Title: The enemy from within: a prophage of *Roseburia intestinalis* systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition

Jeffrey K. Cornuault[1], Elisabeth Moncaut[1], Valentin Loux[2], Aurélie Mathieu[1], Harry Sokol[1,3,4], Marie-Agnès Petit[1], Marianne De Paepe[1]

[1]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

[2]MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

[3] Sorbonne Université, Inserm, Centre de Recherche Saint-Antoine, CRSA, AP-HP, Hôpital Saint Antoine, Service de Gastroenterologie, F-75012 Paris, France.

[4]Department of Gastroenterology, Saint-Antoine Hospital, Assistance Publique – Hôpitaux de Paris, UPMC, Paris, France.

Abstract

Despite an overall stability in time of the human gut microbiota at the phylum level, strong temporal variations in species abundance have been observed. We are far from a clear understanding of what promotes or disrupts the stability of microbiome communities. Environmental factors, like food or antibiotic use, modify the gut microbiota composition, but their overall impacts remain relatively low. Phages, the viruses that infect bacteria, might constitute important factors explaining temporal variations in species abundance. Gut bacteria harbour numerous prophages, or dormant viruses. A breakdown of prophage dormancy can evolve through the selection of ultravirulent phage mutants, potentially leading to important bacterial death. Whether such phenomenon occurs in the mammal's microbiota has been largely unexplored.

Here we studied temperate phage-bacteria coevolution in gnotoxenic mice colonised with *Escherichia coli* and *Roseburia intestinalis*, a dominant symbiont of the human gut microbiota. We show that *R. intestinalis* harbors two active prophages, Jekyll and Shimadzu, and observed the systematic evolution of ultravirulent Shimadzu phage mutants, leading to a collapse of *R. intestinalis* population. In a second step, phage infection drove the fast evolution of host phage-resistance mainly through phage-derived spacer acquisition in a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) array. Alternatively, phage resistance was conferred by a prophage originating from an ultravirulent phage with a restored ability to lysogenize.

Our results demonstrate that prophages are the potential source of ultravirulent phages that can successfully infect most of the susceptible bacteria. This suggests that prophages can play important roles in the short-term temporal variations observed in the composition of the gut microbiota.

Introduction

The mammal's gut microbiota is one of the densest microbial communities on earth. In Humans, its composition is globally conserved among healthy adults (Qin et al. 2010, Arumugam et al. 2011, Consortium 2013), and abnormal compositions, called dysbioses, are correlated with numerous human diseases (Backhed et al. 2004, Larsen et al. 2010, Marchesi et al. 2016). Yet the factors influencing the composition of the gut microbiota in health and disease are far from being entirely understood. For example, in a study on ~1,000 healthy adults, the combined effect of 69 covariates (food, drug, anthropometric factors, etc.) could explain only 16.4% of the variation in microbiota composition (Falony et al. 2016). The impact of host genetics seems to be even lower (Rothschild et al. 2018). In addition, despite overall stability in term of presence or absence of main bacterial genera, strong variations in the relative abundance of species are observed on short time scale, even in healthy individuals submitted to the same diet (Koenig et al. 2011, Flores et al. 2014, Thaiss et al. 2014, Voigt et al. 2015, Sarker et al. 2017).

Bacteriophages, or phages for short, the viruses that infect bacteria, have been suspected to impact intestinal microbiota composition. In spite of clear evidence that phage-mediated selection plays an important role in bacterial strain diversification (reviewed in (Scanlan 2017)), there is as yet little proof that they are important determinants of the microbiota composition (reviewed in (Mills et al. 2013, De Paepe et al. 2014, Manrique et al. 2017)). Some metagenomics studies showed correlations between an increase in specific phages with reductions of particular bacterial taxa (Reyes et al. 2013, Waller et al. 2014). Yet, in most animal experiments using well-defined phage-bacteria pairs, phage-mediated bacterial mortality in the gut environment was limited to a fraction of susceptible bacteria, and was not sufficient to drive the selection of phage-resistant bacteria (Chibani-Chennoufi et al. 2004, Weiss et al. 2009, Maura et al. 2012), suggesting that intestinal physiology complicates phage multiplication (Brussow 2013). In line with this idea, the phage to bacteria ratio is much lower in the gastrointestinal tract than in other environments, suggesting less lytic replication (Silveira and Rohwer 2016). These different pieces of evidence might explain why the study of phages in the gut microbiota lagged behind the study of bacteria, and that phages are still often forgotten in microbiome studies.

Temperate phages, in particular, are generally considered to cause low bacterial mortality. They encode molecular mechanisms enabling them either to reproduce through lytic cycles upon infection, or to lysogenize bacteria, i.e. to establish in a stable dormant state in the infected cell. The dormant phage is called a prophage, the bacterial host a lysogen, and it is estimated that about 80% of bacteria are lysogens in the mammal's gut (Touchon et al. 2016, Cornuault et al. 2018, Kim 2018). In line with this estimation, virions from temperate phages are dominant among intestinal viruses (Reyes et al. 2010, Minot et al. 2011, Knowles et al. 2016). Prophages are generally considered to be beneficial to their bacterial hosts, in particular in the gastrointestinal tract (Manrique et al. 2017, Mahony et al. 2018). The first advantage of lysogeny is superinfection inhibition, i.e. the protection from infection by other phages of the same "immunity" type. Inhibition of the superinfecting phage occurs by the same mechanism that establishes the dormancy of the resident prophage (Berngruber et al. 2010). Secondly, prophages can bring new functions to their host, favouring colonization, growth or stress resistance (reviewed in (Bondy-Denomy and Davidson 2014)). Finally, lysogeny can provide an advantage to the bacterial host by releasing infectious virions that are able to kill susceptible bacterial competitors (Bossi et al. 2003, Brown et al. 2006). Yet, in the gut, this advantage is limited by the frequent lysogenization of susceptible bacteria (Duerkop et al. 2012, De Paepe et al. 2016), as well as by the supposed rarity of susceptible competitors. Therefore, to date, the role of temperate phages in intestinal microbial communities is mostly studied from the perspective of horizontal gene transfer.

Yet some prophages can constitute a major threat for their hosts. In several microbial environments, important induction of prophages has been observed, resulting in the death of a significant part of the bacterial population (Selva et al. 2009, Brum et al. 2016, De Paepe et al. 2016). In addition, prophages of strains used in large-scale industrial fermentations are potential sources of virulent phage mutants infective for the host strain, the so-called ultravirulent mutants (Lucchini et al. 1999, Shimizu-Kadota et al. 2000, Durmaz et al. 2008). However, to the best of our knowledge, even though lysogeny is widespread among intestinal bacteria (Bobay et al. 2014, Touchon et al. 2016, Kim 2018), massive killing of bacteria by a temperate phage has never been documented in intestinal environments. This might result from a potential limited ability of phages to multiply in the gut environment as previously

4

discussed, but also from a surprisingly paucity of studies of temperate phage-bacteria co-evolution in such an important microbial ecosystem (Scanlan, 2017).

The complexity of the gut microbiota prevents any exhaustive characterization of all the virus-bacteria interactions, even with high throughput metagenomic studies. A particular roadblock is the difficulty to predict the bacterial host of the phages detected, in particular for phages with no relatives among reference sequenced viruses (reviewed in (Edwards et al. 2016)). In particular, no phage infecting *Lachnospiracea*, the most abundant bacterial family of the human gut microbiota (Falony et al. 2016), has been deposited in public databases. Here we followed the co-evolution of a *Lachnospiracea* strain*, Roseburia intestinalis* L1-82, and its prophages in gnotoxenic mice. The *Roseburia* genus is of particular interest not only because of its abundance in the gut microbiota, but also because of its important production of short chain fatty acids in the human gut, including butyrate, which have a critical role in the regulation of host immune responses (Rios-Covian et al. 2016). In addition, this genus is depleted in several human diseases, suggesting it might be important for human health (Tamanai-Shacoori et al. 2017). In our study, mice were also colonized with *Escherichia coli*, a sub-dominant member of the human gut microbiota whose abundance increases in several human diseases (Willing et al. 2010). We show that ultravirulent mutants of *R. intestinalis* phage Shimadzu systematically invade this simplified gut ecosystem, and lyse the large majority of susceptible bacteria. Phage mediated bacterial killing drives the fast evolution of host phage-resistance through phage-derived spacer acquisition in a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) array. Our results illustrate that prophages are not always dormant, and can play important roles in the short-term temporal variations observed in the composition of the microbiota.

5

Results

Roseburia intestinalis L1-82 harbors two active prophages.

The sequencing and complete assembly of _R. intestinalis_ L1-82 genome revealed the presence of two complete prophages, that we named Shimadzu and Jekyll (Fig. 1A). The production of virions by these prophages was determined by sequencing encapsidated DNA, i.e. DNA protected from DNAse in a culture supernatant. Resulting reads predominantly aligned on the two identified prophages (60% and 19.4% of total reads respectively), indicating that both prophages produce virions. Of note, read coverage was highly uneven on the Jekyll prophage, suggesting that besides the full size phage genome, a shorter version is also encapsidated (Fig. S1A). Phage genomes were annotated using RAST (Aziz et al. 2008), HHPRED (Zimmermann et al. 2018) and Phagonaute (Delattre et al. 2016) as previously described (Cornuault et al. 2018). Interestingly, the Shimadzu integration site is located within its integrase gene _int_, encoding a large serine recombinase. Consequently, two different integrases are produced during the lytic cycle and during lysogeny (Fig. S1B). The two forms differ by their last 22 amino acids, which might modify the stability of the protein, as described in mycophages (Broussard et al. 2013).

The Jekyll genome had no significant nucleotide similarity with any viral genome of the NCBI RefSeq genome database, indicating that it corresponds to a new viral clade. The Virfam classification tool, that uses virion protein remote homology and gene synteny (Lopes et al. 2014), clustered Jekyll with phages infecting _Bacilli_ species, such as lactobacillus phage LLH. The Shimadzu genome presented homology with the _Faecalibacterium prausnitzii_ phage Lagaffe (Fig. S1C), and grouped within a badly-resolved Virfam cluster comprising _E. coli_ lambda phage. Electronic microscopy showed that Jekyll is a _Siphoviridae_, and Shimadzu a _Myoviridae_ (Fig. 1B), in line with genomic predictions. Quantitative PCR on encapsidated viral DNA indicated that the virion concentration in saturated culture supernatants was 3.10^8 and 8.10^8 virions/mL for Shimadzu and Jekyll respectively. Interestingly, Jekyll, but not Shimadzu, was induced by mytomicin C, a genotoxic agent causing DNA damage, a general signal for prophage derepression (Fig. 1C).

Similarly to _F. prausnitzii_ phage Lagaffe, Shimadzu encodes a Diversity Generating Retroelement (DGR) (Fig. S2). DGRs generate variability in target genes through a reverse transcriptase (RT)-mediated mechanism that introduces nucleotide substitutions at defined

locations, called variable repeats (VR) (reviewed in (Guo et al. 2014)). VR are homologous to a template repeat (TR), located directly upstream the reverse transcriptase, and used as a template to generate variation. As sometimes observed (Wu et al. 2018), two VR could be detected in Shimadzu: the first one, VR_1, is located in a predicted tail gene located just before the reverse transcriptase, a commonly observed position, and the second, VR_2, is located in *orf43*, of unknown function (Fig. 1A and S2B). Analysis of sequence reads obtained from virions collected from *in vitro* culture supernatants indicated that both VR are successfully modified by the DGR: 2 % of reads aligning to the VRs had 10 to 15 mismatches typical of DGR mediated mutagenesis, whereas only 0.1% of reads outside this region exhibited 1 to 2 mismatches.

R. intestinalis and Shimadzu populations exhibit important temporal variations in mice

We then examined free phage and bacterial populations in germ-free mice colonized with *R. intestinalis* L1-82 and either *E. coli* LF82 (experiments 1 and 2) or *E. coli* MG1655 (experiment 3). Each experiment comprised 5 to 8 mice housed in two different cages placed in the same isolator. Phage and bacterial populations were monitored by quantitative PCR and/or microscopy in faeces for 33 days (Fig. S3). *R. intestinalis* became rapidly dominant over *E. coli*, reaching 96 ± 2.5% of the population after 5 days of colonization (Fig. 2A). However, in each cage, a significant collapse of *R. intestinalis* was observed in the majority of mice after 16 to 25 days of colonization (Fig. 2A). Since mice were kept in a constant and closed environment, we hypothesized that one of the two *R. intestinalis* L1-82 prophages could be responsible for this change in bacterial composition. During the first 10 days of colonization, the concentration of Jekyll and Shimadzu virions in mouse faeces was comparable to that observed *in vitro*, around 5 x 10^8 virions/g of faeces (Fig. 2B). Afterward, Jekyll free phage concentration remained constant but Shimadzu concentration suddenly increased by more than a thousand fold, reaching up to 10^{12} virions/g of faeces after 13 to 25 days of colonization, before decreasing and stabilizing around 4 x 10^{10} virions/g (Fig. 2B). Similar results were observed in mice colonized with the two different *E. coli* strains, ruling out potential effects of *E. coli* LF82 virulence genes on phage production.

Shimadzu multiplication is responsible for *R. intestinalis* mortality.

Shimadzu free phage concentration was maximal shortly before the fall of concentration of *R. intestinalis* (Fig. 2C, R_2 = 0.72), strongly suggesting phage-related bacterial mortality. Yet, as

7

discussed earlier, prophages provide resistance to superinfection, i.e. lysogenic bacteria are immune to infection by a phage that is present as a prophage in its genome. The sudden multiplication of Shimadzu could thus result either from the invasion of ultravirulent mutants able to infect lysogenic cells, or from a sudden environmental signal triggering a massive prophage induction. In favour of the first hypothesis, Shimadzu lysis plaques on *R. intestinalis* L1-82 lawn were obtained from mouse faeces collected after 21 days of colonization (Fig. 2D). These lysis plaques, by definition, were formed by ultravirulent mutants. Of note, the number of plaque forming units (PFU) obtained corresponded to 10^2 to 10^6 PFU per gram of faeces, far from the 10^{12} virions estimated by quantitative PCR. Quantification of three phage stocks revealed a 100-fold difference between genome copy numbers determined by qPCR and PFU, suggesting that most virions are uninfectious in these conditions, but not entirely explaining the difference observed in faeces between PFU and virions. This difference could result either from damage of virions during extraction or from genetic heterogeneity in the population of Shimadzu in mouse faeces, with only few phage genotypes being able to infect *R. intestinalis* L1-82 *in vitro*. Nevertheless, the isolation of phages on L1-82 indicated that Shimadzu ultravirulent mutants were present in mice, most probably explaining bacterial mortality.

The bloom of Shimadzu free phage is associated with few dominant mutations.

To identify the mutations responsible of the ultravirulent phenotype, we sequenced five phages, isolated either from experiment 1 (Shivir1) or from experiment 2 (Shivir2, 3 and 4) after 21 days of colonisation, and Shivir5 from extra cage 7 of experiment 2, not fully presented in this report, and isolated after 33 days of colonisation. Each phage genome contained between 34 and 45 mutations, but only four loci were mutated in the five isolates: the two VRs targeted by the DGR (~10-15 mutations per VR), a region comprising two inversely oriented transcriptional regulators, named immunity region, and finally *orf26*, of unknown function (Fig. 3A). To know if these mutation hotspots represented a general trend rather than biases due to phage isolation, we sequenced total phage DNA (virome) isolated from mouse faeces harvested after 21 days of colonisation. Virome 1 corresponds to a pool of faeces from the 8 mice of experiment 1, whereas virome 2 corresponds to a single mouse of experiment 2. In both viromes, mutations in the immunity region, in VR_1 and in VR_2 were respectively present in ~70%, ~65% and 47% of aligning reads. No other locus was mutated in more than 20% of covering reads (Fig. 3A&B). For example, the mutation in *orf26* was present in only 5

% of aligning reads in virome 1, and even less in virome 2, indicating that this mutation is not essential for Shimadzu multiplication in mice. Non-synonymous mutations in all five phage isolates (Fig. 3C) suggests mutations in *orf26* are nevertheless necessary to form lysis plaques *in vitro*, and might explain the discrepancy between PFU and phage copy numbers in faeces discussed above.

In the two viromes, VR_1 and VR_2 variants comprised one to three dominant alleles, with the exception of the VR_2 in virome 1, in which the most abundant allele represented only 3% of reads (Table S1). All ultravirulent isolates from experiment 2 had the same VR_1 allele, which corresponds to the second most dominant allele in the virome 2 (Table S1). By contrast, all isolates carried different VR_2 alleles, and differed from the most abundant ones in the corresponding virome (Table S1). This reflects stronger selection of VR_1 alleles as compared to VR_2, probably resulting from a higher importance of the VR_1 containing tail protein for infection.

The mutated region comprising two inversely oriented transcriptional regulators is most probably the immunity region, that controls prophage maintenance and resistance to superinfection. Indeed, similarly to the immunity region of the model phage lambda, the intergenic region comprises three almost identical repeats, known as operators (O_R) in lambda phage, and a putative promoter identified with the dedicated tool DBTBS (Sierro et al. 2008) (Fig 3B). In phage λ, the two transcriptional regulators *cI* and *cro* are under mutually exclusive expression: during lysogeny, O_{Rs} are fixed by CI repressor, preventing the expression of *cro* and the downstream lytic cycle genes, and vice versa during the lytic cycle. Upon superinfection, CI repressors will also bind the O_R of the incoming phage, preventing the expression of its lysis genes, and conferring immunity. All five ultravirulent phages have a mutation in *cro* and in O_R1 (Fig. 3B&C). These mutations might diminish or even abolish the binding of CI repressor produced by the resident prophage, enabling the incoming phage to multiply in lysogens. Mutations in O_R2 or O_R3 were also observed in some isolates, probably accentuating this phenotype.

Dynamics of mutation selection

To gain insights into the respective role of the mutations described above in Shimadzu multiplication and phage bloom, we then examined their timing of selection (Fig 3C). Since dominant VR_2 alleles in virome 1 represented only 3% of sequences and that each phage

isolate had a different allele, we concluded that the VR_2 containing gene might be less essential for Shimadzu multiplication, and we focused on the VR_1 and the immunity region. Important VR_1 diversification was detected before the onset of the phage bloom in two cages, indicating the selection of mutant alleles prior to phage bloom. This selection suggests that even prior to the bloom, the production of free phage not only results from spontaneous prophage induction but also from an infection process. This infection probably concerns only a small subpopulation of bacteria since neither an increase in free phage titers, nor a bacterial decrease, could be detected at that time. Such VR_1 allele selection could occur if the ancestral tail fiber gene is not permitting efficient infection. We hypothesized that the infected bacterial population could consist of bacteria having lost spontaneously the Shimadzu prophage. Screening of clones isolated either from an *in vitro* culture or from mice after 13 days of colonisation revealed in the two cases a proportion close to 5.10^{-3} cured bacteria, a proportion confirmed by qPCR. We can speculate that these cured bacteria are the susceptible hosts enabling VR_1 allele selection. O_{Rs} mutations became dominant concomitantly with the phage bloom in all mice, definitely associating the major phage multiplication peak with the evolution of ultravirulence.

Ultravirulent phages drive the fast selection of resistant bacteria

Despite the systematic and very important multiplication of Shimadzu in all mice, *R. intestinalis* concentration almost recovered its initial level after 30 days of colonization, and in some mice the bacterial population was barely affected by the ultravirulent phages (Fig. 2). These observations hint at the rapid emergence of resistant bacteria. To investigate this phenomenon, a total of 300 *R. intestinalis* clones were isolated after 16, 19, 25, 28 or 33 days of colonization in experiment 2. Ninety-six of them were grown in the presence or absence of ultravirulent phage Shivir5*,* which showed the best lysis ability of bacterial liquid cultures. For two additional Shivir mutants, a smaller scale analysis (on 10 bacterial clones) by plaque assay revealed an identical susceptibility pattern. The proportion of resistant clones dramatically increased after the phage bloom (Fig. 4A), explaining the recovery of *R. intestinalis* population in mice.

Resistance is mostly related to the acquisition of phage-directed CRISPR-Cas spacer

Bacteria have developed an incredible number of mechanisms to counter phage infections (reviewed in (van Houte et al. 2016)). To get insight into the mechanisms involved, 4 resistant clones of *R. intestinalis* isolated after 28 days of colonization were entirely sequenced. Each of them contained between 38 and 54 mutations, but only gene *sbpE* (RiL182_04007), presenting homology to a *Bacillus subtilis* extracellular solute-binding protein gene, was mutated in all clones. Yet this mutation was dominant in the mouse gut early on, after only 5 to 7 days of colonization (Fig. 4B), and long before the Shimadzu bloom, suggesting that *sbpE* mutations are associated to adaptation to the mouse gastro-intestinal tract rather than to phage resistance.

Sequencing also revealed in three of the four isolates the insertion of a 30 base pair sequence homologous to the Shimadzu genome (Table S2), in a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) array, of type II-C (Fig. 4C). CRISPR and their associated genes, CRISPR-Cas immune systems, provide phage resistance by integrating short phage-derived sequences (spacers), generally at the leader end of the CRISPR array, similarly to what we observed. Type II-C CRISPR-Cas systems function by cleaving invading DNA homologous to spacer sequences. Spacer acquisition was then investigated in the clones tested for Shivir5 susceptibility: more than 80% of the resistant clones had acquired a new spacer, but none of the susceptible clones (Fig. 4C&E), and the proportion of clones with two additional spacers increased with time (Fig. 4F). Sequencing of newly acquired spacers in 10 other clones showed that they all possess one or two Shimadzu-derived spacers, but two clones also possess a spacer that targets the *R. intestinalis* genome (Table S2). The already described cost associated with auto-immunity of CRISPR-cas type II systems was apparent because these two clones displayed altered growth, especially on plates. Finally, spacer sequencing revealed the presence in the same mouse of clones with different spacers, revealing several spacer acquisitions in the same mouse.

To confirm that bacterial resistance through CRISPR-Cas activity is general and does not result from biases in the isolated clones, we monitored spacer acquisition on total bacterial DNA from mouse faeces (Fig. 4G). In each cage, important spacer acquisition was first evidenced at the onset of bacterial recovery (Fig. 4B). In addition, the number of acquired spacer increased with time, exactly as in isolated clones (Fig. 4F). Altogether, these results indicate that CRISPR-Cas activity is the main *R. intestinalis* resistance mechanism against ultravirulent Shimadzu in the mouse gut.

Spacer acquisition is associated with the loss of Shimadzu prophage

Full genome sequencing of the three *R. intestinalis* clones with an additional spacer also revealed their loss of the Shimadzu prophage. Amplification by PCR of either *attB*, the site formed by prophage excision, or *attL*, a site formed by prophage presence (Fig. 4D), revealed the absence of Shimadzu in all 32 clones with a new spacer (Fig. 4E). To evaluate the representativeness of the isolated clones, we determined by quantitative PCR the evolution with time of the global proportion of faecal bacteria with intact *attB* site. This proportion remained close to that estimated by screening of individual clone, 5.10^{-3}, for the first two weeks of colonization (Fig. 4H). Yet, after phage bloom, the proportion of bacteria with an intact *attB* site dramatically increased in all tested mice: Shimadzu-derived spacer acquisition most probably selected cured bacteria because of the important cost associated with autoimmunity.

Resistance is also mediated by a prophage with a modified immunity region

By contrast, the fourth sequenced clone (clone 4 in Fig. 4E), had no additional spacer but a second Shimadzu prophage, inserted in a tRNA at position 3,542,165 of the genome. Sequence analysis revealed that some prophage positions were mutated in 50% of aligning reads, suggesting that the second prophage, named Shim4 (for Shimadzu prophage with modified Immunity region located in clone 4), was mutated at these positions (Fig. 3A). In particular, two different alleles of both VRs could be identified, and the VR_1 mutant allele was identical to that of the ultravirulent Shivir phages isolated from the same experiment (Table S1). No mutation in *orf26* was identified, reinforcing the idea that such mutations are not essential for infection *in vivo*. In the immunity region, the same three point mutations observed in corresponding virome 2 were detected, but the *cro* gene and its promoter were inverted (Fig. 5A). This inversion might lead to an increased expression of CI repressor, enabling to restore the resistance to superinfection of the lysogen. To investigate the generality of this phenomenon, we looked for the presence of such Shim prophage with an inversion of *cro*, in the eight other clones resistant to Shivir5 infection but that had not acquired an additional spacer. We could amplify by PCR the Shimadzu immunity region in all isolates, even in clones with an intact *attB* sites, indicating the presence of a prophage at another genomic position than in the ancestral L1-82. An inversion of *cro* gene was evidenced in all eight isolates (Fig. 4E). These bacterial isolates still produced virions, albeit at a level much lower than the

ancestral L1-82 strain or other lysogens isolated from mice, comforting the hypothesis that they produce more CI repressor in the prophage state than the L1-82 (Fig. 5B and S4).

Altogether, ultravirulent phage resistance of the 40 clones tested can be explained either by spacer acquisition or by lysogenization by a phage with a new immunity type. For clarity, a summary sketch of the results is proposed in Fig. 6.

Discussion

The prophage state is generally considered a dormant state that gives little opportunity for viral evolution, since viruses mutate much faster during lytic replication than during lysogenic cycles (Sanjuan et al. 2010). This might explain why the coevolution between temperate phages and bacteria have been much less studied than between virulent phages and bacteria (Koskella and Brockhurst 2014, Scanlan 2017). Yet, theory predicts that ultravirulent mutants with mutations in the operator that diminishes the binding of the repressor should invade the population when transmission is high, i.e. when the benefit of superinfection outweighs the cost of virulence (Berngruber et al. 2010, Berngruber et al. 2013). Here, we followed the coevolution between *R. intestinalis* L1-82 and its prophages for 33 days in the mouse gut, and demonstrate that Shimadzu ultravirulent mutants systematically invade, leading to major phage-mediated mortality of *R. intestinalis*. In addition, in agreement with theory (Berngruber et al. 2010), we observed in some cases later invasion of compensatory mutations that restore lysogenization ability and provide resistance to the ultravirulent mutants. Since the large majority of bacteria of the human microbiota harbours prophages (Touchon et al. 2016, Kim 2018), this kind of dynamics might regularly occur, leading to massive phage-induced bacterial mortality, and participating in the variations in microbiota composition.

Even though predicted by theory, evolution of ultravirulence was considered to only rarely arise in nature, and, to the best of our knowledge, had only been observed in industrial setups, where large bacterial population size favours the appearance of mutants (Shimizu-Kadota et al. 2000, Capra et al. 2011, Mercanti et al. 2011). Based on results with lambda phage, the best-known example of lysogeny regulation, this rarity was proposed to stem from the fact that a double mutant would need to be generated, and the probability of acquiring two

13

simultaneous mutations is very small during lysogeny (Lenski and May 1994, Koch 2007). Because of this presumed rarity, the evolution of ultravirulence is rarely, if ever, considered when referring to lysogeny in host-associated microbiota. Similarly to what was shown with lambda, all ultravirulent Shimadzu mutants had at least two mutations in the immunity region, in operators O_R1 and O_R3. Yet, when a single strain dominates at high concentration, as in gnotoxenic mice, our results show that ultravirulent mutants with several simultaneously acquired mutations can be generated. Single-strain stability, i.e. the presence of a single dominant strain, was observed for numerous species of the gut microbiota (Truong et al. 2017), and in particular for *Roseburia* (Moss et al. 2017), suggesting that the environmental conditions necessary for ultravirulence evolution should be present in natural gut ecosystems. Most probably, some specific phage genetic features are also required. In particular, whether the phage DGR system and/or the presence of spontaneously cured bacteria facilitated ultravirulence evolution is a very interesting open question that deserve further investigation in a future work.

Here phage infection resulted in the fast selection of resistant bacteria in the mouse gut, contrary to what is generally observed *in vivo* (Chibani-Chennoufi et al. 2004, Maura et al. 2012, Reyes et al. 2013, Nale et al. 2016). This absence of genetic resistance selection was interpreted as a result of physiological bacterial resistance to phage infection in the mouse intestine, resulting from its complex structure and slow bacterial growth (Miedzybrodzki et al. 2012, Brussow 2013, Berngruber et al. 2015, Maslov and Sneppen 2015). In our experiments, almost all bacterial clones isolated after the emergence of ultravirulent phages were resistant, indicating very efficient phage-mediated selection and thereby particularly efficient multiplication of Shimadzu in the mouse gastrointestinal tract. To the best of our knowledge, our work constitutes the first examination of phage-bacteria interactions with a dominant member of the gut microbiota, and such outcome might be not so rare with this kind of bacteria. Ultravirulent prophage mutants might be responsible from some of the unexplained variations in bacterial populations in animal experiments.

Despite this very efficient phage infection, bacterial population did not drop so much, and this is due to rapidly acquired resistance by CRISPR-Cas immunity. CRISPR-Cas systems are widely distributed in prokaryotes, but only a small proportion of these systems have been identified to be active in bacteria. Here we demonstrate very good efficiency of the minimal type II-C

CRISPR-Cas system of *R. intestinalis*. Considering that among strains acquiring spacers only those cured of Shimadzu were conserved, and that cured bacteria were about 10^7 per mice, we can estimate a rate of spacer acquisition above 10^{-7}, similar to that of the best active CRISPR-Cas systems (reviewed in (McGinn and Marraffini 2019)). This high efficiency might be driven by the presence of defective Shimadzu particles, suggested by the discrepancy between Shimadzu PFU and the number of free-phage determined by qPCR. Indeed, there is evidence that spacer acquisition occurs most efficiently during the infection by defective viral particles (Hynes et al. 2014). Yet, interestingly, phage-derived spacer acquisition did not lead to phage extinction. This could result from a high phage mutation rate, enabling to generate escape mutants, but also from the presence of lysogens with an additional Shim-like prophage, that continuously produce free phage.

Interestingly, because the bacteria harbour a prophage closely related to the invading phage, spacer acquisition is highly detrimental, and leads to a kind of suicide. In the case of ultravirulent phage mutants, CRISPR–Cas does not act as a true immunity system, but rather functions as an abortive infection system leading to infected cell death. Luckily for *R. intestinalis*, a subpopulation of spontaneously cured bacteria, around 5.10^{-3}, is sufficient for phage-derived spacer acquisition and recovery of the bacterial population. However, contrary to Shimadzu, most prophages are almost impossible to cure (our own experience). In some cases, this might be related to the presence of toxin-antitoxin genes, as in the Jekyll prophage of this study, that stabilize the presence of prophages through post-segregational killing (Lehnherr et al. 1993, Romero et al. 2009, Goeders and Van Melderen 2014, Niu et al. 2015). In consequence, such subpopulation of cured bacteria does not always exist, probably precluding type II CRISPR-Cas efficiency against superinfection by ultravirulent mutants.

In the present study, constitutive resistance to Shivir conferred by phage receptor mutations was not observed. This is surprising in the light of studies showing that phage receptor mutations tend to dominate, especially when phage populations are high (Westra et al., 2015). The fitness cost of phage receptor mutations, in the case of Shimadzu, may be even higher than the one associated to CRISPR-Cas activity. But an alternative explanation could be that such receptor mutants may also be systematically countered by Shimadzu escape mutants, thanks to the DGR-mediated high variability of its tail gene which possibly encodes the receptor binding protein.

15

Finally, our study illustrates how a prophage can be an important source of mortality in bacterial populations. Therefore, prophages could be compared to "genetic time bombs", meaning that lysogenic bacteria live under the constant threat of superinfection by an ultravirulent mutant. Since phages are known to alter competition among bacterial strains or species (Bohannan and Lenski 2000, Joo et al. 2006, Koskella et al. 2012, Li et al. 2017), such evolution could have major effects on the composition of an ecosystem. In particular, *Roseburia* species are very important in the human gut microbiota, notably due to its important butyrate production, a metabolite with critical role in the regulation of host immune responses (Rios-Covian et al. 2016). Phage-mediated killing of *Roseburia* bacteria might therefore participate to the unexplained variations in microbiota composition observed even in defined communities (Koenig et al. 2011, Flores et al. 2014, Thaiss et al. 2014, Voigt et al. 2015, Sarker et al. 2017), and, in certain conditions, lead to important changes and a rupture of intestinal homeostasis.

Material and methods

Bacterial cultures. *R. intestinalis* L1-82 (DSM 14610), was grown at 37°C in LYBHI medium (BHI medium (Merck) supplemented with 0.5% yeast extract (Difco), cellobiose at 0.5 mg/ml (Sigma), maltose at 0.5 mg/ml (Sigma) and cysteine at 0.5 mg/ml (Sigma) in an anaerobic chamber filled with 90% N_2, 5% CO_2 and 5% H_2. *Escherichia coli* LF82 was grown in LB medium (Difco) at 37°C. *E. coli* was grown on LB agar plates (1.5% agar). *R. intestinalis* was plated on LYBHI agar plates (LYBHI medium with 1.5% agar and hemin at 0.5 %), supplemented with 2µg/mL of ciprofloxacine to suppress *E. coli* growth when necessary.

Sequencing of encapsidated DNA in L1-82 culture supernatant. 500 ml of a L1-82 culture grown overnight were centrifuged at 5,200 g for 15 min at 4 °C. Supernatant was recovered and centrifuged at 5,200g for 30 min at 4 °C. This step was repeated with 1 h of centrifugation and the samples were concentrated with PEG and NaCl as described below. Phage pellet was resuspended in 2 ml of SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris pH 7.5 (Sigma) and treated with 0.25 µg of RNAse A (Sigma) and Dnase I (Sigma) at 37 °C for 1 h. Phage DNA was then extracted using Promega kit Wizard® DNA Clean-up system. Sequencing was

performed with the Ion proton sequencing technology at the INRA Metagenopolis platform. Alignment of reads was performed using bowtie2 (-N 1 –L 32) and then visualized with Tablet using default parameters (Milne, 2013).

Mitomycin C induction. Overnight cultures were diluted 100-fold and grown at 37°C until absorbance at 600nm was between 0.2 and 0.4. 135 µL of culture was distributed in a 96-well plate (Cellstar, 96 Well Cell Culture Plate, U-bottom) and 15µL of H_2O or Mitomycin C (Sigma)) was added at the appropriate dilution. Plate were sealed (AMPLIseal™ Greiner Bio-one) in anaerobic condition and incubated at 37°C in a TECAN SUNRISE. Absorbance at 600nm was measured every 15 min for 10 hours. Phage genome counts in supernatant were determined 6 hours after induction by quantitative PCR as described below.

Animals and experimental design. Germfree 5 to 8 week-old C3H/HeN mice (female) from the germfree rodent breeding facilities of Anaxem-Micalis (INRA, Jouy-en-Josas, France) were kept in flexible-film isolators (Getinge-La Calhène, Vendôme, France) in standard cages (2 to 4 mice/cage) with sterile wood shavings as animal bedding. Mice were given free access to autoclaved tap water and to a standard diet, R03-40 (Scientific Animal food and Engineering, Augy, France), sterilized by gamma irradiation. Isolators were maintained under controlled conditions of light (12h), temperature (20 to 22°C) and humidity (45 to 55%). To obtain gnotobiotic *E. coli*/*R. intestinalis*-diassociated mice, mice were orally inoculated with 100µL of a saturated *E. coli* LF82 culture ($2.5x10^9$ CFU). 48 hours later, 10^{10} CFU of *R. intestinalis* in 200µL were administered by intragastric gavage of mice pretreated with sodium bicarbonate (0.2 M, 0.1mL by intragastric gavage, 10 min of wait before inoculation of bacteria). All procedures were carried out according to European Community Rules of Animal Care and with authorization 1234-2015101315238694 from French Veterinary Services.

Phage extraction and isolation from faeces. Fresh or frozen faeces thawed on ice for 10 min were diluted 40-fold in cold PBS. Resuspended faeces were kept on ice for 5 min with regular agitation prior to centrifugation for 10 min at 5,251 g at 4°C. Supernatants were recovered and filtered through a 0.22 µm filter (PALL Corporation Acrodisc PF syringe filter). For subsequent quantification by PCR, virions were precipitated with PEG 8000 (Sigma) at a final concentration of 10% and NaCl (Sigma) at a final concentration of 1M. After overnight incubation at 4°C, phage particles were harvested by centrifugation (5,251 g for 1 hour at 4°C

17

with a swinging rotor) and resuspended in 100 µl of SM buffer (100 mM NaCl, 8 mM MgSO4, 50 mM Tris pH 7.5 (Sigma)). Samples were treated with 10 U of Turbo DNase (Ambion) for 1 hour at 37 °C in Turbo DNase buffer and then incubated at 95°C for 30 min in 0.2 ml PCR tubes for enzyme inactivation and capsid disruption. For phage isolation, the fecal filtrates were mixed with 150 µL of an exponentially growing *R. intestinalis* culture (Optical density at 600nm comprised between 0.2 and 0.5) in an anaerobic chamber. 1 ml of melted Top BHI agarose (LYBHI supplemented with 0.25% agarose and 0.1% cysteine (Sigma)) at 37°C was added and the mix was immediately poured on 6 cm Petri Dishes containing LYBHI agar. Plates were then incubated overnight at 37°C in the anaerobic chamber. A single lysis plaque was then mixed with an exponentially growing bacterial culture and top agar containing and poured on an LYBHI Agar plate as described above. After 16 hours of incubation at 37°C, lysis was confluent, and 2 ml of SM buffer were poured on the top of the plates and incubated at 4°C for two hours. The phage supernatant was then recovered and filtered at 0.2 µm.

Bacterial DNA extraction from faeces. Bacterial pellets obtained during free phage extraction from faeces were resuspended in 500 µL of LB (Difco). The suspension was centrifuged 1 min at 500 g to remove debris. Supernatant was recovered and mixed with 250 µL of lysis buffer (200mM NaCl/20mM EDTA, 5% SDS), 250 µL of phenol:choloform:isoamyl alcohol (25:24:1; [pH 8.0], Sigma aldrich) and half of a tube of silica beads (100µM MP Bio, Lysing matrix B). Bacteria were lysed using Fast-prep MPBio (5.5, 3x30s, 5 min between each cycle). Samples were then centrifuged (13,000 g, 3min, 20°c) and the aqueous phase recovered. After addition of 400 µL of chloroform:isoamyl alcohol (24:1), the mix was vortexed vigorously (10s) and centrifuged (13,000 g, 3min, 20°C). DNA was precipitated with 2 volumes of ethanol and 0.3M of sodium acetate and resuspended in 100 µL of water.

Phage and bacteria quantification by qPCR. qPCR on 100-fold diluted samples was performed using the Takyon ROX SYBR Mastermix blue dTTP kit (Eurogentec) and the StepOnePlus real time PCR system (Applied Biosystem). Phage and bacteria amounts were determined using specific primer pairs (Table S3). Numbers of copies in standards were calculated using DNA quantification (Qbit) and the strain genome size. The reaction mix was the following: 7.5 µl Takyon mix, 0.9 µl H_2O, 0.3 µl of each primer (200 nM final concentration), 6 µl of DNA diluted in H_2O. The PCR conditions were: (95°C 15s, 58°C 45s, 72°C 30s) 45 cycles, 72°C 5 min, followed by melting curves. Results were analysed using the StepOne Software 2.3.

Isolation of *R. intestinalis* clones from mouse faeces. One or two faeces were recovered from mice, sealed immediately in screw cap micro tube, transferred to an anaerobic chamber (see bacterial culture), diluted in 0.5 mL of PBS 1X. 20µL of this bacterial suspension was then streaked on GHYBHI plates (brain-heart infusion medium supplemented with 0.5% yeast extract, 1.5% agar, cellobiose (1 mg/ml [Sigma]), maltose (1 mg/ml [Sigma]), cysteine (0.5 mg/ml [Sigma]), Hemin 0,5% and ciprofloxacin at 2 µg/mL). After 24 to 48h of growth at 37°C, isolated clones were grown 24h at 37°C in liquid LY-BHI medium, frozen with 20% of glycerol and kept at -80°C until analysis.

Quantification of Shimadzu cured bacteria. In an anaerobic chamber, bacteria either from a growing culture (OD600 = 0.3) or from a frozen faeces were rapidly diluted in cold PBS to prevent reinfection and plated on LYBHI agar plates. Individual clones were used to inoculate two 96 well microplates per condition. After a 16 hours incubation at 37°C, PCR with either attB or attL primers (table S3) was realized on 1 µl of bacterial cultures, and the proportion of attB positive clones determined. For quantification by PCR, quantitative PCR was performed on bacterial DNA as described above with attB and sigA primers, and the proportion of cured bacteria determined as the attB over sigA copy numbers.

Shimadzu susceptibility assay in 96 wells microplates. A 96 well microplate (Cellstar, 96 Well Cell Culture Plate, U-bottom) was inoculated with 96 different clones and duplicated for conservation at -80°C and subsequent tests. For each test, a microplate was used to inoculate a microplate filled with 200 µl of LYBHI. Overnight cultures were diluted 50-fold and grown at 37°C for two hours in 50 µl. 5 µl of a phage lysate at a concentration of 10^9 pfu/ml was added and the microplate incubated for 30 mn at 37°C. Cultures were then diluted 4-fold in warm LYBHI and incubated at 37°C for 16 hours. Absorbance at 600nm was then measured in a TECAN SUNRISE, and compared to the absorbance of cultures grown in similar conditions but with no phage. Clones were considered susceptible to the phage if the absorbance at 600nm of the culture with phage was more than 2-fold lower than that of the culture without phage. The experiment was repeated two times for all clones and a third time for clones that gave incongruent results.

Shimadzu susceptibility assay on agar plates. A 5 µl drop of phage suspension containing ~10^5 PFU was spotted on a lawn of *R. intestinalis* in top agarose as described above, and incubated

19

at 37°C for 16 hours. Confluent lysis at the position of the drop indicated bacterial susceptibility, whereas 10 or less PFU indicated bacterial resistance.

Bacterial numeration by microscopy. 10-fold dilutions of freshly passed faeces in PBS were spread on an agarose pad on a microscope slide and imaged at 100x magnification by a microscope (Zeiss Axioskop2Plus; camera Zeiss axiocam ICc1), in phase contrast. *E. coli* and *R. intestinalis* cells were manually counted (Fig. S3). We analysed more than 50 cells on 3 different images for each condition. In parallel, *E. coli* CFU were numerated on LB agar plates. The concentration of *R. intestinalis* cell was obtained by multiplying the concentration of *E. coli* deduced from plating by the ratio of *R. intestinalis* on *E. coli* cells obtained by microscopy.

Transmission electronic microscopy (TEM). For Jekyll virion observation, 100 mL of *R. intestinalis* culture supplemented with 1µg/mL of mitomycin C were centrifuged and the supernatant filtered through a 0.2 µm filter, concentrated with PEG and 0.5M NaCl as described above and then purified by ultracentrifugation (20,000g, 2 h, 4 °C) on a 3 layers iodoxanol gradient (concentration of 0, 20% and 45% of iodoxanol). Virions were recovered at the interface between the 20% and the 45% layers. Shimadzu virions were extracted from mice faeces resuspended in PBS. After centrifugation for 2 min at 10,000g, the supernatant was filtered and then concentrated using an ultrafiltration device Amicon Ultra 2 mL with a cut-off of 100 000 Da (Millipore)**.** For both phages, droplets of phage preparations were directly placed on Formvar carbon-coated grids for 5 min. The grids were stained with 1% uranyl acetate and then viewed for TEM using a HITACHI HT 7700 (Elexience, France) at 80 kV. Microphotographies were acquired with a charge-coupled device camera AMT.

***R. intestinalis* L1-82 sequencing, assembly and annotation.** Genomic DNA was extracted from a bacterial culture with a standard phenol:chloroform protocol preceded by a lysis step with lysozyme and SDS, as previously described (Fouet et al., 1990). Sequencing was realised by the Genotoul GIS, using the PacBio technology. Long reads were quality controlled whith NanoPlot v1.8.1 (De Coster et al. 2018). Sequencing produced 202,572 reads of N50 12,694 nucleotides for a total run of 1,214,051,359 bases. This corresponds to a sequencing depth of 270x. Assembly was performed using Canu 1.6 (Koren et al. 2017) with default parameters except genomeSize=4.5m and pacbio=raw. Two contigs were produced by the assemblers, one corresponding to the chromosome, the other of 21,201 nucleotides with a GC% of 0.01%. The

second was considered spurious and removed from the analysis. A run of Illumina paired-end sequencing (5,827,730 paired-reads of 151 bp) was conducted to correct the remaining error. Reads were first mapped on the PacBio assembly using BWA 0.7.12-r1039 (Li and Durbin 2009), then the assembly was corrected using Pilon 1.22 (Walker et al. 2014). 290 corrections of the initial assembly were performed, mostly indels correction. A final single circular contig of 4 493 348 nucleotides was obtained, and its first nt was chosen at the shift of the genome GC skew, near the *dnaA* gene. The genome was then annotated with Prokka (Seemann 2014), and prophage genes and boundaries were manually edited. The new version of the genome has been deposited on the EBI database under accession number ERZ773328. On Shimadzu phage, promoters of *cI* and *cro* genes were searched with DBTBS (http://dbtbs.hgc.jp/).

Detection of mutations in the sequenced bacterial isolates. Research of SNPs in bacterial genomes was made using the reference L1-82 genome, using reads generated by Illumina sequencing by GATC-Eurofin and analysed using the PATRIC web interface (https://patricbrc.org) (Wattam et al. 2017). Mapping was made using BWA-MEM and SNP were searched with freebayes with default parameters. Prophage insertion in clone 4 was first evidenced by the twice higher read coverage on the region corresponding to the Shimadzu prophage, following reads alignment using bowtie2 (-N 1 –L 32) and visualization with Tablet using default parameters (Milne 2013). The position of the second prophage was determined by recovering, at each prophage border, the sequence of 10 clipped reads. These "hybrid" reads were then aligned with BLASTn to the L1-82 genome, revealing the insertion site. Spacer insertion was detected as in other clones, as described below.

Shivirs and virome sequencing. 5 ml of ultravirulent phage stocks (concentration above 10^9 PFU/ml) were concentrated with 10% PEG8000 and 0.5M NaCl. Alternatively, for virome sequencing, virions were extracted from two to three mouse faeces as described above for qPCR. In both cases, viral DNA was then extracted with a standard phenol:chloroform protocol. DNA was sequenced either by GATC-Eurofins with the Illumina technology (2x125 nt), or, in the case of virome 1, with IonTorrent (300 nt reads) in the INRA metagenopolis unit.

Shivir mutation analysis. Raw reads were quality filtered using Trimmomatic (Bolger et al., 2014) in order to remove paired reads with bad qualities (with length < 100 bp, with remnant of Illumina adapters and with average quality values below Q20 on a sliding window of 4 nt).

The filtering process yielded an average of 1 to 3 million paired-end reads (2 × 150 bp) per sample. Point mutations and small indel were detected by alignment of reads using bowtie2 (-N 1 –L 32) and visualized with Tablet using default parameters (Milne 2013). Since Shivir were multiplied on L1-82 strain that carry the ancestral Shimadzu prophage, ~20% of reads corresponded to the wild-type Shimadzu sequence. Putative larger deletions or insertions were searched by assembling the filtered reads using SPAdes (Bankevich et al., 2012) with the meta option and increasing kmer values (-k 21,33,55,77,99,127). The resulting contig was then compared by BLASTn to the ancestral Shimadzu.

Analysis of viromes. A first search of SNP was performed similarly to the Shivir analysis. The more thorough analysis of the VR regions was performed as follow. First, using Tablet, all reads covering the VR region of interest and 20 extra nt on each side (for 300 nt-long reads, virome 1), or at least one side (for 125 nt-long reads, virome 2) were downloaded. We reasoned that the various VR alleles in a given virome should group together in a phylogenetic tree. These reads were therefore used for a multi-alignment with MUSCLE (Edgar 2004), which was then trimmed with Gblocks, and given to PhyML on the LIRRM interface (Dereeper et al. 2008) to build a tree (default parameters, when more than 500 reads were present, two subpopulations were made and treated in parallel). Several clear flat branches per tree were obtained, corresponding to the main alleles of a given VR. The number of leafs of each such cluster was counted, and by substraction, the remaining reads were considered as singleton alleles. A read of each main cluster was then used to determine the amino-acid sequence of each VR allele. To determine allele proportions, a mapping of virome reads on each VR allele was performed with bowtie2, allowing 0 mismatch (-N 0 –L 20).

Detection of spacer acquisition. Spacer acquisition was determined by a PCR test (primers in table S3). PCR products were observed by electrophoresis on a 2% agarose gel.

Detection of mutations by Sanger sequencing. Sanger sequencing of the *R. intestinalis sbpE* and CRISPR loci was performed by GATC-Eurofin on DNA substrates generated by PCR amplification of bacterial DNA purified from mouse faeces with primers in table S3, using standard procedures. Sequences of the Shimadzu VR_1 and immunity region were similarly obtained from PCR amplification products of viral faecal DNA. Manual inspection of

22

sequencing traces allowed to detect peak mixtures, indicative of the presence of mutants in the population.

Author contributions. JKC, EM and MDP performed the experiments. VL assembled and annotated the *R. I.* L1-82 genome. JKC, MAP and MDP performed the mutation analyses. AM, HS and EM developed protocols. JKC and MDP analyzed the results and wrote the manuscript, and all authors revised and approved the manuscript.

References

Arumugam, M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, *et al.* (2011). "Enterotypes of the human gut microbiome." Nature **473**(7346): 174-180.

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, *et al.* (2008). "The RAST Server: rapid annotations using subsystems technology." BMC Genomics **9**: 75.

Backhed, F., H. Ding, T. Wang, L. V. Hooper, G. Y. Koh, A. Nagy, C. F. Semenkovich, *et al.* (2004). "The gut microbiota as an environmental factor that regulates fat storage." Proc Natl Acad Sci U S A **101**(44): 15718-15723.

Berngruber, T. W., R. Froissart, M. Choisy and S. Gandon (2013). "Evolution of virulence in emerging epidemics." PLoS Pathog **9**(3): e1003209.

Berngruber, T. W., S. Lion and S. Gandon (2015). "Spatial structure, transmission modes and the evolution of viral exploitation strategies." PLoS Pathog **11**(4): e1004810.

Berngruber, T. W., F. J. Weissing and S. Gandon (2010). "Inhibition of superinfection and the evolution of viral latency." J Virol **84**(19): 10200-10208.

Bobay, L. M., M. Touchon and E. P. Rocha (2014). "Pervasive domestication of defective prophages by bacteria." Proc Natl Acad Sci U S A **111**(33): 12127-12132.

Bohannan, B. J. M. and R. E. Lenski (2000). "The Relative Importance of Competition and Predation Varies with Productivity in a Model Community." Am Nat **156**(4): 329-340.

Bondy-Denomy, J. and A. R. Davidson (2014). "When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness." J Microbiol **52**(3): 235-242.

Bossi, L., J. A. Fuentes, G. Mora and N. Figueroa-Bossi (2003). "Prophage contribution to bacterial population dynamics." J Bacteriol **185**(21): 6467-6471.

Broussard, G. W., L. M. Oldfield, V. M. Villanueva, B. L. Lunt, E. E. Shine and G. F. Hatfull (2013). "Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches." Mol Cell **49**(2): 237-248.

Brown, S. P., L. Le Chat, M. De Paepe and F. Taddei (2006). "Ecology of microbial invasions: amplification allows virus carriers to invade more rapidly when rare." Curr Biol **16**(20): 2048-2052.

Brum, J. R., B. L. Hurwitz, O. Schofield, H. W. Ducklow and M. B. Sullivan (2016). "Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics." ISME J **10**(2): 437-449.

Brussow, H. (2013). "Bacteriophage-host interaction: from splendid isolation into a messy reality." Curr Opin Microbiol **16**(4): 500-506.

Capra, M. L., D. J. Mercanti, L. C. Rossetti, J. A. Reinheimer and A. Quiberoni (2011). "Isolation and phenotypic characterization of Lactobacillus casei and Lactobacillus paracasei bacteriophage-resistant mutants." J Appl Microbiol **111**(2): 371-381.

Chibani-Chennoufi, S., J. Sidoti, A. Bruttin, E. Kutter, S. Sarker and H. Brussow (2004). "In vitro and in vivo bacteriolytic activities of Escherichia coli phages: implications for phage therapy." Antimicrob Agents Chemother **48**(7): 2558-2569.

Consortium, T. H. M. P. (2013). "Structure, function and diversity of the healthy human microbiome." Nature **486**(7402): 207-214.

Cornuault, J. K., M. A. Petit, M. Mariadassou, L. Benevides, E. Moncaut, P. Langella, H. Sokol, *et al.* (2018). "Phages infecting Faecalibacterium prausnitzii belong to novel viral genera that help to decipher intestinal viromes." Microbiome **6**(1): 65.

De Coster, W., S. D'Hert, D. T. Schultz, M. Cruts and C. Van Broeckhoven (2018). "NanoPack: visualizing and processing long read sequencing data." Bioinformatics.

De Paepe, M., M. Leclerc, C. R. Tinsley and M. A. Petit (2014). "Bacteriophages: an underestimated role in human and animal health?" Front Cell Infect Microbiol **4**: 39.

De Paepe, M., L. Tournier, E. Moncaut, O. Son, P. Langella and M. A. Petit (2016). "Carriage of lambda Latent Virus Is Costly for Its Bacterial Host due to Frequent Reactivation in Monoxenic Mouse Intestine." PLoS Genet **12**(2): e1005861.

Delattre, H., O. Souiai, K. Fagoonee, R. Guerois and M. A. Petit (2016). "Phagonaute: A web-based interface for phage synteny browsing and protein function prediction." Virology **496**: 42-50.

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard*, et al.* (2008). "Phylogeny.fr: robust phylogenetic analysis for the non-specialist." Nucleic Acids Res **36**(Web Server issue): W465-469.

Duerkop, B. A., C. V. Clements, D. Rollins, J. L. Rodrigues and L. V. Hooper (2012). "A composite bacteriophage alters colonization by an intestinal commensal bacterium." Proc Natl Acad Sci U S A **109**(43): 17621-17626.

Durmaz, E., M. J. Miller, M. A. Azcarate-Peril, S. P. Toon and T. R. Klaenhammer (2008). "Genome sequence and characteristics of Lrm1, a prophage from industrial Lactobacillus rhamnosus strain M1." Appl Environ Microbiol **74**(15): 4601-4609.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res **32**(5): 1792-1797.

Edwards, R. A., K. McNair, K. Faust, J. Raes and B. E. Dutilh (2016). "Computational approaches to predict bacteriophage-host relationships." FEMS Microbiol Rev **40**(2): 258-272.

Falony, G., M. Joossens, S. Vieira-Silva, J. Wang, Y. Darzi, K. Faust, A. Kurilshikov*, et al.* (2016). "Population-level analysis of gut microbiome variation." Science **352**(6285): 560-564.

Flores, G. E., J. G. Caporaso, J. B. Henley, J. R. Rideout, D. Domogala, J. Chase, J. W. Leff*, et al.* (2014). "Temporal variability is a personalized feature of the human microbiome." Genome Biol **15**(12): 531.

Fouet, A., S. F. Jin, G. Raffel and A. L. Sonenshein (1990). "Multiple regulatory sites in the Bacillus subtilis citB promoter region." J Bacteriol **172**(9): 5408-5415.

Goeders, N. and L. Van Melderen (2014). "Toxin-antitoxin systems as multilevel interaction systems." Toxins (Basel) **6**(1): 304-324.

Guo, H., D. Arambula, P. Ghosh and J. F. Miller (2014). "Diversity-generating Retroelements in Phage and Bacterial Genomes." Microbiol Spectr **2**(6).

Hynes, A. P., M. Villion and S. Moineau (2014). "Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages." Nat Commun **5**: 4399.

Joo, J., M. Gunny, M. Cases, P. Hudson, R. Albert and E. Harvill (2006). "Bacteriophage-mediated competition in Bordetella bacteria." Proc Biol Sci **273**(1595): 1843-1848.

Kim, M. S. B., J. W. (2018). "Lysogeny is prevalent and widely distributed in the murine gut microbiota." ISME J **12**(4): 1127-1141.

Knowles, B., C. B. Silveira, B. A. Bailey, K. Barott, V. A. Cantu, A. G. Cobian-Guemes, F. H. Coutinho*, et al.* (2016). "Lytic to temperate switching of viral communities." Nature **531**(7595): 466-470.

Koch, A. L. (2007). "Evolution of temperate pathogens: the bacteriophage/bacteria paradigm." Virol J **4**: 121.

Koenig, J. E., A. Spor, N. Scalfone, A. D. Fricker, J. Stombaugh, R. Knight, L. T. Angenent*, et al.* (2011). "Succession of microbial consortia in the developing infant gut microbiome." Proc Natl Acad Sci U S A **108 Suppl 1**: 4578-4585.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman and A. M. Phillippy (2017). "Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation." Genome Res **27**(5): 722-736.

Koskella, B. and M. A. Brockhurst (2014). "Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities." FEMS Microbiol Rev **38**(5): 916-931.

Koskella, B., D. M. Lin, A. Buckling and J. N. Thompson (2012). "The costs of evolving resistance in heterogeneous parasite environments." Proc Biol Sci **279**(1735): 1896-1903.

Larsen, N., F. K. Vogensen, F. W. van den Berg, D. S. Nielsen, A. S. Andreasen, B. K. Pedersen, W. A. Al-Soud*, et al.* (2010). "Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults." PLoS One **5**(2): e9085.

Lehnherr, H., E. Maguin, S. Jafri and M. B. Yarmolinsky (1993). "Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained." J Mol Biol **233**(3): 414-428.

Lenski, R. E. and R. M. May (1994). "The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses." J Theor Biol **169**(3): 253-265.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, X. Y., T. Lachnit, S. Fraune, T. C. G. Bosch, A. Traulsen and M. Sieber (2017). "Temperate phages as self-replicating weapons in bacterial competition." J R Soc Interface **14**(137).

Lopes, A., P. Tavares, M. A. Petit, R. Guerois and S. Zinn-Justin (2014). "Automated classification of tailed bacteriophages according to their neck organization." BMC Genomics **15**: 1027.

Lucchini, S., F. Desiere and H. Brussow (1999). "Comparative genomics of Streptococcus thermophilus phage species supports a modular evolution theory." J Virol **73**(10): 8647-8656.

Mahony, J., G. A. Lugli, D. van Sinderen and M. Ventura (2018). "Impact of gut-associated bifidobacteria and their phages on health: two sides of the same coin?" Appl Microbiol Biotechnol **102**(5): 2091-2099.

Manrique, P., M. Dills and M. J. Young (2017). "The Human Gut Phage Community and Its Implications for Health and Disease." Viruses **9**(6).

Marchesi, J. R., D. H. Adams, F. Fava, G. D. Hermes, G. M. Hirschfield, G. Hold, M. N. Quraishi*, et al.* (2016). "The gut microbiota and host health: a new clinical frontier." Gut **65**(2): 330-339.

Maslov, S. and K. Sneppen (2015). "Well-temperate phage: optimal bet-hedging against local environmental collapses." Sci Rep **5**: 10523.

Maura, D., M. Galtier, C. Le Bouguenec and L. Debarbieux (2012). "Virulent bacteriophages can target O104:H4 enteroaggregative Escherichia coli in the mouse intestine." Antimicrob Agents Chemother.

McGinn, J. and L. A. Marraffini (2019). "Molecular mechanisms of CRISPR–Cas spacer acquisition." Nature Reviews Microbiology **17**: 7-12.

Mercanti, D. J., D. Carminati, J. A. Reinheimer and A. Quiberoni (2011). "Widely distributed lysogeny in probiotic lactobacilli represents a potentially high risk for the fermentative dairy industry." Int J Food Microbiol **144**(3): 503-510.

Miedzybrodzki, R., J. Borysowski, B. Weber-Dabrowska, W. Fortuna, S. Letkiewicz, K. Szufnarowski, Z. Pawelczyk*, et al.* (2012). "Clinical aspects of phage therapy." Adv Virus Res **83**: 73-121.

Mills, S., F. Shanahan, C. Stanton, C. Hill, A. Coffey and R. P. Ross (2013). "Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota." Gut Microbes **4**(1): 4-16.

Milne, I. S., G. Bayer, M. Cock, PJA. Pritchard, L. Cardle, L. Shaw, PD. and Marshall, D. (2013). "Using Tablet for visual exploration of second-generation sequencing data." Briefings in Bioinformatics **14**(2): 193-202.

Minot, S., R. Sinha, J. Chen, H. Li, S. A. Keilbaugh, G. D. Wu, J. D. Lewis*, et al.* (2011). "The human gut virome: inter-individual variation and dynamic response to diet." Genome Res **21**(10): 1616-1625.

Moss, E. L., S. B. Falconer, E. Tkachenko, M. Wang, H. Systrom, J. Mahabamunuge, D. A. Relman*, et al.* (2017). "Long-term taxonomic and functional divergence from donor bacterial strains following fecal microbiota transplantation in immunocompromised patients." PLoS One **12**(8): e0182585.

Nale, J. Y., J. Spencer, K. R. Hargreaves, A. M. Buckley, P. Trzepinski, G. R. Douce and M. R. Clokie (2016). "Bacteriophage Combinations Significantly Reduce Clostridium difficile Growth In Vitro and Proliferation In Vivo." Antimicrob Agents Chemother **60**(2): 968-981.

Niu, Y. D., S. R. Cook, J. Wang, C. L. Klima, Y. H. Hsu, A. M. Kropinski, D. Turner*, et al.* (2015). "Comparative analysis of multiple inducible phages from Mannheimia haemolytica." BMC Microbiol **15**: 175.

Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen*, et al.* (2010). "A human gut microbial gene catalogue established by metagenomic sequencing." Nature **464**(7285): 59-65.

Reyes, A., M. Haynes, N. Hanson, F. E. Angly, A. C. Heath, F. Rohwer and J. I. Gordon (2010). "Viruses in the faecal microbiota of monozygotic twins and their mothers." Nature **466**(7304): 334-338.

Reyes, A., M. Wu, N. P. McNulty, F. L. Rohwer and J. I. Gordon (2013). "Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut." Proc Natl Acad Sci U S A **110**(50): 20236-20241.

26

Rios-Covian, D., P. Ruas-Madiedo, A. Margolles, M. Gueimonde, C. G. de Los Reyes-Gavilan and N. Salazar (2016). "Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health." Front Microbiol **7**: 185.

Romero, P., N. J. Croucher, N. L. Hiller, F. Z. Hu, G. D. Ehrlich, S. D. Bentley, E. Garcia*, et al.* (2009). "Comparative genomic analysis of ten Streptococcus pneumoniae temperate bacteriophages." J Bacteriol **191**(15): 4854-4862.

Rothschild, D., O. Weissbrod, E. Barkan, A. Kurilshikov, T. Korem, D. Zeevi, P. I. Costea*, et al.* (2018). "Environment dominates over host genetics in shaping human gut microbiota." Nature **555**(7695): 210-215.

Sanjuan, R., M. R. Nebot, N. Chirico, L. M. Mansky and R. Belshaw (2010). "Viral mutation rates." J Virol **84**(19): 9733-9748.

Sarker, S. A., B. Berger, Y. Deng, S. Kieser, F. Foata, D. Moine, P. Descombes*, et al.* (2017). "Oral application of Escherichia coli bacteriophage: safety tests in healthy and diarrheal children from Bangladesh." Environ Microbiol **19**(1): 237-250.

Scanlan, P. D. (2017). "Bacteria-Bacteriophage Coevolution in the Human Gut: Implications for Microbial Diversity and Functionality." Trends Microbiol **25**(8): 614-623.

Seemann, T. (2014). "Prokka: rapid prokaryotic genome annotation." Bioinformatics **30**(14): 2068-2069.

Selva, L., D. Viana, G. Regev-Yochay, K. Trzcinski, J. M. Corpa, I. Lasa, R. P. Novick*, et al.* (2009). "Killing niche competitors by remote-control bacteriophage induction." Proc Natl Acad Sci U S A **106**(4): 1234-1238.

Shimizu-Kadota, M., M. Kiwaki, S. Sawaki, Y. Shirasawa, H. Shibahara-Sone and T. Sako (2000). "Insertion of bacteriophage phiFSW into the chromosome of Lactobacillus casei strain Shirota (S-1): characterization of the attachment sites and the integrase gene." Gene **249**(1-2): 127-134.

Sierro, N., Y. Makita, M. de Hoon and K. Nakai (2008). "DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information." Nucleic Acids Res **36**(Database issue): D93-96.

Silveira, C. B. and F. L. Rohwer (2016). "Piggyback-the-Winner in host-associated microbial communities." NPJ Biofilms Microbiomes **2**: 16010.

Tamanai-Shacoori, Z., I. Smida, L. Bousarghin, O. Loreal, V. Meuric, S. B. Fong, M. Bonnaure-Mallet*, et al.* (2017). "Roseburia spp.: a marker of health?" Future Microbiol **12**: 157-170.

Thaiss, C. A., D. Zeevi, M. Levy, G. Zilberman-Schapira, J. Suez, A. C. Tengeler, L. Abramson*, et al.* (2014). "Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis." Cell **159**(3): 514-529.

Touchon, M., A. Bernheim and E. P. Rocha (2016). "Genetic and life-history traits associated with the distribution of prophages in bacteria." ISME J **10**(11): 2744-2754.

Touchon, M., A. Bernheim and E. P. Rocha (2016). "Genetic and life-history traits associated with the distribution of prophages in bacteria." ISME J.

Truong, D. T., A. Tett, E. Pasolli, C. Huttenhower and N. Segata (2017). "Microbial strain-level population structure and genetic diversity from metagenomes." Genome Res **27**(4): 626-638.

van Houte, S., A. Buckling and E. R. Westra (2016). "Evolutionary Ecology of Prokaryotic Immune Mechanisms." Microbiol Mol Biol Rev **80**(3): 745-763.

Voigt, A. Y., P. I. Costea, J. R. Kultima, S. S. Li, G. Zeller, S. Sunagawa and P. Bork (2015). "Temporal and technical variability of human gut metagenomes." Genome Biol **16**: 73.

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo*, et al.* (2014). "Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement." PLoS One **9**(11): e112963.

Waller, A. S., T. Yamada, D. M. Kristensen, J. R. Kultima, S. Sunagawa, E. V. Koonin and P. Bork (2014). "Classification and quantification of bacteriophage taxa in human gut metagenomes." ISME J **8**(7): 1391-1402.

Wattam, A. R., J. J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, *et al.* (2017). "Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center." Nucleic Acids Res **45**(D1): D535-D542.

Weiss, M., E. Denou, A. Bruttin, R. Serra-Moreno, M. L. Dillmann and H. Brussow (2009). "In vivo replication of T4 and T7 bacteriophages in germ-free mice colonized with Escherichia coli." Virology **393**(1): 16-23.

Willing, B. P., J. Dicksved, J. Halfvarson, A. F. Andersson, M. Lucio, Z. Zheng, G. Jarnerot, *et al.* (2010). "A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes." Gastroenterology **139**(6): 1844-1854 e1841.

Wu, L., M. Gingery, M. Abebe, D. Arambula, E. Czornyj, S. Handa, H. Khan, *et al.* (2018). "Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey." Nucleic Acids Res **46**(1): 11-24.

Zimmermann, L., A. Stephens, S. Z. Nam, D. Rau, J. Kubler, M. Lozajic, F. Gabler, *et al.* (2018). "A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core." J Mol Biol **430**(15): 2237-2243.

Supplementary table S1

			sequence from VR1 (amino acid)	%
Shimadzu wt		VR1	RCYWWSASVRSGGSTNCVFVNVSGHSSNWGASGGLYVPV	
		from TR	RCNWWTASVRSGNSTNCVNVNNNGNSNNWNASNELYVPC	
Experiment 1	Virome, pool cages 1&2	allele WT	RCYWWSASVRSGGSTNCVFVNVSGHSSNWGASGGLYVPV	27,70
		allele 2	RCNWWTASVRSGNSTICVHVDHSGHSYNWGASYELCVPV	26,90
		allele 3	RCNWWTASVGSGNSTVCVNVSINGYSDSWHASHELYVPV	15,10
		allele 4	RCGWWTASVRSGASTRCVYVGSYGNSNYWGASNELCVPV	4,88
	Phage isolate, cage 1	Shivir 1	RCYWWSASVRSGGSTNCVGVNNSGYSGIWLASGELYVPV	
Experiment 2	Virome A2, mice 2 cage 3	allele WT	RCYWWSASVRSGGSTNCVFVNVSGHSSNWGASGGLYVPV	39,62
		allele 8	RCNWWTASVRSGNSSNCVHVDYSGHSGNWDASHELCVPV	16,15
		allele 6	RCGWWTASVRSGNSTNCVDAASSGHSNHWGASHELYVPV	33,85
	Phage isolates, cages 3 & 4	Shivir2, cage 4	RCNWWTASVRSGNSSNCVHVDYSGHSGNWDASGELYVPV	
		Shivir3, cage 3	RCNWWTASVRSGNSSNCVHVDYSGHSGNWDASHELCVPV	
		Shivir4, cage 3	RCNWWTASVRSGNSSNCVHVDYSGHSGNWDASHELCVPV	
		Shivir5, cage 7	RCNWWTASVRSGNSSNCVHVDYSGHSGNWDASHELCVPV	
	Prophages, cage 3	Shimm1, clone4	RCNWWTASVRSGNSSNCVHVDYSGHSGNWAASHELCVPV	
		Shimm2, clone296	RCNWWTASVRSGNSSNCVHVDYSGHSGNWDASHELCVPV	
		Shimm3, clone297	RCNWWTASVRSGNSSNCVHVDYSGHSGNWDASHELCVPV	

			sequence from VR2 (amino acid)	%
Shimadzu wt		VR2 wt	SGGSTGCVSVSYHGHSYDWSASTELYVPV	
		from TR	SGNSTNCVNVNNNGNSNNWNASNELYVPV	
Experiment 1	Virome, pool cages 1&2	allele WT	SGGSTGCVSVSYHGHSYDWSASTELYVPV	61,46
		allele 2	SGSSTHCVYVDDNGHSGSWNASNELYVPV	1,67
		allele 3	SGNSTNCVRVTDNGHSNGWIASHELYVPV	3,13
		allele 4	SGGSANCVGVSDSGHSDHWNAGNELYVPV	3,33
	Phage isolate, cage 1	Shivir1	GGGSTRCVYVNGNGNSGDWGASDELCVPV	
Experiment 2	Virome A2, mice 2 cage 3	allele WT	SGGSTGCVSVSYHGHSYDWSASTELYVPV	52,15
		allele 8	SGNSTGCVYVGSNGYSDNWVAGSELCVPV	26,32
		allele 6	SGGSAYCVHVTNNGNSIHWGASGELYVPV	7,18
	Phage isolates, cages 3 & 4	Shivir2, cage 4	SGGSTGCVSVSYHGHSNHWGASGELYVPV	
		Shivir3, cage 3	SGTSTYCVTVNSYGYSDDWIASSELCVPV	
		Shivir4, cage 3	SGGSTGCVSVSYHGHSYDWSASGELYVPV	
		Shivir5, cage 7	SGGSTHCVAVSGDGNSNIWGASSELYVPV	
	Prophage cage 3	Shimm1, clone4	SGNSSNCVHVDYSGHSGNWDASHELCVPV	

Table S1. Amino acid sequences of the protein segment containing the DGR VRs, in viromes and in Shimadzu isolates. For the Shimadzu WT, two sequences are indicated: the first (VR1) corresponds to the actual protein, while the second correspond to the translation of the template repeat (TR). Color indicates amino acids modified relative to this TR sequence. For the viromes, the percentage (%) of each allele is indicated.
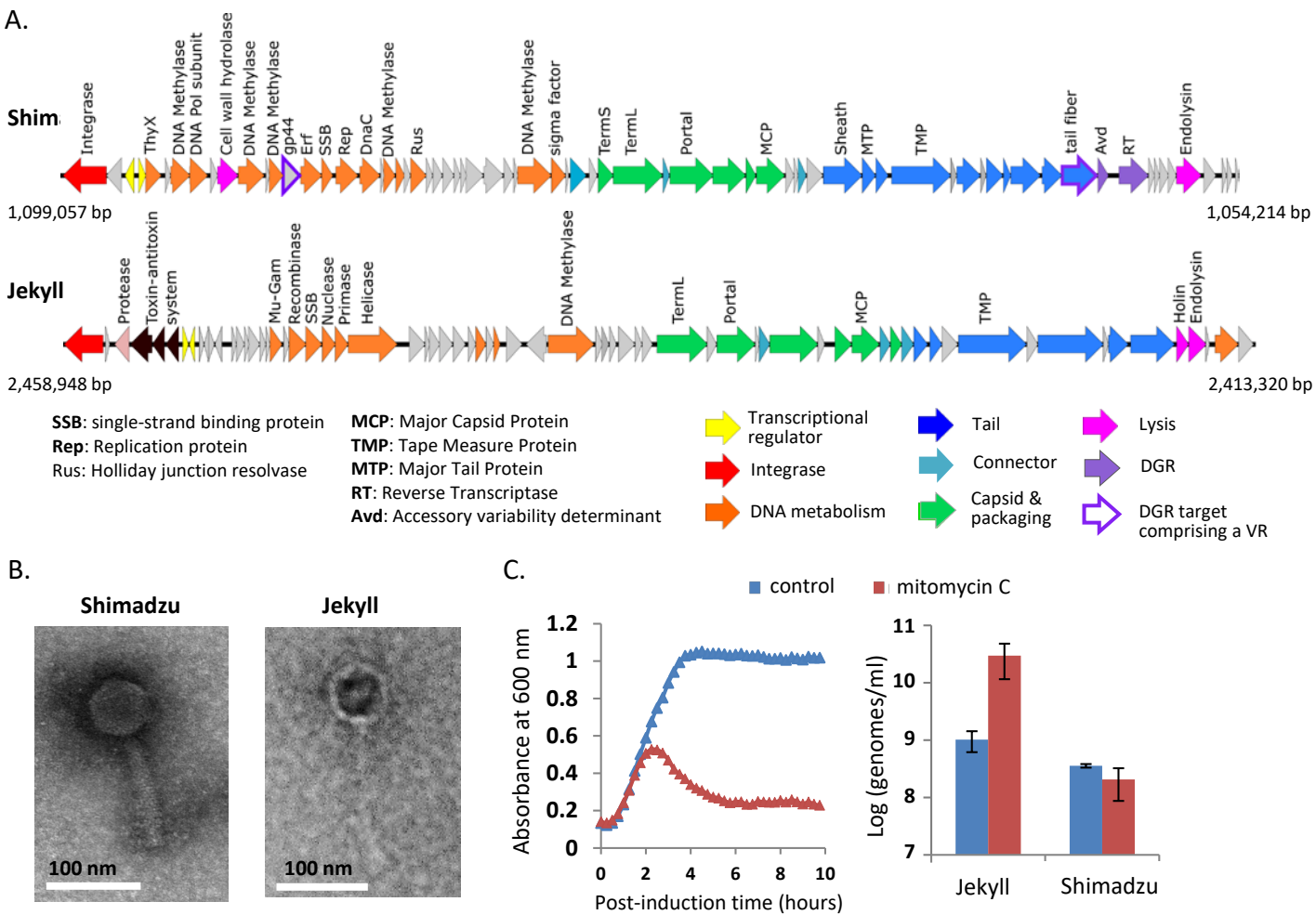
29

Table S2. Sequences and corresponding protospacer positions of acquired spacers

Clone	mouse	Sequence	Function targeted gene	Protospacer position
A1-3-D7	A1	GGAGGAGCTTCGCAATGTGATCGCCGAGGT	hyp prot with VR_2	33,275 - 33,304
A2-1-D7	A2	TAATTGATTGATAAATTTTCAACCATGCCA	Replication initiation protein	34,989 - 35,018
A2-2-D7	A2	CGAGACCGGCCTACTGGATCCAAACGGTTGC	Cytosine methyltransferase	32,200 - 32,230
A2-3-D7	A2	TAATTGATTGATAAATTTTCAACCATGCC	Replication initiation protein	34,989 - 35,018
C1-3-D7	C1	GGAGGAGCTTCGCAATGTGATCGCCGAGGT	hyp prot with VR_2	33,275 - 33,304
C1-5-D7	C1	TGACGGTCTGGCCCCTACCCTTACGACTTG	Intergenic region	32,506 - 32,535
		CCGCAGCGTGCAACTACTACCGCTCCCAGA	Resolvase	37,964 - 37,990
C1-1-D7	C1	TAATTGATTGATAAATTTTCAACCATGCCA	Replication initiation protein	34,983 - 35,010
C2-2-D7	C2	GACATTATGCAATGATTGTGATCTTGTCT	hyp prot	41,360 - 41,388
		CTTCCAAAAGACAGCCGCGTTATGCTGATTA	genome RI L1-82 (2,132,045-2,132,066)	
C2-3-D7	C2	GACATTATGCAATGATTGTGATCTTGTCTA	hyp prot	41,360 - 41,388
		CTTCCAAAAGACAGCCGCGTTATGCTGATTA	genome RI L1-82 (2,132,045-2,132,066)	
C2-4-D7	C2	GCATTCTAAGACTGAGCCACCACATCAGGC	intergenic region	44,722 - 44,751
C2-5-D7	C2	GCATTCTAAGACTGAGCCACCACATCAGGC	intergenic region	44,722 - 44,751
C3-4-D7	C3	CACTCTGATCCAGCAGCGTGAAGCCGAAA	gp16 (start integrase)	35,290 - 35,318

Table S3: Oligonucleotides used in this study.

	Targeted gene	Forward primer	Reverse primer
quantitative PCR	*R. intestinalis sigA*	GAAGATGGACGCGCAAGAAC	TGCTTGGATGACGCAGCTTA
	Shimadzu *tmp*	GACAGAGGCGACCACTATCG	GGAGACTGAGAGAGTCCCGT
	Jekyll *termL*	CCCAAAGATTGGCGAGGACT	GCGGCGACTCTCCATAATCA
	Shimadzu *attB*	GCGCTCCGTGAACATTCTATT	AAAGGTTCAGCTTTTGCGCC
Standard PCR and sequencing	Shimadzu Immunity region	GGCGCGATCTTTTTGCTCTT	GAGCTCCTCGAGCTTCTCAC
	attB shimadzu	CGGTTGGTGTTTCCTGTATATG	GTCTATTGTTCGAGCAATTC
	attP shimadzu	CGGTTGGTGTTTCCTGTATATG	CCCAAGTACAAACGCATACATAACC
	R. intestinalis sbpE	TCAGATGGCATTCCCGCAGAG	AAGGACAGCGATGACGGACAG
	Shimadzu VR_1	TCCAGTTACCCGGCTACACG	GCGCACCCTCATGGAATACAG
	CRISPR spacer	ATGAATCGTCAAGAGGCTCAAG	GGGATTTGTGCTCCCTATTATACC

Figure 1

A.

Shim...

Integrase, ThyX, DNA Methylase, DNA Pol subunit, Cell wall hydrolase, DNA Methylase, DNA Methylase, gp44, Erf, SSB, Rep, DnaC, DNA Methylase, Rus, DNA Methylase, sigma factor, TermS, TermL, Portal, MCP, Sheath, MTP, TMP, tail fiber, Avd, RT, Endolysin

1,099,057 bp   1,054,214 bp

Jekyll

Protease, Toxin-antitoxin system, Mu-Gam, Recombinase, SSB, Nuclease, Primase, Helicase, DNA Methylase, TermL, Portal, MCP, TMP, Holin, Endolysin

2,458,948 bp   2,413,320 bp

SSB: single-strand binding protein
Rep: Replication protein
Rus: Holliday junction resolvase

MCP: Major Capsid Protein
TMP: Tape Measure Protein
MTP: Major Tail Protein
RT: Reverse Transcriptase
Avd: Accessory variability determinant

Transcriptional regulator
Integrase
DNA metabolism

Tail
Connector
Capsid & packaging

Lysis
DGR
DGR target comprising a VR

B.

Shimadzu

Jekyll

100 nm

100 nm

C.

- control
- mitomycin C

Figure 1: *Roseburia intestinalis* L1-82 prophages. A) Genetic maps of Shimadzu and Jekyll prophages. Positions in *R. intestinalis* L1-82 genome are indicated. **B)** Transmission electron microscopy images of Shimadzu and Jekyll virions. **C)** Phage genome concentrations in bacterial cultures with or without mitomycin C (0.5 μg/ml). Mean ± standard deviation of three measures.
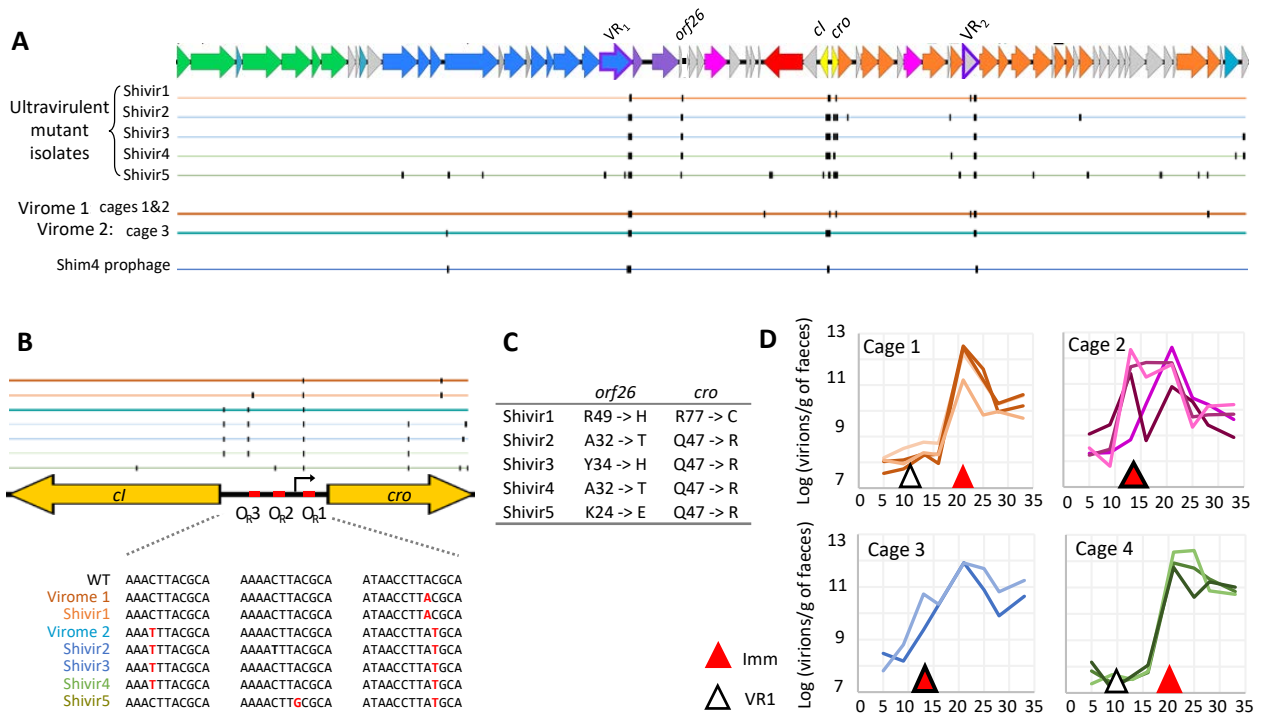
Figure 2
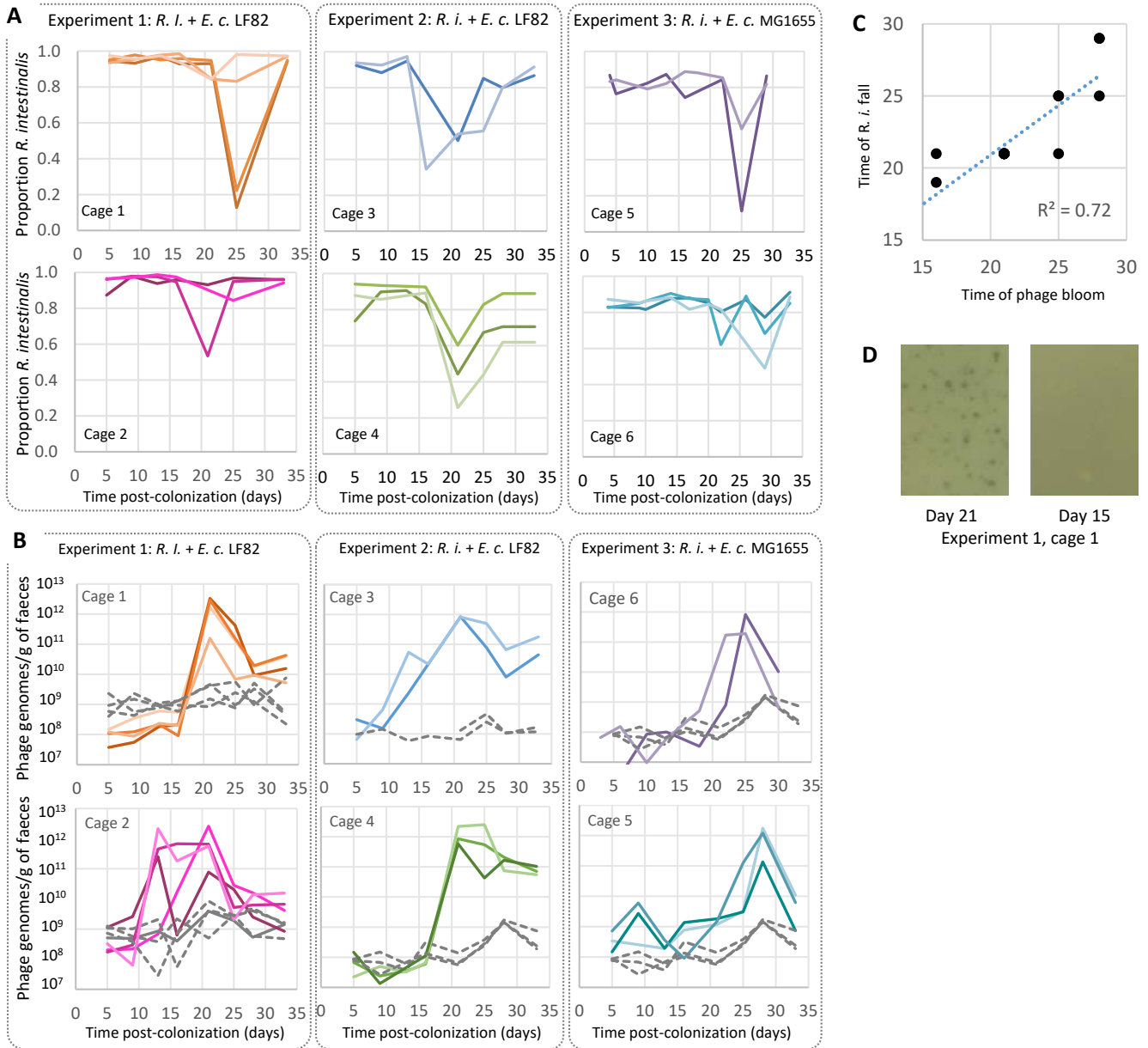
Figure 2: **Evolution of *R. intestinalis* and phage populations with time in mouse faeces**. Each panel corresponds to a different cage, hosting two to four mice. In each experiment, two cages (1 & 2, 3 & 4, or 5 &6) were placed in the same isolator. Mice from cages 1 to 4 were colonized with the LF82 strain of *E. coli*, whereas mice in cages 5 and 6 were colonized with the *E. coli* MG1655 strain. **A)** Evolution of the proportion of *R. intestinalis* over total bacteria. These proportions were estimated by quantitative PCR in experiments 1 and 2, and by microscopy in experiment 3. **B)** Evolution of the concentration of Shimadzu (solid line) and Jekyll (dashed lines) free-phage genomes. **C)** Time of *R. intestinalis* population collapse as a function of time of Shimadzu bloom. **D)** After 21 days of colonization, lysis plaques could be obtained from mouse faeces (left panel), but not from faeces collected at earlier time points (right panel).

Figure 3

Figure 3: Mutations of ultravirulent Shimadzu phages. A) Representation of mutation positions in virome 1 (all mice from experiment 1) and 2 (one mice from cage 3 in experiment 2), and in ultravirulent phage isolates. Color of the line indicates the cage of isolation. **B)** Upper panel: Mutation positions in the immunity region of Shimadzu, comprising the repressor genes *cl* and *cro* and the intergenic regions with three operators. Lower panel: mutations in the operators. **C)** Amino acid mutated in *orf26* and *cro* in ultravirulent isolates. **D)** Time of first detection of mutations in the operators and in the VR1 in total viral DNA from mouse faeces.

Figure 4
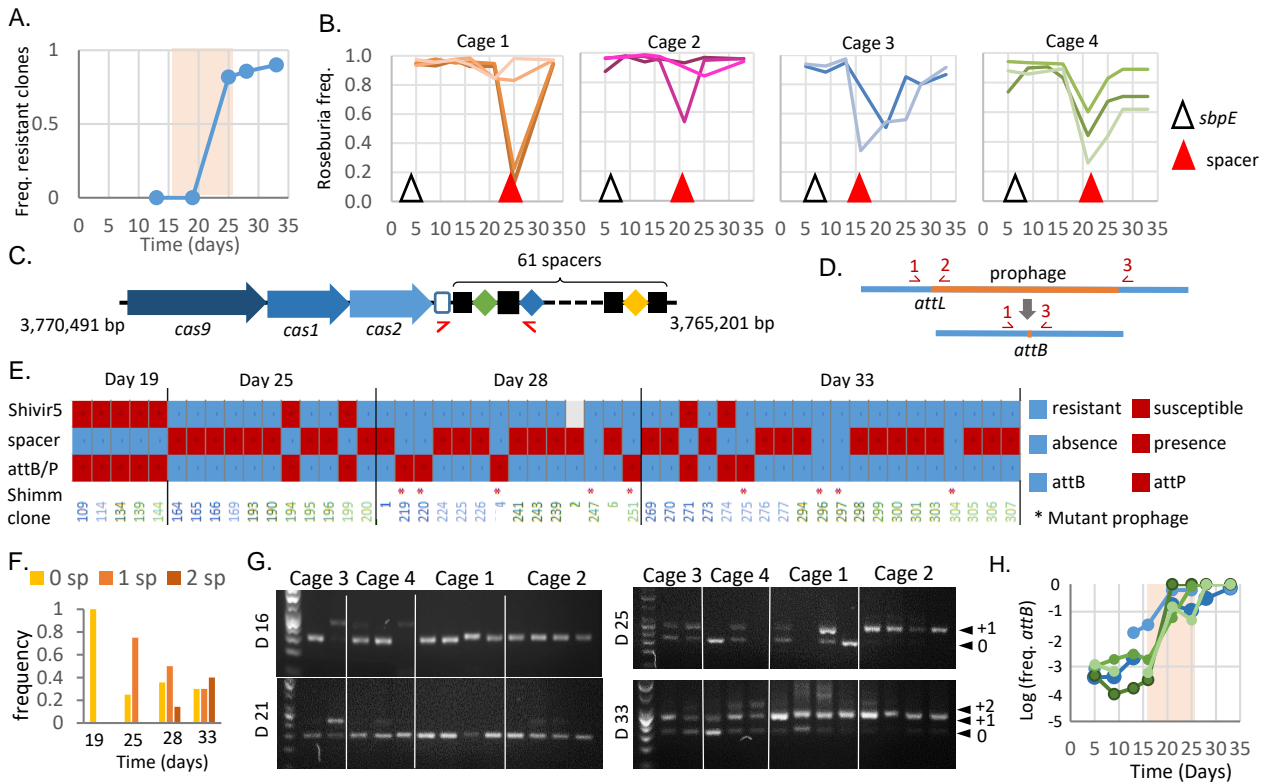
Figure 4: Phage resistance is associated to spacer acquisition or lysogenization by a prophage with a new immunity type. **A)** Proportion of bacterial clones resistant to Shimadzu with respect to isolation time from cages 3 & 4. The pale pink region indicates the Shimadzu bloom. **B)** Time of first detection of mutations in *sbpE* and spacer acquisition in total bacterial DNA from mouse faece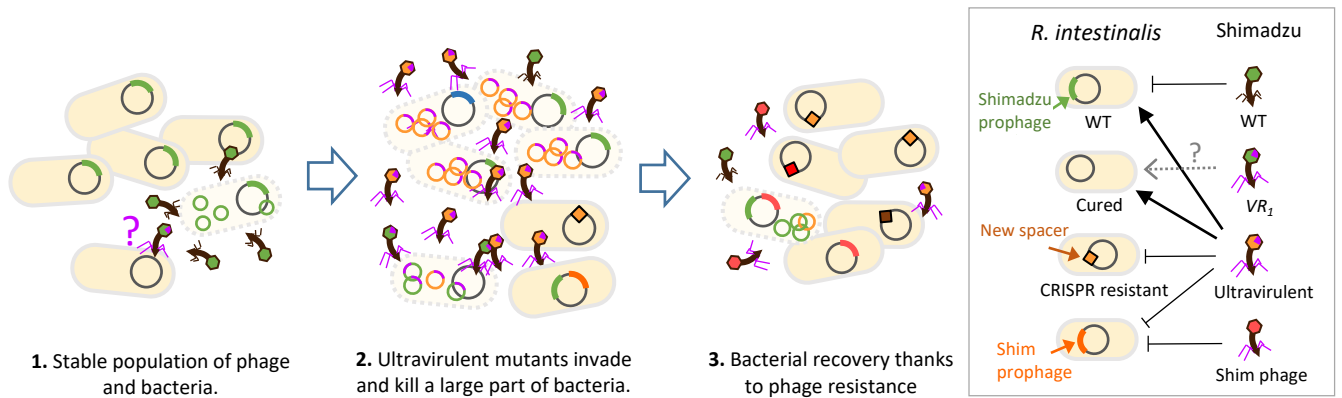s. **C)** Type II-C CRISPR-cas system of *R. intestinalis* L1-82 and position of primers used for detection of spacer acquisition. Positions in the bacterial genome are indicated. **D)** Schematic representation of Shimadzu *attL* and *attB* sites, and their detection by PCR. **E)** For 52 isolated clones, are indicated: Shivir5 resistance (upper line), spacer acquisition (middle line), presence of *attB* or *attL* sites (lower line), and presence of a Shim prophage (*). The color of clone number corresponds to the mouse from which it has been isolated, as in Fig. 2. Sequenced clones correspond to numbers 1, 2, 4 and 6. Clones isolated at earlier time points were all identical to those of day 19 and therefore not included in the figure. **F)** Frequency of clones with 0, 1 or 2 new spacers as a function of isolation time. **G)** Spacer acquisition in total *R. intestinalis* bacteria from faeces at four different time points (day 16, day 21, day 25 and day 33). DNA ladder is a 50 pb scale. Smaller PCR fragments correspond to the wild-type size (indicated by "0"), the acquisition of 1 spacer to "+1" and 2 spacers to "+2". **H)** Population level of Shimadzu cured *R. intestinalis,* in mice from different cage. Cage is color coded as in B panel.

Figure 5
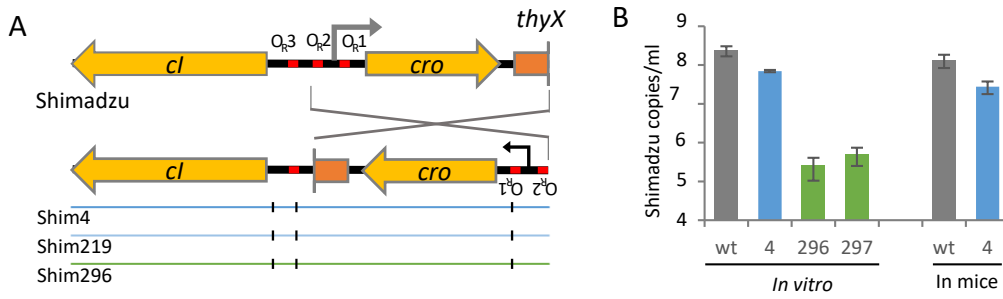
Figure 5. Prophage Shim has an inverted *cro* region. **A)** Zoom on the immunity region, showing the *cro* inversion in Shim prophage and the 3 point mutations. **B)** Spontaneous Shimadzu virion production in clones 4, 296, 297 and in L1-82 ancestral clone, *in vitro* of after 3 days of mice colonization. Clone 4 is a dilysogen harbouring Shimadzu and Shim, the other clones only contain a Shim copy. Mean ± s.d of three measures.

Figure 6

1. Stable population of phage and bacteria.

2. Ultravirulent mutants invade and kill a large part of bacteria.

3. Bacterial recovery thanks to phage resistance

Figure 6. Summary sketch of the results. In a first step, stable populations of free phage and bacteria are observed. Yet, complex dynamics might already occur, as 0.1% of bacteria are cured of the Shimadzu prophage, and phages with mutant VR_1 alleles are selected**.** The second step corresponds to the free phage bloom: ultravirulent Shivir mutants invade and multiply on bacteria. Phages with a restored lysogenisation ability (Shim phages) are generated. In the third step, resistant bacteria invade, thanks to CRISPR spacer acquisition and to lysogenization by a new immunity type prophage.

Figure S1

A.

Jekyll

Labels (left to right): Integrase, Mu-Gam Recombinase, SSB, Methyltransferase, Terminase, Major Capsid protein, TMP, Holin, Endolysin, RexA

1 Kbp

B

Prophage	1	MPTSKNTARKYAAGGSSDNSCGVEPLRAVIYARYSSSGQREESIEGQLRDCYEFAKKHGIIVIGEYIDKAMTGRVDRRPDFQRMMKDSEKGRFNCVLLWK
Free phage	1	MPTSKNTARKYAAGGSSDNSCGVEPLRAVIYARYSSSGQREESIEGQLRDCYEFAKKHGIIVIGEYIDKAMTGRVDRRPDFQRMMKDSEKGRFNCVLLWK
Prophage	101	MDRFARNRYDSAMYKYKLKKNGIRIFYAKETIPDGPEGIILESVMEGYAEYYSENLAQNVKRGNYDSALELKTLGKTCLGLKTGPDGRYMIDQAEAAIVR
Free phage	101	MDRFARNRYDSAMYKYKLKKNGIRIFYAKETIPDGPEGIILESVMEGYAEYYSENLAQNVKRGNYDSALELKTLGKTCLGLKTGPDGRYMIDQAEAAIVR
Prophage	201	RIFEEYAEGERAKDIYERLNSEGYRTSRGGKFNKNSLRRILSNKKYIGVYEYEDIYVENGIPAIITDRDLFERVQKMLKINHDAPARGKAQNFLLTTKLF
Free phage	201	RIFEEYAEGERAKDIYERLNSEGYRTSRGGKFNKNSLRRILSNKKYIGVYEYEDIYVENGIPAIITDRDLFERVQKMLKINHDAPARGKAQNFLLTTKLF
Prophage	301	CGLCGSPMIGDGGTSHTGKAYAYYSCTKRKRGRSCKKESVPKDWIEDLVVGELVKIVHNDELIEQIADRVMEYQKREKDQSGLHALEIRQKENEKAISNM
Free phage	301	CGLCGSPMIGDGGTSHTGKAYAYYSCTKRKRGRSCKKESVPKDWIEDLVVGELVKIVHNDELIEQIADRVMEYQKREKDQSGLHALEIRQKENEKAISNM
Prophage	401	LAAIEAGIITPSTKTRLMELEADRADIEKGIAHELLAEPEFERDQIIYFLERFRSGDINDEAYRIMLVDTFLNSVYLYDDDHLVLVMNYSGENCKVDLKL
Free phage	401	LAAIEAGIITPSTKTRLMELEADRADIEKGIAHELLAEPEFERDQIIYFLERFRSGDINDEAYRIMLVDTFLNSVYLYDDDHLVLVMNYSGENCKVDLKL
Prophage	501	VEGAVSGDGCKGSAFAPSSALNGAKPNPVYFFK-------KVFAVVVNFTRK--
Free phage	501	VEGAVSGDGCKGSAFAPSSAL---KAEPGVLFQESVRRGGKLHAEIAKNKSVRA

C.

Shimadzu

Lagaffe

Shimadzu labels (left to right): Integrase, Methyltransferase, DNA Pol, Cell wall hydrolase, Methyltransferase, Methyltransferase, Erf, SSB, Rep_O, DnaI, Methyltransferase, Resolvase, Methyltransferase, Tail completion, TermS, TermL, Portal, MCP, Sheath, MTP, TMP, Baseplate J, P2_J, Avd, RT, Endolysin

65%

Figure S1. Peculiar genomic features of the two prophages. A) Read coverage of Jekyll prophage. The region comprising capsid genes present an 8-fold higher coverage than the rest of the genome. B) Alignment of the Shimadzu integrase sequence in the prophage (upper line) and free phage (lower line) forms. The last 20 amino acids are different. C) Shimadzu and Lagaffe (GenBank MG711461) genomic comparison. Blast was made using BlastN. The two phages share 60% of their proteins and overall good synteny.

Figure S2

A.
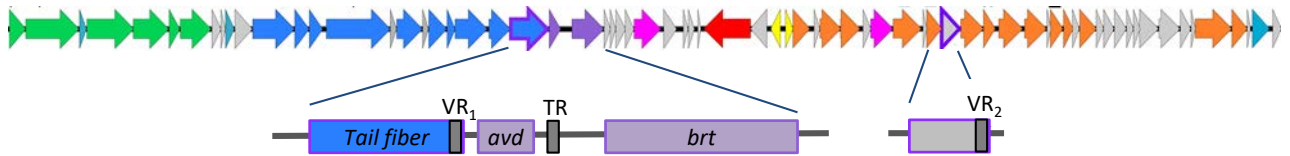
B.

VR$_1$ (18,935 - 19,063)

GCGGCCGCTGCTACTGGTGGTCCGCTTCTGTCAGGAGTGGCGGCTCGACTAATTGTGTGTTCGTCAACGTCAGCGGGCACTCCAGCAACTGGGGCGCCAGCGGCGGGCTTTACGTCCCGGTCTGCTTCC

TR (19,582 - 19,710)

GCGGCCGCTGCAACTGGTGGACCGCTTCTGTCAGGAGTGGCAACTCGACTAATTGCGTGAACGTCAACAACAACGGGAACTCCAACAACTGGAACGCCAGCAACGAGCTTTACGTCCCGGTCTGATTCC

VR$_2$ (33,472 - 33,585)

TAC- - - TGGACCGCTTCTGTCAGGAGTGGCGGCTCGACTGGTTGCGTGTCCGTCAGCTACCACGGGCACTCCTACGACTGGAGCGCCAGCACCGAGCTTTACGTCCCGGTCTGCTTCG

Figure S2. Diversity Generating Retroelement of Shimadzu. A) general organisation of the DGR in the Shimadzu genome. Color legend is as in Fig. 1. B) Sequences of the template repeat (TR), of the variable repeat VR$_1$ in a tail gene (VR1), and of the second variable repeat VR$_2$ in *orf44*. As already described, mainly TR adenines are modified in both VR (in red). Their positions in the Shimadzu genome are indicated (free phage form of the genome).

Figure S3
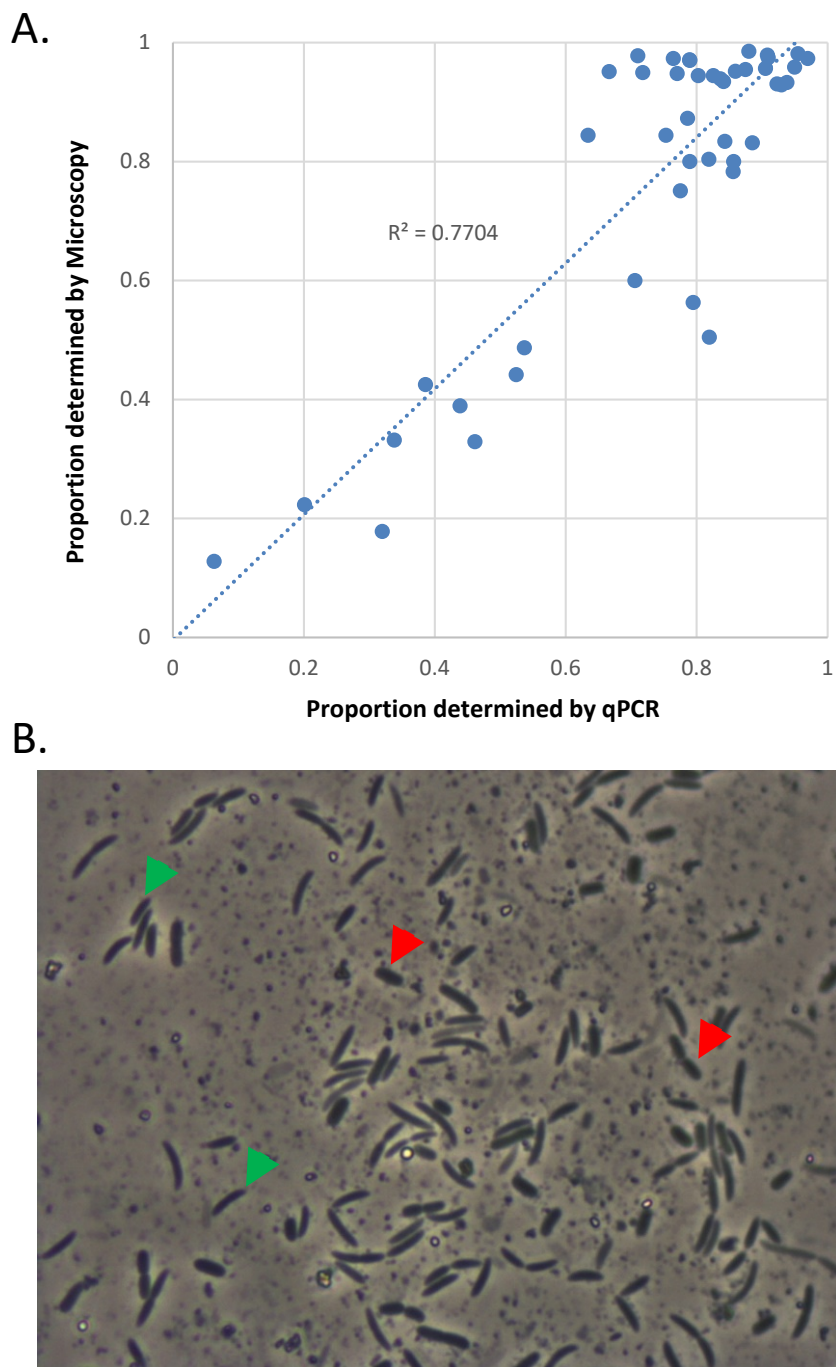
Figure S3: Comparison of quantification by qPCR and by microscopy. **A)** Proportion of *R. intestinalis* determined by quantitative PCR (qPCR) in function of the proportion determined by microscopy. An overall good correlation is observed. **B)** Observation of faeces by microscopy. *R. intestinalis* cells (green arrows) are clearly distinguishable from *E. coli* cells (red arrows).

Figure S4
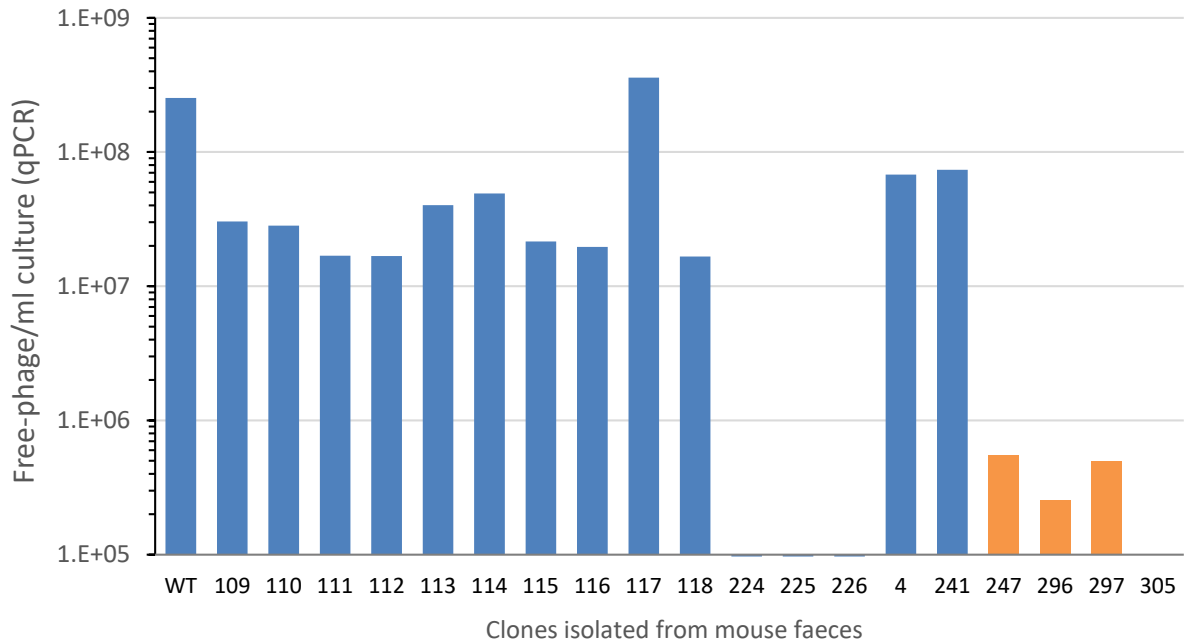
Figure S4: *In vitro* free phage production of clones isolated from mouse faeces. Shimadzu genome copy number in 16 hours saturated cultures. Clone numbers are those of Fig. 5. Orange bars correspond to clones in which a Shim prophage was detected. Clones 224, 225 226 and 305 are devoid of a Shimadzu prophage.