms initially undergo an arms race dynamic in these conditions, with bacteria and phage continually evolving resistance and infectivity, while retaining resistance/infectivity to previously encountered populations (Buckling and Rainey 2002). Both increased resistance and infectivity ranges are associated with a growth rate cost relative to ancestral genotypes, and the arms race dynamic can shift to a fluctuating dynamic (temporal changes in the frequency of specialised genotypes) as a result (Hall et al. 2011b). Our initial approach was to focus on the phenotypic and genomic characterization of single genotypes from replicate populations evolved in the presence or absence of phages at the final time point of a 400 generation evolution experiment (Hall et al. 2011b) in order to to link phenotype with genotype (which is not possible with whole population genomic characterisation) and investigate genome-wide effects of bacterial adaption in multiple independently-evolved genomes from both experimental treatments. We also sequenced additional phage-resistant genotypes from all coevolved populations after approximately 70 generations, and from a single population at multiple earlier time points, in order to track molecular evolution over time.

Results

Phenotypic evolution

Coevolved bacteria showed extensive and variable resistance to both sympatric and allopatric coevolving phage, whilst bacteria evolved in the absence of phage (evolution treatment) were sensitive to all coevolved phage genotypes they were tested against (fig. 1A and supplementary fig. S1). Fitness (measured as bacterial growth relative to the ancestral genotype in pairwise competition experiments) in the absence of phage was approximately 25% higher in control compared to both coevolved hosts ($t_{10} = 5.24$, $p < 0.0002$) and the ancestor, ($t_5 = 10.98$, $p < 0.0001$), with no difference in fitness between coevolved hosts and the ancestor ($t_5 = 0.1014$, $p = 0.9232$; fig. 1B). These data indicate that coevolution had the net effect of constraining adaptation to the abiotic environment (M9 King's Media B).

Genome evolution

Initially we sampled a single genotype from each population at T60 (coevolved (n = 6) and evolved (n = 6) for whole genome sequence analysis. Coevolved T60 genotypes had on average ~10× as many non-synonymous mutations as T60 evolved genotypes (respectively, 50 ± 36.42 and 4.3 ± 0.81; mean ± S.D; $t_{10} = 3.07$, $p = 0.0059$; fig. 2A), as well as a much greater variation in the total number of mutations (including both non-synonymous and synonymous mutations) (fig. 2B and Table 1). Genetic divergence among coevolved genotypes was much greater than that among evolved genotypes (fig. 3): the mean number of genes that were differentially mutated between a given pair of genotypes was 95.7 (s.e. = 42.1) and 7.6 (s.e. = 3.0) respectively; this distance measure equals the square of the Euclidean distance, as established as an appropriate measure of genetic divergence (Excoffier et al. 1992; Paterson et al. 2010).

We suspected that the high mean and variance in the number of mutations might have resulted from the evolution of mutator phenotypes in some coevolved populations, as previously observed in this system (Pal et al. 2007). Indeed, 5 out of 6 of the T60 coevolved genotypes (compared with 0 out of 6 of the evolved genotypes) had mutation rates at least 10-fold higher than the ancestor (Table 1); notably, one coevolved non-mutator genotype had a comparable number of mutations to the evolved genotypes. For 3 of these genotypes the mutator phenotype was associated with mutations in one or more Mismatch Repair (MMR) genes (*mut*S*, uvr*D or *uvr*B) (Table 1), but for two genotypes the definitive cause of the mutator phenotype was not apparent.

Coevolved and evolved genotypes had different sets of mutations: the same genes were more likely to be mutated within- than between- treatments (permutation test: $F_{1,10} = 1.23$, $p = 0.004$, see also fig. 3). To determine whether bacteria in both treatments had distinct sets of mutations, as opposed to evolved bacteria simply acquiring fewer mutations than coevolved genotypes, we repeated our analysis but restricted it to either genes mutated only in coevolved populations (n = 279) or to genes mutated only in control populations (n = 15). In both cases, genotypes were still more similar within- than between-treatments (genes mutated in coevolved treatment: $F_{1,10} = 1.17$, $p = 0.004$; genes mutated in evolved treatment: $F_{1,10} = 2.33$, $p = 0.026$; fig. 3 and fig. 4A-B).

Genetics of host resistance evolution during antagonistic coevolution

Due to the selection of mutator phenotypes in coevolved populations, many of the mutations observed in this treatment affected a wide range of traits, many of which showed no obvious link with phage resistance (i.e. associated with LPS or cell wall biogenesis). For example 82% of the 201 non-redundant genes with amino acid changing mutations to which we could

give a functional gene classification were assigned to one of fifteen functional categories (other than those associated with resistance) including many linked to metabolic processes such as amino acid transport and metabolism, carbohydrate metabolism, and inorganic ion transport, supplementary fig. S2. Nonetheless, some of these genes may have a more cryptic role in phage resistance and we also considered the possibility that on-going coevolution may have resulted in additional resistance mechanisms not related to LPS: two genotypes coevolved for 400 generations also had mutations in one of two different putative genes (based on homologies with *E. coli*) required for phage reproduction and growth (Czyz et al. 2001; Qimron et al. 2006).

Despite the lack of clear association between many mutations and phage resistance, mutations that were present in coevolved but not evolved genotypes do shed light on the genetic bases of resistance in this system. Of particular note are the many mutations in genes encoding the Lipopolysaccharide (LPS) component of the outer membrane that were present in all coevolved genotypes (see supplementary Table 1): Phi2 shares significant homologies with *Escherichia coli* bacteriophage T7, which is known to interact with LPS (Tamaki et al. 1971). Thirty-six (18%) non-redundant mutations in the long-term coevolution dataset (including data from the mutator genotypes) were associated with LPS and cell envelope biogenesis. To further investigate the importance of LPS for phage resistance we sequenced earlier non-mutator genotypes, which had accumulated fewer mutations. Specifically, we sequenced one genotype from each of the 6 coevolved populations, after approximately 70 bacterial generations or 10 transfers (T10), as well as 4 additional genotypes from a single population from multiple earlier time points of coevolution (after approximately 14 (T2), 28 (T4), 42 (T6), 56 (T8) and 70 (T10) generations). All genotypes had evolved some phage resistance (supplementary fig. S3A) and all 4 of the earliest genotypes only had mutations in

genes associated with LPS biogenesis (supplementary fig. S3B and supplementary table S2). Of the 6 genotypes coevolved for 70 generations, (each with a unique resistance phenotype, supplementary fig. S1), 10 out of the 15 non-synonymous mutations present in this dataset were in genes associated with LPS biogenesis (supplementary table S2). To determine if these mutations in LPS-associated genes actually affect LPS structure, we determined LPS length for coevolved and evolved genotypes together with the ancestral host. The LPS length of a bacterium is related to the number of O-antigen repeated units comprising the polymer. The chain length refers to the number of these repeated units which can vary by strain. As an example, for *E. coli*, the number of repeated units can result in LPS that is very short (1 to 7 units), short (7 to 16 units), intermediate (10 to 18 units) or long (16 to 25 units) (Franco et al. 1998). While the ancestral and evolved genotypes had the same short LPS of just a few O-antigen repeats, coevolved genotypes had one of 4 different LPS types: very short, short, intermediate or long (supplementary table S3).

Mutations associated with evolved populations

Given the significant increase in competitive fitness observed for genotypes evolved in the absence of phages (figure 1B), it is likely that at least some of the genes mutated in these evolved (but not coevolved) genotypes play a key role in bacterial growth under these experimental conditions. A total of 32 mutations were identified in the 6 evolved genotypes. Of these 32 mutations, 17 (53%) occurred in just four genes (PFLU_0185, *alg*U, *wws*F, and PFLU_4418). Moreover, 5/32 (15%) of the mutations were associated with a region of unknown function (PFLU_0596 to PFLU_0598) indicating that this site may be under selection and encode some function important for adaption to the abiotic environment. Collectively, 68% of all mutations detected in the evolved genotypes were restricted to just these five common sites and all evolved genotypes shared one or more mutations, see fig. 4B.

For example, two genotypes (E1 and E2) had mutations in all five commons sites, one genotype had mutations in four out of the five commons sites (E4) and one genotype had mutations in three out of five common sites (E6). PFLU_0185 contains both a GGDEF and an EAL domain, both of which affect bacteria growth (Ryjenkov et al. 2005; Wang et al. 2010); *wws*F has high homologies to *fts*Q which encodes an essential cell division protein in *E. coli* (Chen et al. 1999), and PFLU_4418 has high homologies to *par*A, which encodes an ATPase in *P. aeruginosa* (Lasocki et al. 2007). Given the low number of mutations but large increase in fitness for evolved genotypes, these mutations are likely to have conferred a significant selective advantage.

Mutations in these genes were notably absent in the coevolved genotypes with the exception of a single mutation in PFLU_0185 detected in one coevolved genotype (C4) and another mutation in PFLU_4418 detected in one other coevolved genotype (C1). Only one gene with mutations was detected in more than one genotype in both treatment groups and this was *algU*. Mutations in *algU* were present in 5/6 evolved genotypes (fig. 4B) and 4/6 coevolved genotypes. However, as this gene encodes an alternative sigma factor that plays a regulatory role in a number of different cellular functions including the bacterial stress response, biofilm formation, conversion to mucoidy (a phenotype associated with phage resistance (Scanlan and Buckling 2012)) and motility (Schurr et al. 1995; Garrett et al. 1999), it may be under different selection pressures in the coevolution and evolution treatments. *algU* is an exception and the number of genes with shared mutations between the two treatments was very low (the number of genes with mutations identified in the control line that were also present in the co-evolution comprised only a small percentage (0.02%) of the mutations identified in the co-evolution dataset). A key question that arose from our whole genome sequence analysis is

why weren't a greater number of these putative beneficial mutations present in coevolved populations?

Coevolution of evolved genotypes and evolution of coevolved genotypes

To provide some insight into why a greater number of mutations that are linked to a growth rate advantage were not present in coevolved hosts, we determined how fitness of T60 evolved genotypes changed when they were subsequently coevolved with phages ("evolved-coevolved" T10 genotypes). As a control for this experiment we also evolved the T60 coevolved bacteria in the absence of phage ("coevolved-evolved T10" genotypes), for the same period of time. Coevolving populations may not have enough time to acquire these beneficial mutations because of clonal interference or reduced population size. As such, fitness of the "evolved-coevolved T10" genotypes should be higher than ancestral bacteria coevolved for the same amount of time ("coevolved T10"). By contrast, if the growth-beneficial mutations had a relatively small benefit, or were disadvantageous as a result of coevolution with parasites, fitness of "evolved-coevolved T10" should be comparable to "coevolved T10" genotypes. Similarly, we would not expect any increase in fitness of the "coevolved-evolved T10" genotypes if the growth-beneficial mutations had a relatively small benefit, or were disadvantageous as a result of coevolution with parasites. We therefore coevolved the ancestral bacterium and the evolved T60 genotypes (E1-E6, evolved for 400 generations) in the presence of phages and evolved the coevolved T60 genotypes (C1-C6, coevolved for 400 generations) in the absence of phage for an additional 70 generations (10 transfers), before comparing their competitive fitness scores in the absence of viruses. Fitness of the evolved genotypes was dramatically reduced following coevolution (evolved T60 versus "evolved-coevolved T10" $t_{10} = 3.214$, $p < 0.01$; fig. 5), and was not significantly different from the "coevolved T10" bacteria ($t_{10} = 0.5074$, $p = 0.3115$; fig. 5). In our control

experiment, fitness of the "coevolved-evolved T10" genotypes was no different following evolution in the absence of phage for 70 generations (see supplementary Figure S4).

Asymmetry in evolutionary potential between bacteria and bacteriophage

We have previously shown that viral infectivity evolution appears to be restricted to just one or possibly two loci (Paterson et al. 2010; Scanlan et al. 2011). Moreover, there is strong positive correlation between the number of amino acid changes in phage infectivity loci and phage host range (Hall et al. 2011a; Scanlan et al. 2011) demonstrating a constraint on the potential for phages to evolve elevated infectivity ranges. To investigate if similar genetic constraints operate on the evolution of bacterial resistance ranges, we looked at the genetic data from the T2 -T10 bacterial hosts and compared it with the dynamics of genetic change observed for these previously characterized phage they had coevolved with (Scanlan et al. 2011). In contrast to phages (Hall et al. 2011a; Scanlan et al. 2011), there was no significant correlation between the number of mutations and resistance ranges evolved during early coevolution and resistance was linked to mutations at multiple different loci, indeed, single mutations at different loci involved in LPS and cell wall biogenesis confer highly variable resistance ranges (supplementary fig. S3A-B)

Discussion

Here we show that coevolution with viral parasites drives genome-wide evolution and genetic divergence of their bacterial hosts. Our results also indicate a novel cost of host-parasite coevolution in that coevolution with parasites constrains the acquisition of mutations important for host adaptation to the abiotic environment. An analogous finding has been reported for coevolving viruses adapting to temperature stress (Zhang and Buckling 2011), although the genomics underlying this was not investigated.

Although the evolution of mutators in at least 4/6 coevolved populations precluded the identification of genes specifically associated with resistance evolution, sequence analysis of an additional 12 hosts from earlier on in the coevolutionary process enabled us to characterise the genetics of resistance evolution through time and across different populations and also link genotype to phenotype. Resistance evolution to phage in this study system was due to mutations associated with a range of genes involved in LPS or cell wall biogenesis. Additionally, single mutations at different loci identified in hosts isolated from early on in the coevolutionary process conferred different resistance phenotypes and resistance ranges. We also tested all evolved hosts for phage resistance; they all remained sensitive to phage, indicating that there was no correlation between adaption to the abiotic environment and resistance evolution as has been shown previously for other bacteria and phage combinations (Meyer et al. 2010).

Structural analysis of evolved and coevolved bacteria supported our genetic analysis and shows that LPS structure is indeed a crucial determinant of host resistance in this system. LPS (also commonly referred to as endotoxin) is a complex trait with numerous different

genes contributing to the making, processing, assembly and export of this structure which forms an integral component of bacterial cell structure and functionality. LPS is also a significant virulence factor in a number of human pathogens; 100s of LPS variants have been reported for single species and variation in LPS can account for differences in both virulence and antimicrobial resistance (Banemann et al. 1998; Fierer and Guiney 2001). Our data clearly shows that changes in LPS structure are a key determinant of resistance/susceptibility to coevolving phages and that coevolution with phage drives variation in LPS structure. We observed four general types of LPS banding, ranging from short to very long and although it is not known exactly how these changes in LPS structure affect resistance, it is possible that mutations conferring a very short LPS phenotype result in the loss of the phage receptor whereas mutations conferring a long LPS phenotype mask the receptor site or result in structural changes that prevent access to the receptor. Together with our genetic analysis (where we observed different mutations at multiple loci linked to LPS and cell wall biogenesis), these data indicate that there is considerable genetic and structural plasticity in LPS for our host bacterium. This helps explain how LPS, as the parasite binding site, can support long-term coevolutionary dynamics in this system.

Despite the fact that we observed a high number of non-synonymous mutations with no obvious link to resistance genes for our T60 hosts (> 80%), the number of genes with mutations shared between the coevolved and evolved genotypes was extremely low. This general absence of mutations in genes beneficial for the abiotic environment in the coevolved T60 genotypes represents a cost of resistance to parasites that is distinct from the well-documented cases of antagonistic pleiotropy (Bohannan and Lenski 2000). Correspondingly, our subsequent experiments revealed that mutations present in the evolved genotypes no longer conferred any fitness benefit after coevolution with parasites. Indeed, fitness of these

subsequently coevolved control lines was the same as that of lines coevolved for the same amount of time but without the 400 generations prior adaptation to the abiotic experimental conditions. Moreover, in our control experiment coevolved bacteria did not show a significant increase in fitness after evolution in the absence of phages. This strongly suggests that absence of growth-beneficial mutations in coevolving hosts was not simply the result of a reduction in the rate of acquisition of beneficial mutations (through clonal interference (Felsenstein 1974; Gerrish and Lenski 1998) or reduced population size), but the result of negative epistasis between growth-promoting and resistance mutations. This negative epistasis might have arisen because resistance mutations may act as a limiting factor in maximizing growth rate (by for example, limiting the rate of uptake of key nutrients), hence increases in growth rate resulting from other mutations would have little effect. While the operation of epistasis between resistance and other mutations is consistent with previous work, definitive proof of this hypothesis would require the re-construction of genotypes with both the growth-promoting mutations and resistance mutations (Weinreich et al. 2006; Chou et al. 2011; Khan et al. 2011; Plucain et al. 2014), as well as detailed functional analyses of these mutations. Unfortunately, given the vast number of mutations observed here, these experiments are beyond the scope of the current work.

The increased molecular evolution and among-replicate divergence of coevolved hosts is consistent with Van Valen's "Red Queen" hypothesis (van Valen 1974, 1973) and associated empirical studies that antagonistic coevolution accelerates molecular evolution (Hedrick 1994; Obbard et al. 2006; Paterson et al. 2010). The proximate mechanism for these findings in our study was likely to have been primarily the increased mutation rates of most of the coevolved genotypes, as has been observed previously (Pal et al. 2007). The evolution of increased mutation rates during coevolution with phages is presumably because of linkage

between mutator alleles and the adaptive resistance mutations that they rapidly generate (Pal et al. 2007). Crucially, the majority of mutations resulted in amino acid changes in genes with no obvious link to resistance in our T60 genotypes, suggesting a causal link between coevolution with parasites and functional changes in non-resistance traits. It is unclear what selective advantage, if any, these additional mutations might confer, as many are possibly deleterious, but some of them may buffer against the pleiotropic costs of resistance mutations; another example of epistasis (Maisnier-Patin and Andersson 2004; Poon et al. 2005). Comparable functional changes in non-resistance traits are likely to arise under conditions where coevolution has been shown to select for increased recombination in eukaryotic systems (Morran et al. 2011). Note that the increased divergence between coevolved replicates may also have been driven to some extent by genetic bottlenecks imposed by selective sweeps of resistant mutants (Buckling and Rainey 2002). By contrast, bottlenecks resulting from genetic drift are unlikely to have played a major role, as coevolved population sizes under almost identical experimental conditions were consistently high (never below 10^8), and populations were only diluted 100-fold at each transfer.

Finally, our results may also help to explain why bacteria appear to have an evolutionary advantage over phages in this system (Buckling and Rainey 2002). We have previously shown that there is strong positive correlation between the number of amino acid changes in phage infectivity loci and phage host range (Hall et al. 2011a; Scanlan et al. 2011), and that multiple mutations are required to infect certain highly resistant hosts. Our analyses of genetic data from bacteria (T2-T10) that had coevolved with previously characterized phage (Scanlan et al. 2011) show that the evolution of broad resistance range can occur much more readily, sometimes through single mutations, than the evolution of broad infectivity ranges.

Materials and Methods

Experimental design. We established two treatments with six replicate microcosms per treatment from isogenic stocks of host bacterium *Pseudomonas fluorescens* SBW25 and lytic viral parasite bacteriophage Phi2 (coevolution with phage and evolution with no phage) as outlined previously (Hall et al. 2011b). In brief: microcosms containing 6 mL M9KB medium (M9 salt solution supplemented with 10 g L glycerol and 20 g L proteose peptone) were inoculated with ~ 10^8 cells of *P. fluorescens* SBW25 and for the coevolution treatment c. 10^5 particles of bacteriophage Phi2 (Buckling and Rainey 2002; Hall et al. 2011b). Six replicates for each treatment (coevolution and evolution) were maintained by serial transfer, with 100-fold dilution every 48 h, 48 h is the equivalent of one transfer (1T) and one transfer is equal to approximately seven bacterial generations. Microcosms were incubated statically at 28 °C, and vortexed for 1 min (11 000 *g*) to homogenize the culture prior to 1 % v:v transfer to fresh media. Populations were maintained by serial transfer for ~400 bacterial generations (Hall et al. 2011b), and a sample of each was frozen at 80 °C in 20% v:v glycerol every 10 transfers.

Isolation of bacteria and phage. Bacteria and phage were isolated as described earlier (Hall et al. 2011b) and all genetic and phenotypic analysis was conducted at the genotype level. For genetic and phenotypic analysis we isolated twenty-four bacteria representative of the diversity naturally occurring in this system; one from each population of the coevolution (n=6) and control treatments at T60 (n=6), one from each population of the coevolution treatment at T10 (n=6), and one from T2, T4, T6, T8, T10 and T30 from a single population

(n=6). Host resistance assays, mutation rate analysis, whole genome sequencing and fitness assays were all conducted on these single bacterial genotypes, see methods below.

Host resistance assays. Host resistance assays were conducted according to previously optimised methods (Hall et al. 2011b). Each T10 and T60 host genotype, together with the genotypes obtained from control populations, was tested for resistance against 180 different phage phenotypes (ten phage genotypes from each of the six populations at three different time-points 6×3×10 = 180). The resistance profile of each host was assayed using a pin replicator to apply phage to growing lawns of host bacteria that were made using the soft agar overlay method (each host was tested individually and in triplicate). Bacteria were scored as resistant if no plaques were observed in any of the three replicates. The phenotypes of bacteria from early coevolution (T2 to T10 for population C6 only) were characterized in a similar manner in an earlier study (Scanlan et al. 2011).

Coevolution of evolved genotypes and evolution of coevolution genotypes. We took the six single evolved and coevolved genotypes (i.e. those initially evolved for 60 transfers in the absence and presence of phages) isolated from our evolved (E1 to E6) and coevolved (C1 to C6) selection populations at T60 and used these to initiate our "Evolved-Coevolved" (n = 6) and Coevolved-Evolved (n = 6) experimental selection lines, by coevolving each of these genotypes with phage and without for ten transfers (~70 bacterial generations), respectively, as described above. At the end of this experiment, a single genotype from each population was isolated and used to assess fitness.

Competition experiments. Competition experiments were performed by inoculating microcosms with equal densities of a marked strain of ancestral *P. fluorescens* SBW25-*lacZ* and the competitor phenotype, and grown (competed) overnight at 28 °C (Hall et al. 2011b).

474 The numbers of both competitor and marked strains were enumerated by plating onto Luria–
475 Bertani agar plates supplemented with X-gal at the start and end of the assay. The addition of
476 X-Gal allowed us to enumerate the marked strain of *P. fluorescens* SBW25-*lacZ*, which
477 forms distinct blue colonies on this medium. The starting and final densities of both the
478 marked strain and competing phenotype were calculated as CFUs/ml based on direct count
479 data after 48 hours incubation. Each assay was performed in triplicate and fitness (W) was
480 taken as the ratio of the estimated Malthusian parameters (m) of each competing type,
481 $m=\ln(N_f/N_0)$, where N_0 is the starting density and N_f is the final density (Lenski 1991).

482

483 *Fluctuation tests.* We used fluctuation tests to estimate bacterial mutation rates for all isolates
484 as previously described (Pal et al. 2007). Six microcosms per bacterial isolate were
485 inoculated with 100–1000 bacterial cells and were allowed to grow for 24 h shaking in a 28
486 °C incubator. Final cell density was determined by plating dilutions on non-selective solid
487 medium (KB). The number of mutants was estimated by plating 60 µl of each culture on solid
488 selective medium (KB agar plates supplemented with rifampicin (100 µg ml $^{-1}$) or
489 streptomycin (50 µg ml $^{-1}$)). Jones median estimator was used to calculate mutation rate from
490 the average and median frequency of mutant colonies (Rosche and Foster 2000).

491

492 *Bacterial whole genome sequencing and variant detection.* The twenty four different bacterial
493 genotypes that we had characterised phenotypically were used for whole genome sequence
494 analysis. DNA was extracted from pure cultures using the Qiagen DNeasy Blood and Tissue
495 Kit (Qiagen, U.K.) according to the manufacturer's instructions for bacterial DNA extraction.

DNA was eluted in sterile distilled water; DNA quality and quantity were checked by using a Nanodrop and agarose gel electrophoresis prior to sequencing.

Paired-end sequencing was carried out by Source Bioscience (Nottingham, U.K.) using the IlluminaHISeq2000 platform. Inevitable decay in sequence quality towards the end of a sequenced read can introduce higher frequencies of base-call errors, reducing the quality and number of reads that map to a reference sequence, and the efficacy of any downstream analysis. Initial inspection of the reads performed using the FastQC report tool (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) showed that the base-call quality scores (supplementary fig. S5; values calculated across all samples) increasingly dropped below the Q30 threshold value (indicating >= 0.001 probability of an incorrect base-call) as the read-length approached 60bp. Reads were therefore trimmed to 55bp prior to further analysis, to increase the percentage and overall quality of mappable reads. Trimmed reads were mapped to the complete **SBW25** *Pseudomonas fluorescens* reference genome (http://www.ncbi.nlm.nih.gov/nuccore/NC_012660.1) using BWA (Li and Durbin 2009) with default "aln" and "sampe" parameters. On average 98.33 ± 1.7% (mean ± SD) of all reads were mapped, achieving 138.0 ± 49.64x (mean ± SD) average coverage (supplementary table S4). All sequence data is available upon request.

Mapped reads were then processed for variant detection. Potential PCR duplicates were flagged and removed from the output using the *MarkDuplicates* command within the Picard toolkit (http://picard.sourceforge.net). Conversion to pileup format was carried out using samtools *mpileup* (/http://samtools.sourceforge.net/). Taking pileup as input, a combination of samtools and Varscan (Koboldt et al. 2012) were then used to analyse the mapped reads in search of statistically significant differences to SBW25. VarScan (http://varscan.sourceforge.net/), calls variants using a heuristic method and statistical test based on the number of aligned reads supporting each genotype. The VarScan command

m*pileup2cns* calls consensus genotypes at all positions with sufficient coverage (setting --variants 1 filters out non-variant positions). Most parameters were kept at default, with the exception of --min-coverage (minimum number of reads to cover position (30)), --min-reads2 (minimum number of bases at that position that differ from the reference nucleotide (8)), --min-ave-qual (minimum average quality of the bases covering that position (20)), and --p-value (0.01, (Fisher's Exact Test). Results were exported in VCF format. In addition to using stringent thresholds to minimize any false positive SNP detection we used PCR (n = 60) followed by Sanger sequencing as a further control measure to verify the mutations that were called were true positives. We also selected a number of indels from our dataset (n = 15) and verified their presence using PCR and Sanger sequencing as a quality control measure. All PCR verification checks were positive indicating that the number of false positives in the dataset was minimal.

Sequence analysis. Mutations and indel data were compiled and analysed (including functional gene assignment) using a combination of Artemis, EXPasy portal tools, BLAST and data available from the Pseudomonas Genome Database (http://www.pseudomonas.com). To provide a visual overview of genetic distance at the nucleotide distance between different genotypes and to complement analysis of genetic divergence performed at the gene level (see statistical analysis) we constructed a simple distance tree using the Neighbor-Joining method (Saitou and Nei 1987). To create a data file for tree construction we took every nucleotide position where a mutation was identified relative to the ancestral or WT sequence for all T60 genotypes in the dataset. For each genotype this resulted in a contiguous nucleotide string with the mutation present in that genotype(s) or the corresponding WT base given for all other sequenced genotypes where this mutation was absent. This was done for all T60 sequenced genomes, n = 12. The tree is drawn to scale, with branch lengths in the same units

as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of differences method (Nei and Kumar 2000) and are in the units of the number of base differences per sequence. The analysis was conducted in MEGA5 and involved 13 nucleotide sequences with a total of 426 positions in the final dataset. Circular genome plots (shown in fig. 4A and 4B) were constructed in DNAplotter.

Statistical analyses. We tested whether treatment (coevolution/evolution) affected the types of mutations observed using permutational analysis of variance (Zapala and Schork 2006). For each gene, every genotype was scored as being mutated (1) or not mutated (0), before the Euclidean distance between each pair of genotypes was calculated (Excoffier et al. 1992). The probability that the observed ratio of average within- and between- group distances could arise by chance alone was assessed by *F*-test comparing observed values to random permutations of the raw data (Anderson 2001) using *adonis* {*vegan* in R v2.11.1. All other statistical analyses were performed using R or SPSS. In cases where we made multiple tests, such as comparing fitness of control (evolved) bacteria to coevolved and ancestral bacteria separately, we only accepted statistical significance when $P < \alpha=0.05$ adjusted by sequential Bonferroni correction.

Structural analysis of LPS. LPS was extracted from WT, coevolved and evolved genotypes using a hot-phenol method and analyzed as described in (Davis and Goldberg 2012).

Acknowledgements

This work was funded by European Research Council and NERC (UK). We are also thankful to Steve Paterson for useful comments on the manuscript.

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595