The Gut Microbiota Facilitates Drifts in the Genetic Diversity and Infectivity of Bacterial Viruses

Luisa de Sordi, Varun Khanna, Laurent Debarbieux

Title:

The gut microbiota facilitates drifts in the genetic diversity and infectivity of bacterial viruses

Authors:

Luisa De Sordi[1], Varun Khanna[2], Laurent Debarbieux[1,3*]

Affiliations:

[1] Department of Microbiology, Institut Pasteur, Paris F-75015 France.

[2] Hub Bioinformatique et Biostatistique, Institut Pasteur – C3BI, USR 3756 IP CNRS, Paris F-75015, France.

[3] Lead Contact

[*] Correspondence: laurent.debarbieux@pasteur.fr

Summary

The intestinal microbiota and human health are intimately linked but interactions between bacteria and bacteriophages in the context of the mammalian intestine remain largely unexplored. We used comparative population genomics to study a tripartite network consisting of a virulent bacteriophage, its bacterial host, and a phage-insensitive bacterial strain both *in vitro* and within the murine gut. The bacteriophage adapted to infect the insensitive strain when the three partners co-existed in the gut of conventional mice, but not dixenic mice or in planktonic cultures. The molecular changes associated with modifications in the bacteriophage host spectrum included single amino-acid substitutions and an unusual homologous intra genomic recombination event within the genome of the bacteriophage. An intermediate bacterial host isolated from the murine microbiota mediated bacteriophage adaptation. Our data indicate that by offering access to new hosts, the microbiota shifts the genetic diversity of bacteriophages thereby promoting long-term persistence of bacteriophage populations.

Introduction

Viruses are the most abundant biological entities in the intestinal microbiome, with a marked predominance of bacteriophages (Minot *et al.*, 2013; Norman *et al.*, 2014; Virgin, 2014). Culture-independent metagenomic approaches have been successful in characterising viral communities, but these methods are largely uninformative about their relationships with the matching bacterial hosts in the gut (Breitbart *et al.*, 2003; Minot *et al.*, 2011; Mirzaei and Maurice, 2017). This is particularly true for virulent bacteriophages: their sequences, unlike those of temperate ones, cannot be deduced from bacterial genomes and identifying their vehicles of replication in the intestinal microbiome is often unachievable. For instance, the CrAssphage is the most widespread bacteriophage in the human population, but its host is still not known (Dutilh *et al.*, 2014).

All humans carry a unique set of intestinal bacteriophages (Faith *et al.*, 2013). However, healthy individuals share a pool of conserved bacteriophages, which are different in type, number and abundance, from those found in patients suffering from Inflammatory Bowel Disease (IBD) (Manrique *et al.*, 2016; Norman *et al.*, 2015). More, significant bacteriophage changes in relative abundance, richness and diversity were recently reported to be associated with the development of human autoimmune disease (Zhao *et al.*, 2017). Indeed, there is increasing evidence that bacteriophages make a substantial contribution to the eubiosis of the human microbiome, although the mechanisms are yet to be defined (De Paepe *et al.*, 2014; Manrique *et al.*, 2017). In that perspective, it is no wonder that viral communities homogenise between donors and patients undergoing successful faecal microbiota transplant (FMT) (Chehoud *et al.*, 2016; Zuo *et al.*, 2017). However, it remains remarkable that filtered (bacteria-free) FMT

provides curative results comparable to those of traditional FMT in resolving gut infections by *Clostridium difficile* (Ott *et al.*, 2016). This implies that the viral community can re-establish itself by replicating on the new hosts available and thereby contribute to reshaping the bacterial population towards a healthy condition.

Bacteriophage survival depends on the ability to perpetuate the antagonistic coevolution with bacterial hosts. This requires maximising the viral infection capacity towards each new population encountered (Paterson *et al.*, 2010). When the ecological conditions are perturbed, this relationship can be destabilised with consequent increased extinction risk for the bacteriophage (Wright *et al.*, 2016). To limit this risk, viral genome plasticity generates genomic diversity to face the spatial and temporal variations typical of antagonistic interactions (Buckling and Rainey, 2002; Vos *et al.*, 2009).

We hypothesised that the bacterial heterogeneity in the intestinal tract, would expand the genomic diversification of the coevolving viral population, promoting bacteriophage adaptation to different hosts. Using three experimental conditions (*in vitro* and in dixenic and conventional mice) we identified molecular events allowing one virulent bacteriophage to adapt to previously insensitive host bacteria, a phenomenon observed exclusively in the intestine.

We propose for bacteriophages to persist within this highly competitive environment they exploit the microbiota as a base to jump between different bacterial hosts.

Results

The intestinal environment promotes bacteriophage host-jump

Our experimental approach involved a tripartite network consisting of: *(i)* a novel virulent bacteriophage, P10, related to the myovirus Felix O1; *(ii)* its host, the pathogenic *Escherichia coli* strain LF82 and *(iii)* the commensal *E. coli* strain MG1655, to which P10 cannot bind, and is therefore non-infective (up to 10^{10} pfu were tested) (Figure S1). The behaviour of these three partners was monitored for 24 days in parallel in three conditions: *in vitro* (planktonic cultures), in the gut of mice only colonised with strains LF82 and MG1655 (dixenic) and in the gut of conventional mice (Figure 1A and see Figure S2 for bacteriophage population densities). We screened for "host-jumps" or, in other words, a gain of function by P10; the function being the ability to infect the previously insensitive host, strain MG1655. No endogenous bacteriophage infecting either of the strains was isolated from mouse intestine from the control group. Samples were analysed by plaque assay and sequencing of total bacteriophage populations (see below). Interestingly, bacteriophage P10 variants adapted to strain MG1655 (hereafter named ad_P10) and still infecting strain LF82 were isolated, but only from the gut of conventional mice (20% rate). No P10 variants able to jump hosts were obtained from dixenic mice or *in vitro* (Figure 1A and S2).

Adapted bacteriophages have extensively diversified host range

We investigated whether the ability of ad_P10 in infecting a previously insensitive host was associated with a more generalised effect on the bacteriophage host range. We isolated 16 ad_P10 bacteriophages and tested them on a collection of 105 Enterobacteriaceae strains

(Figure 2 and Table S1). Almost half of the ad_P10 isolates had a reduced host range than the wild type strain (up to 50% reduction), and most others showed marginal difference; one exceptional ad_P10 variant had expanded its host range to 11 more strains than the wild-type P10 (Figure 2).

An unusual genome recombination event associated with host jump in virulent bacteriophages

To gain insight into the host jump mechanism and its variable effect on the host range of adapted bacteriophages, the genomes of two ad_P10 isolates were sequenced. In both genomes, a stretch of 216 or 280 nucleotides of the hypervariable region of the tail fiber gene *gp91* underwent duplication, resulting in two interspersed repeats. The segment recombined within the hypervariable domain of the adjacent long tail fiber gene *gp90,* causing the deletion of either 237 or 273 nucleotides. This recombination was probably promoted by the presence of homologous flanking regions of 17 to 22 base pairs (Figure 3A). We named this event a "homologous intra genomic recombination" (hIGR) and investigated its frequency by sequencing the corresponding region in 20 randomly chosen ad_P10 isolates: it was present in 10 of the 20 isolates tested. Recombination is a major factor in genomic evolution, so we tested whether similar events had taken place in other bacteriophage genomes. A search in the database of all annotated bacteriophage tail fibers revealed that three other bacteriophages of the FelixO1-like family displayed the same tail fiber protein organisation as ad_P10: 60- to 70-amino acids in the C-terminal hypervariable region were duplicated in the adjacent tail fiber-

encoding gene sequence (Figure 3B). Similar hIGR events were also found, but at different positions, in the tail fiber genes of other bacteriophages (Figure 3B and Table S2).

Population genomic analyses indicates that the patterns of mutation differ between *in vitro* and *in vivo* populations

We used population genomics analyses to compare P10 populations evolving *in vitro*, in dixenic mice and in conventional mice. The number of single-nucleotide polymorphisms (SNPs) accumulated over 24 days (17 to 28 SNPs, at frequency higher than 1%; Table S3) did not differ significantly (p=0.67) and did not depend on the presence of strain MG1655 in the system (p=0.1) or on the host-jump event (p=0.93) (Mann-Whitney test). This suggests that the adaptation was not a consequence of a higher mutation rate in the intestinal environment than in laboratory cultures.

We found that gene *gp37*, which is homologous to *rIIA* (one of the *r* genes for lysis inhibition in T4 (Paddison *et al.*, 1998)) was a hotspot for mutations in all populations, implicating this gene in general viral fitness that was not investigated further (Figure 1B-C). Some genomic regions were differentially mutated *in vitro* versus *in vivo*. In liquid cultures, P10 persistently infected only one host, strain LF82, and none of the SNPs in these populations were highly frequent (< 20%), suggesting that they had a limited role in viral fitness. The genes most frequently affected in this condition included those encoding enzymes involved in nucleic acid metabolism, possibly fine-tuning the infection mechanism (Figure 1B-C and Table S3).

By contrast, numerous mutations were only found in bacteriophages replicating in the intestinal tract. Many were present at high frequency and were located in the genomic region

136 predicted to encode for structural proteins. Most of the high-frequency SNPs, like the hIGR
137 described above, were in the two tail fiber genes (*gp90* and *gp91*).
138
139 **A single point mutation in a tail fiber gene is sufficient for a host jump**
140 Tail fiber gene *gp91* was only modified in ad_P10 populations (those that showed a host-jump)
141 with a particular SNP (A→ G at position 55079) at a frequency of 100% (confirmed by PCR
142 sequencing of 20 individual ad_P10 bacteriophages; Figure 1B-C and Table S3). It was absent
143 from non-adapted populations from all three conditions tested. This variant results in a single-
144 amino acid change at position 284 (Y284H). We tested its involvement in host specificity and
145 constructed a recombinant P10 bacteriophage (rP10) carrying this mutation (Y284H): this single
146 SNP was sufficient to confer the ability to infect strain MG1655 (Figure S3). However, this
147 happened at the expense of bacteriophage replication in its original host, strain LF82 (decrease
148 of nearly 100-fold) without conferring highly efficient replication in strain MG1655 (Figure S3).
149 However, consecutive replication in this new host rapidly led to adaptation such that rP10
150 replication in strain MG1655 was ten times higher than P10 in strain LF82 (Figure S3). As few as
151 four replication passages *in vitro* were sufficient for this massive gain of infectivity on the new
152 host. Genomic sequencing showed that this improvement was accompanied by additional
153 mutations in the genome of rP10 that likely contributed to increasing phage replication
154 efficiency on the new host. We counted 7 additional SNPs, of which 4 were non-synonymous
155 and one deletion (1bp) (Table S4). Two of the non-synonymous mutations were located within
156 each of the tail fiber genes, one in the DNA polymerase gene and the others in genes of
157 unknown function (Table S4).

Bacterial diversity within the microbiota favours viral host-jump

We investigated why host-jumps were only observed for bacteriophages replicating in the gut of conventional mice. We hypothesised that this event is mediated by the infection of one or more intermediate bacterial hosts within the intestinal microbiota leading to the observed modifications of the bacteriophage genetic diversity. We isolated a gut bacterial strain that was infected by ad_P10 at high efficacy of plating (EOP) of 1 from faeces of conventional mice not exposed to any bacterial inoculum (see methods). By contrast, wild-type P10 was only poorly able to form plaques (EOP=10^{-2}) on this strain. This supports the idea that the P10 host-jump is concomitant with an increased lytic efficacy against an alternative host in the same environment. Sequence analysis identified this bacterium as a murine *E. coli* strain that we named Mouse *E. coli* (MEc) 1. We studied the role of this putative intermediate host in the host-jump of P10 resulting in ad_P10 by using co-cultures *in vitro*: we mixed P10 with strains MEc1 or MG1655, with or without the presence of strain LF82, or with a mix of the three *E. coli* strains. After seven days of incubation, the efficacy with which the wild-type P10 population infected MEc1 increased 10-fold whenever this strain was present (Figure 4A). If strain MG1655 was also present, the host jump was observed, although it was delayed to day 24 when the preferential host of P10, LF82, was in the system (Figure 4A-B). By contrast, no host-jump was observed when P10 was mixed *in vitro* in the presence of LF82 and MEc1, alone or in combination, nor, as expected, in the presence of MG1655 alone (Figure 4A-B). This indicates that host-jumps become prevalent when there is a strong pressure conferring advantage on the

new infecting population, such as in conditions where the density of the preferential hosts is low or null.

Discussion

To persist in the environment, obligate parasites, such as viruses, rely on the availability of suitable hosts. Genetic mutations during viral replication can generate viral variants that can acquire the ability to infect new hosts (Poirier and Vignuzzi, 2017). This is illustrated by several human pandemics caused by viruses originally incapable of infecting human cells (for example HIV, flu virus H1N1, and Ebola virus). Our work demonstrates that the abundance of potential hosts favours shifts in viral infectivity: this suggests that the diversity of intestinal microbiota enables the persistence of bacteriophages by promoting diversification due to the availability of divergent infection paths.

Numerous studies have described the behaviour of bacteria and bacteriophages *in silico* and *in vitro*, defining population dynamics, emergence of bacterial resistance and bacteriophage adaptation (Betts *et al.*, 2014; Kerr *et al.*, 2006; Morgan *et al.*, 2005; Weitz *et al.*, 2013; Wichman *et al.*, 2005). However, how these phenomena apply to the homeostasis of human viral communities remains largely unknown (Scanlan, 2017).

The host jump that we observed in the intestinal tract had a large effect on the host spectrum of the P10 population. In the murine intestinal tract, this resulted in a host expansion from one to two or three (possibly more) bacterial strains coevolving with P10. However, the infectivity towards a set of strains that has not coevolved with the bacteriophage, was very variable and often restricted, compared to the wild-type P10. There are at least two possible explanations for these observations: either the host-jump emerged several times in different bacteriophage variants, or a single event was followed by additional mutations affecting their host range phenotype.

Environmental studies have shown that the structure of bacteriophage-host infection networks are typically nested, providing a stable community at the site of sampling (Beckett and Williams, 2013; Flores *et al.*, 2011; Koskella and Brockhurst, 2014). Our results confirm that this pattern is not limited to bacterial hosts isolated from the same environment; it also applies to the ability to infect other potential host strains, reflecting a more generalised, space- and time-independent pattern of behaviour.

Host-jumps were concomitant with the accumulation of mutations in the P10 genome. These mutations largely differed from those found *in vitro* and were mostly in the tail fiber genes. Amongst those, we report a previously undescribed hIGR and provide evidence for similar hIGRs in other bacteriophages, with implications for viral fitness and possibly adaptation to different hosts. Modular exchange between different bacteriophages is known to drive genomic evolution by homologous, illegitimate or site-specific recombination (Brussow *et al.*, 2004; Ignacio-Espinoza and Sullivan, 2012; Yap and Kreuzer, 1991). Here we provide a demonstration of intra-genomic recombination in bacteriophages not harbouring genes homologous to known recombinases, such as all FelixO1-like bacteriophages. Possibly, these bacteriophages exploit the host recombinase machinery, for example that associated with integrated prophages (LF82 harbours four of them (Miquel *et al.*, 2010)), to mediate the evolution of their own genomes which contain “ready-to-use” recombination sites. Such recombination would be a novelty in virulent bacteriophages and further investigation is required to determine whether it can be found elsewhere than in tail fiber genes.

In all adapted bacteriophages, host-jump was associated to accumulation of point mutations. Although many localized in the tail fiber genes, one particular SNP was sufficient in itself to

drive host-jump. However, it did not confer a selective infectious advantage over the wild-type bacteriophage population which continues to replicate on the available original host (optimal fitness). This provides an explanation of why it was not selected and amplified in planktonic cultures or in dixenic mice in the presence of only two bacterial strains. This is consistent with the results of previous experimental and theoretical coevolutionary studies (Ashby *et al.*, 2014). In antagonistic coevolution, the arms-race dynamics between viruses and hosts lead to increased viral infectivity via a constant supply of genomic mutations. However this comes at a cost in terms of fitness, preventing widely infective variants from fixing (Buckling and Rainey, 2002; Gandon *et al.*, 2008; Poullain *et al.*, 2008) and favouring fluctuating selection on alternative host genotypes (Gomez and Buckling, 2011; Hall *et al.*, 2011). Our study widen this concept to infection of different host strains (host-jump) generating viral diversity directly proportional to host diversity.

The kinetics of events leading to host jump is not known. The single SNP we report might have appeared first, followed by other adaptive mutations, as we observed with the specific recombinant variant we generated, or it could have been the other way around. These events are likely to be affected by the uneven spatial distribution of bacteriophage hosts in the intestine. Indeed, their accessibility differs between sites and their density and diversity can vary providing different opportunities for bacteriophage spatial adaptation. Here, the available hosts can drive fluctuating selection of differentially infectious bacteriophage variants resulting from the accumulation of genomic mutations. A schematic representation of fitness landscape showing viral subpopulations and the available hosts is presented in Figure 4C. The extreme endpoints are the extinction of the least fit and least competitive variants, and the expansion,

and probably improved persistence, of those acquiring the ability to infect previously insensitive hosts. Although we did not found significant differences in the accumulation of SNPs when comparing our three experimental conditions, our results should not suggest that the surrounding environment has no impact on genomic evolution. Accumulation of deleterious mutation leads to inevitable extinction, but this happens at a more rapid pace *in vivo* (in the gut), being this an "open" system where the least fit can be eliminated, compared to *in vitro* closed conditions (in flasks). Therefore, mutations found in populations isolated from the intestinal tract are more likely to only reflect advantageous genomic variants.

We propose that these interactions play key roles in maintaining bacteriophage and microbial diversities in the intestinal tract. In this context, disruption of such bacteriophage-bacteria interactions may play crucial role in causing dysbiosis associated to antibiotic treatments or inflammatory disease. Our work shows that the microbiota is a neglected source of evolutionary pressure: it promotes drifts in viral genomic diversity and host ranges by facilitating host jumping and providing an opportunity to persist in this ecological niche.

Author Contributions: Conceptualization, L.D.S. and L.D.; Methodology, L.D.S., V.K., L.D.; Investigation, L.D.S.; Formal Analysis, L.D.S. and V.K.; Writing, L.D.S., V.K. and L.D.; Funding Acquisition, L.D.

Acknowledgments:

This work was supported by the DigestScience Foundation (Lille, France). We thank Guillaume Achaz, Mart Krupovic, Sylvain Gandon, Eduardo Rocha and Marco Vignuzzi for the critical reading of an early version of this manuscript. We also thank the Centre de Gnotobiologie and the sequencing platform of Institut Pasteur for the use of their facilities and managing help. We thank the teams of Christel Neut (Communauté d'Universités et d'Etablissements Lille Nord de France, Inserm U995, Lille, France) and Nicolas Barnich (Université Clermont Auvergne, Clermont-Ferrand, France) for sharing their *E. coli* collections. We are grateful to Matthieu Galtier, Marine Henry and Damien Maura for setting up the co-colonisation protocol and to Chantal Archambeau and Estelle Atse for media and buffers preparation.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Ashby, B., Gupta, S., and Buckling, A. (2014). Effects of epistasis on infectivity range during host-parasite coevolution. Evolution *68*, 2972-2982.

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

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243-1247.

Beckett, S.J., and Williams, H.T. (2013). Coevolutionary diversification creates nested-modular structure in phage-bacteria interaction networks. Interface Focus *3*.

Betts, A., Kaltz, O., and Hochberg, M.E. (2014). Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages. Proc Natl Acad Sci U S A *111*, 11109-11114.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., and Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol *185*, 6220-6223.

Brussow, H., Canchaya, C., and Hardt, W.D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev *68*, 560-602, table of contents.

Buckling, A., and Rainey, P.B. (2002). Antagonistic coevolution between a bacterium and a bacteriophage. Proc Biol Sci *269*, 931-936.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. (2005). ACT: the Artemis Comparison Tool. Bioinformatics *21*, 3422-3423.

Chehoud, C., Dryga, A., Hwang, Y., Nagy-Szakal, D., Hollister, E.B., Luna, R.A., Versalovic, J., Kellermayer, R., and Bushman, F.D. (2016). Transfer of Viral Communities between Human Individuals during Fecal Microbiota Transplantation. MBio *7*, e00322.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80-92.

Da Re, S., Le Quere, B., Ghigo, J.M., and Beloin, C. (2007). Tight modulation of Escherichia coli bacterial biofilm formation through controlled expression of adhesion factors. Appl Environ Microbiol *73*, 3391-3403.

Da Re, S., Valle, J., Charbonnel, N., Beloin, C., Latour-Lambert, P., Faure, P., Turlin, E., Le Bouguenec, C., Renauld-Mongenie, G., Forestier, C., *et al.* (2013). Identification of commensal Escherichia coli genes involved in biofilm resistance to pathogen colonization. PLoS One *8*, e61628.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.

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

Debarbieux, L., Leduc, D., Maura, D., Morello, E., Criscuolo, A., Grossi, O., Balloy, V., and Touqui, L. (2010). Bacteriophages can treat and prevent Pseudomonas aeruginosa lung infections. J Infect Dis *201*, 1096-1104.
Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., *et al.* (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun *5*, 4498.
Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C., Leibel, R.L., *et al.* (2013). The long-term stability of the human gut microbiota. Science *341*, 1237439.
Flores, C.O., Meyer, J.R., Valverde, S., Farr, L., and Weitz, J.S. (2011). Statistical structure of host-phage interactions. Proc Natl Acad Sci U S A *108*, E288-297.
Galtier, M., De Sordi, L., Neut, C., and Debarbieux, L. (2016). Bacteriophages targeting adherent invasive Escherichia coli strains as a promising new treatment for Crohn's disease. J Crohns Colitis.
Gandon, S., Buckling, A., Decaestecker, E., and Day, T. (2008). Host-parasite coevolution and patterns of adaptation across time and space. J Evol Biol *21*, 1861-1866.
Gomez, P., and Buckling, A. (2011). Bacteria-phage antagonistic coevolution in soil. Science *332*, 106-109.
Hall, A.R., Scanlan, P.D., Morgan, A.D., and Buckling, A. (2011). Host-parasite coevolutionary arms races give way to fluctuating selection. Ecol Lett *14*, 635-642.
Hyman, P., and Abedon, S.T. (2009). Practical methods for determining phage growth parameters. Methods Mol Biol *501*, 175-202.
Ignacio-Espinoza, J.C., and Sullivan, M.B. (2012). Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. Environ Microbiol *14*, 2113-2126.
Kerr, B., Neuhauser, C., Bohannan, B.J., and Dean, A.M. (2006). Local migration promotes competitive restraint in a host-pathogen 'tragedy of the commons'. Nature *442*, 75-78.
Koskella, B., and Brockhurst, M.A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol Rev *38*, 916-931.
Manrique, P., Bolduc, B., Walk, S.T., van der Oost, J., de Vos, W.M., and Young, M.J. (2016). Healthy human gut phageome. Proc Natl Acad Sci U S A *113*, 10400-10405.
Manrique, P., Dills, M., and Young, M.J. (2017). The Human Gut Phage Community and Its Implications for Health and Disease. Viruses *9*.
Mead, R. (1988). The Design of Experiments: Statistical Principles for Practical Applications. Cambridge: Cambridge University Press.
Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Rapid evolution of the human gut virome. Proc Natl Acad Sci U S A *110*, 12450-12455.
Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. Genome Res *21*, 1616-1625.
Miquel, S., Peyretaillade, E., Claret, L., de Vallee, A., Dossat, C., Vacherie, B., Zineb el, H., Segurens, B., Barbe, V., Sauvanet, P., *et al.* (2010). Complete genome sequence of Crohn's disease-associated adherent-invasive E. coli strain LF82. PLoS One *5*.
Mirzaei, M.K., and Maurice, C.F. (2017). Menage a trois in the human gut: interactions between host, bacteria and phages. Nat Rev Microbiol *15*, 397-408.
Morgan, A.D., Gandon, S., and Buckling, A. (2005). The effect of migration on local adaptation in a coevolving host-parasite system. Nature *437*, 253-256.

Norman, J.M., Handley, S.A., Baldridge, M.T., Droit, L., Liu, C.Y., Keller, B.C., Kambal, A., Monaco, C.L., Zhao, G., Fleshner, P., *et al.* (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. Cell *160*, 447-460.

Norman, J.M., Handley, S.A., and Virgin, H.W. (2014). Kingdom-agnostic metagenomics and the importance of complete characterization of enteric microbial communities. Gastroenterology *146*, 1459-1469.

Norrander, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene *26*, 101-106.

Ott, S.J., Waetzig, G.H., Rehman, A., Moltzau-Anderson, J., Bharti, R., Grasis, J.A., Cassidy, L., Tholey, A., Fickenscher, H., Seegert, D., *et al.* (2016). Efficacy of Sterile Fecal Filtrate Transfer for Treating Patients With Clostridium difficile Infection. Gastroenterology.

Paddison, P., Abedon, S.T., Dressman, H.K., Gailbreath, K., Tracy, J., Mosser, E., Neitzel, J., Guttman, B., and Kutter, E. (1998). The roles of the bacteriophage T4 r genes in lysis inhibition and fine-structure genetics: a new perspective. Genetics *148*, 1539-1550.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A.J., Thomson, N.R., Quail, M., Smith, F., Walker, D., Libberton, B., *et al.* (2010). Antagonistic coevolution accelerates molecular evolution. Nature *464*, 275-278.

Pickard, D.J. (2009). Preparation of bacteriophage lysates and pure DNA. Methods Mol Biol *502:3-9.*, 10.1007/1978-1001-60327-60565-60321_60321.

Poirier, E.Z., and Vignuzzi, M. (2017). Virus population dynamics during infection. Curr Opin Virol *23*, 82-87.

Poullain, V., Gandon, S., Brockhurst, M.A., Buckling, A., and Hochberg, M.E. (2008). The evolution of specificity in evolving and coevolving antagonistic interactions between a bacteria and its phage. Evolution *62*, 1-11.

Saussereau, E., Vachier, I., Chiron, R., Godbert, B., Sermet, I., Dufour, N., Pirnay, J.P., De Vos, D., Carrie, F., Molinari, N., *et al.* (2014). Effectiveness of bacteriophages in the sputum of cystic fibrosis patients. Clin Microbiol Infect *20*, O983-990.

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

Thomason, L.C., Costantino, N., and Court, D.L. (2007). E. coli genome manipulation by P1 transduction. Curr Protoc Mol Biol *Chapter 1*, Unit 1 17.

Virgin, H.W. (2014). The virome in mammalian physiology and disease. Cell *157*, 142-150.

Vos, M., Birkett, P.J., Birch, E., Griffiths, R.I., and Buckling, A. (2009). Local adaptation of bacteriophages to their bacterial hosts in soil. Science *325*, 833.

Weitz, J.S., Poisot, T., Meyer, J.R., Flores, C.O., Valverde, S., Sullivan, M.B., and Hochberg, M.E. (2013). Phage-bacteria infection networks. Trends Microbiol *21*, 82-91.

Wichman, H.A., Millstein, J., and Bull, J.J. (2005). Adaptive molecular evolution for 13,000 phage generations: a possible arms race. Genetics *170*, 19-31.

Wright, R.C., Brockhurst, M.A., and Harrison, E. (2016). Ecological conditions determine extinction risk in co-evolving bacteria-phage populations. BMC Evol Biol *16*, 227.

Yap, W.Y., and Kreuzer, K.N. (1991). Recombination hotspots in bacteriophage T4 are dependent on replication origins. Proc Natl Acad Sci U S A *88*, 6043-6047.

Zhao, G., Vatanen, T., Droit, L., Park, A., Kostic, A.D., Poon, T.W., Vlamakis, H., Siljander, H., Harkonen, T., Hamalainen, A.M., *et al.* (2017). Intestinal virome changes precede autoimmunity in type I diabetes-susceptible children. Proc Natl Acad Sci U S A.

Zuo, T., Wong, S.H., Lam, K., Lui, R., Cheung, K., Tang, W., Ching, J.Y.L., Chan, P.K.S., Chan, M.C.W., Wu, J.C.Y., *et al.* (2017). Bacteriophage transfer during faecal microbiota transplantation in Clostridium difficile infection is associated with treatment outcome. Gut.

FIGURE LEGENDS

Figure 1. **Bacteriophage host-jump only occur in the gut of conventional mice.**

(A) Schematic representation of three experimental conditions, where bacteriophage P10 was mixed with two *E. coli* strains (MG1655 and LF82), *in vitro* (IV), *in vivo* in dixenic mice (DM) or conventional mice (CM) for 24 days. Inoculation with bacteria and P10 is shown as black arrows. Grey arrows indicate control (Ct) sampling prior to inoculation, followed by sampling of bacteria and P10. The orange box shows recovery of adapted (ad_) bacteriophage only in conventional mice. The relative bacteriophage densities are shown in Figure S2.

(B) Comparative population genomics showing diversity of bacteriophage mutations. Symbols show a map of SNPs present at a frequency higher than 5% in P10 populations after 24 days *in vitro* (IV), in conventional mice (CM) an in dixenic mice (DM), in the presence of one strain (1S), LF82, or two strains (2S), LF82 and MG1655. Two adapted (ad_) populations were sequenced separately, only one is shown here for clarity. These showed the highest frequency of mutations, mainly gathered in the tail fiber genes *gp90* and *gp91*. For a full list of SNPs and positions in all populations refer to table S3. Relevant genes names are indicated along with functional clusters (NA, nucleic acids) whose positions have been deduced from genes annotation.

(C) Mutation Y284H in gene *gp91* (position 55079; highlighted in orange) is the only genomic modification common to all ad_P10 clones. The figure shows a genomic map of bacteriophage P10 with enlargement of the terminal region of the Gp91 protein. The alignment of sequences from P10 and 20 isolated ad_P10 (numbered 1 to 20) is deduced by PCR amplification of the genomic region encompassing all amino-acid substitutions (highlighted in pink and yellow) found within this protein.

Figure 2. Broad host range variations are associated to bacteriophage adaptation.

Ad_P10 (n=16) were tested for infectivity (light blue) versus 105 Enterobacteriaceae (see Table S1) and compared to the ancestral P10 bacteriophage (purple). The shared acquired infectivity towards strain MG1655 has a very variable effect on the global host range. Only positive results are shown for clarity.

Figure 3. Homologous intra genomic recombination (hIGR) events are widespread in bacteriophage tail fibers.

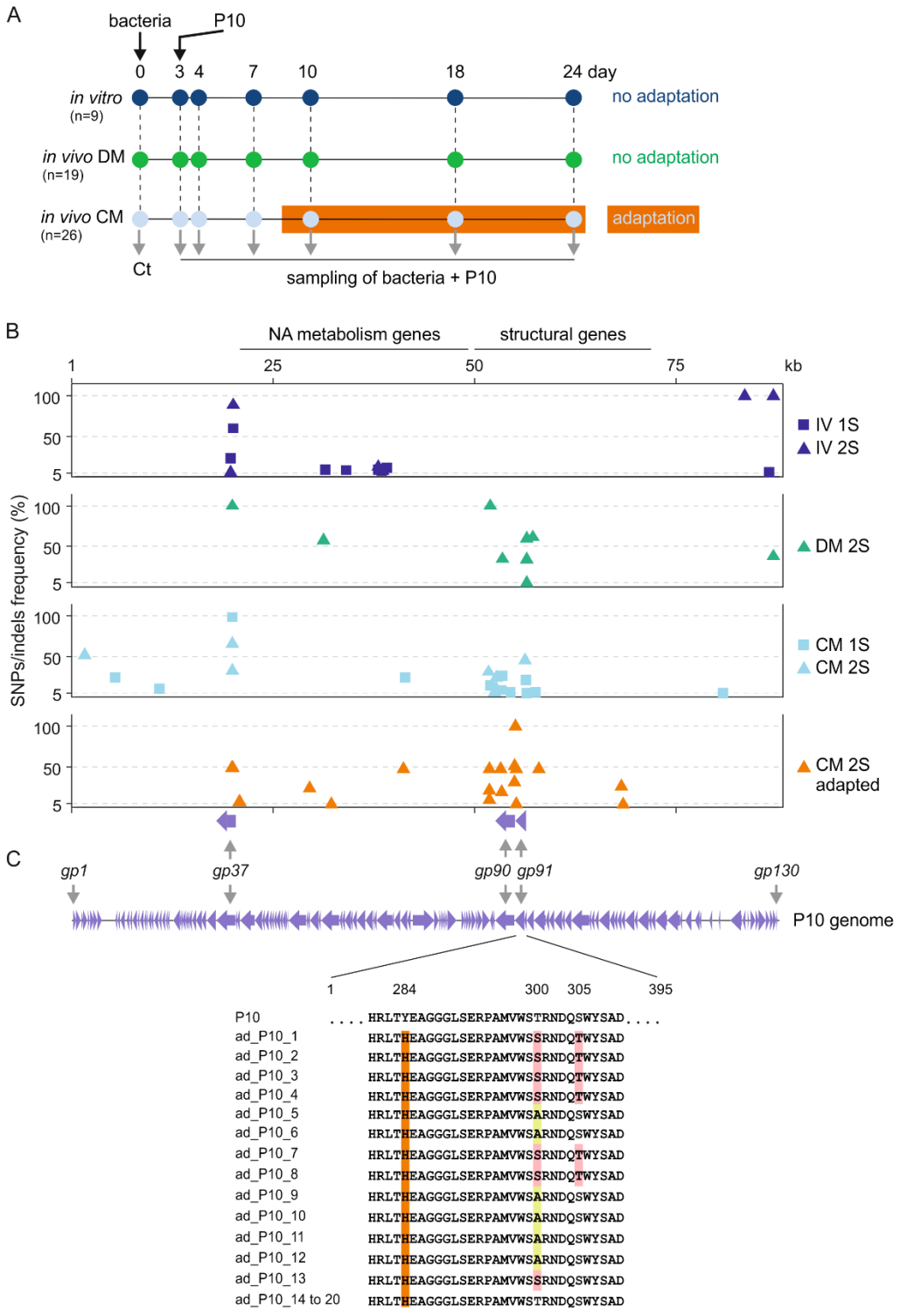
(A) Schematic representation of the interspersed duplication (red arrow) of the hypervariable region of *gp91* (red box) in *gp90* coupled by the deletion of the corresponding region in *gp90* (orange box) in ad_P10 compared to the wild type P10. Sequences of recombination sites (rs) of *gp90* (black) and *gp91* (green) are indicated as well as their location (black arrows). (B) Examples of hIGR from annotated tail fibers of other bacteriophages with duplicated regions shown in yellow (for a full list of local alignments see Table S2).

Figure 4 Intermediate hosts from the microbiota alter bacteriophage infectivity.

(A-B) The intermediate host MEc1 is sufficient to promote P10 host-jump *in vitro*. Data are shown as titers or P10 bacteriophages recovered at day 7 (A) or day 24 (B) following their incubation in flasks in presence of two strains LF82 + MG1655 (black), or LF82 + MEc1 (blue), or MEc1 + MG1655 (green) or the three of them (red) (see methods). Samples were enumerated on three bacterial hosts, LF82 (circles), MEc1 (square) and MG1655 (triangle). Adaptation of P10 to infect MG1655 only happened when MEc1 is in the mix. The initial dose of

464 bacteriophage was $2*10^6$ pfu added to a three-days-old bacterial culture (see methods). Scales
465 are logarithmic on the base 10. Bars show the mean of three independent experiments ± s.e.m.
466 (C) Schematic representation of a theoretical fitness landscape showing the dynamic
467 fluctuations in bacteriophage infectious populations relative to host abundance. Peaks
468 represent bacteriophage sub-populations with fitness shown as a coloured gradient reflecting
469 the viral infectivity and consecutive abundance. Increased bacterial diversity impacts on viral
470 population sizes and diversity, favouring host jumps (black arrow).
471
472

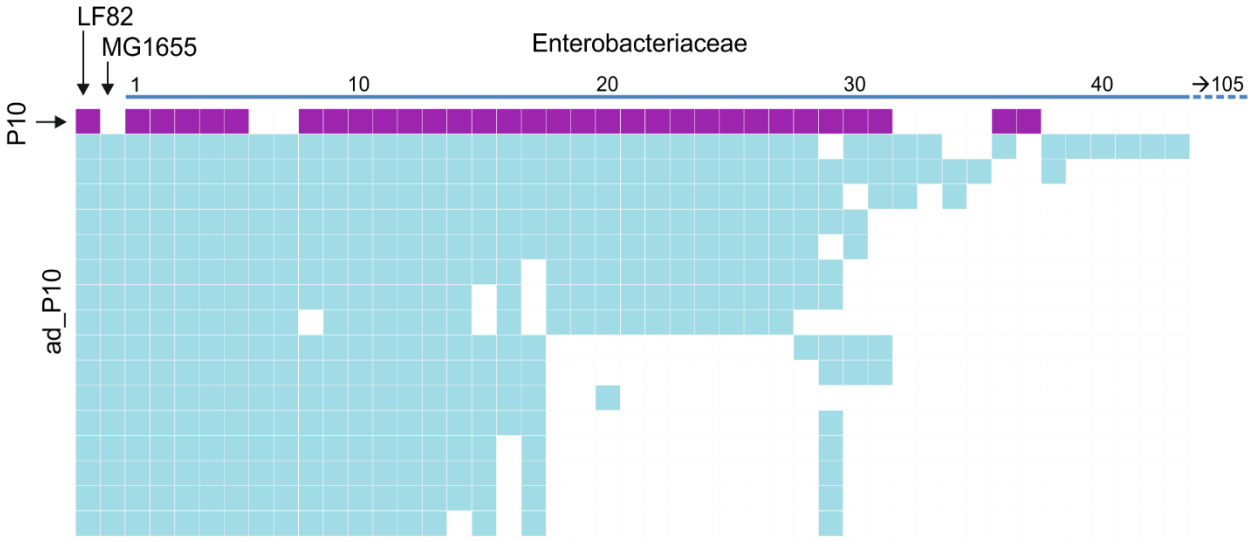
473 Figure 1



474

475

476 Figure 2



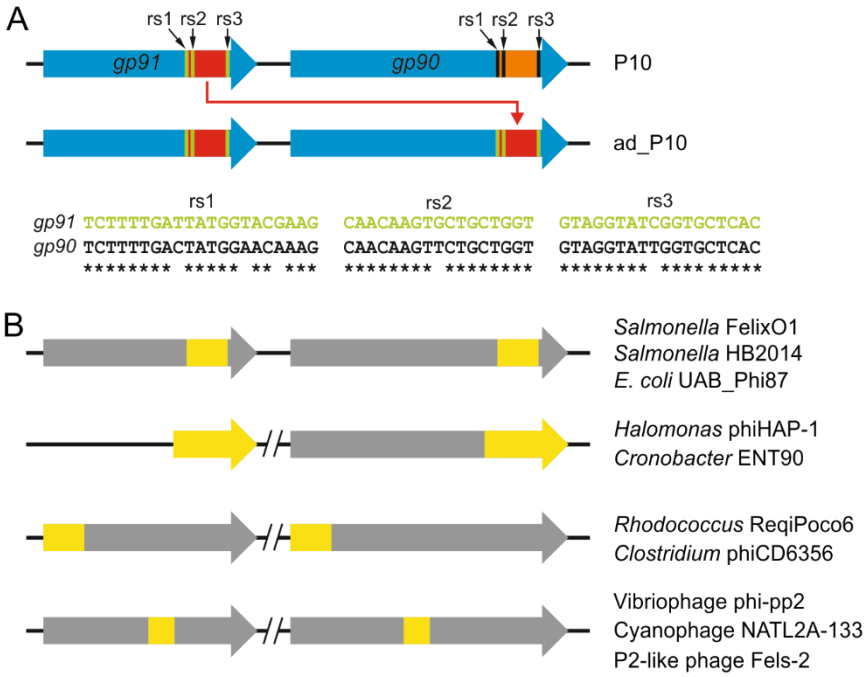
477

478

479

480

481 Figure 3

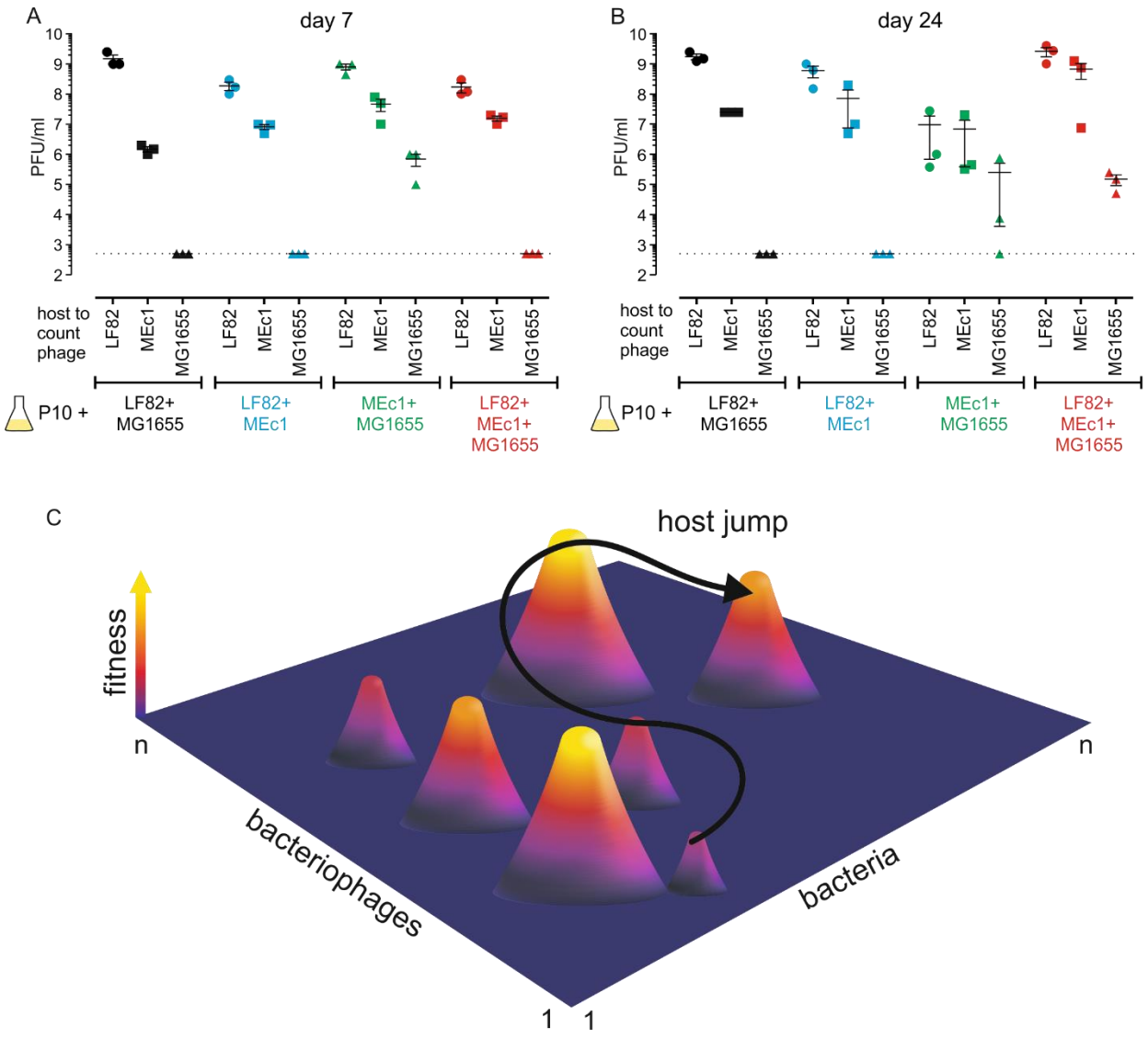


482

483

484

485 Figure 4



486

487

488

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Laurent Debarbieux (laurent.debarbieux@pasteur.fr)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Microorganisms

Bacteriophage P10 (its full name is LF82_P10 – study accession number PRJEB18073) was isolated and purified from wastewater, with an enrichment technique (Debarbieux *et al.*, 2010) using *E. coli* strain LF82. Adsorption of P10 to strains LF82 and MG1655 was measured by adding 2 x 10^5 pfu to exponentially growing cultures (OD_{600}=0.3) and redrawing aliquots every 60 seconds during 10 minutes to titer the unabsorbed bacteriophage (Hyman and Abedon, 2009). To distinguish both *E. coli* strains we introduced two different antibiotic and fluorescent markers. Strain LF82S (Galtier *et al.*, 2016) was disrupted at the *ampC* locus by insertion of a DsRed fluorescence marker gene along with a kanamycin cassette by lambda Red mediated homologous recombination to produce strain LF82 DsRed, named LF82 in this manuscript. Ampicillin resistance and GFP marker genes were P1 transduced from *E. coli* MG1655 *λatt-ampgfp* (Da Re *et al.*, 2007) into strain *E. coli* MG1655 resistant to streptomycin (Da Re *et al.*, 2013) producing strain MG1655 GFP (named MG1655 in this manuscript). The techniques used have been described in detail elsewhere (Datsenko and Wanner, 2000; Thomason *et al.*, 2007). Strains were routinely cultured in lysogeny broth (LB), or on LB agar or Drigalski agar (Bio-Rad,

Hercules, CA) plates, at 37°C. When required, streptomycin (100 µg/mL), kanamycin (50 µg/mL) or ampicillin (50 µg/mL) (Sigma, St. Louis, MO) was added.

Animals and Ethics

BALB/cYJ mice (seven-week-old females) supplied by Charles River Laboratories and CHO axenic mice bred at the Institut Pasteur (Paris, France) were housed in an animal facility in accordance with Institut Pasteur guidelines and European recommendations. Groups of co-housed mice are indicated in the method details section. Food and drinking water were provided ad libitum. Protocols were approved by the veterinary staff of the Institut Pasteur animal facility (Ref. #10.565) and the National Ethics Committee (Ref. #2015-0040).

METHOD DETAILS

Experimental conditions of coevolution

Bacteriophage P10, strains LF82 and MG1655 were mixed in planktonic cultures (*in vitro* setting) or introduced in mice (*in vivo* settings) as indicated in Figure 1A and as detailed below.

In vitro

Approximately 10^8 cfu of each bacterial strains grown to stationary phase were inoculated both individually and combined, in 1ml of LB supplemented with streptomycin (50 µg/µL) incubated statically at 37°C. Three days later 2 x 10^6 pfu of P10 were added to the mix. Every seven days bacterial and bacteriophage titers and infectivity towards both initial bacterial strains were obtained from 100 µL of the sample and cultures were refreshed starting from a 10-fold

dilution in fresh medium. Three technical replicates of three independent experiments were performed.

In vivo

Ten groups of 1 to 5 mice for each *in vivo* model (dixenic and conventional mice) were force-fed strains LF82 and MG1655 (approximately 10^8cfu of each strain per mouse, grown to stationary phase), followed, three days later, by a single bacteriophage administration (2 x 10^6 pfu per mouse). In conventional mice, streptomycin (5 g/L) was added to drinking water from three days before bacterial administration when it was lowered to 0.5 g/L and renewed weekly. Absence of detectable endogenous *E. coli* and bacteriophages infecting LF82 or MG1655 was confirmed prior to bacteriophage administration by plaque assay. A control group consisted in 5 mice caged together and receiving only streptomycin. Whenever a mouse displayed an adaptation event, those sharing the same cage also carried such bacteriophage variants. Due to likeliness of coprophagy in these animals, each cage was considered to have hosted a single adaptation event.

Bacteriophage P10 was also independently mixed with strain LF82 or MG1655 *in vitro* and in conventional mice during 24 days and no plaques were ever recovered on lawns of MG1655.

<u>Bacteria and bacteriophage quantification</u>

Bacterial and bacteriophage samples from liquid cultures or freshly collected fecal samples were weighed and homogenized in PBS, serially diluted in PBS and 4 µL of each dilution were spotted onto LB or Drigalski agar supplemented with streptomycin and kanamycin or ampicillin for the quantification of strains LF82 or MG1655, respectively. Bacteriophage titers were

553 determined by spotting the same dilutions onto LB agar plates covered with a bacterial lawn of
554 the host strain and supplemented with the appropriate antibiotics. The plates were incubated
555 for 18 hrs at 37°C. The detection limit was 5 x 10^2 cfu or pfu per millilitre of culture or gram of
556 faeces.
557
558 Host range of adapted bacteriophages
559 Ad_P10 (n=16) were tested for infectivity, as described above, against different *E. coli* strains
560 (n=105) listed in Table S1 and compared to the profile of wild type P10.
561
562 Analysis of genomic sequences
563 Sequencing of bacteriophages and bacteriophage populations (study accession number
564 PRJEB18073) was performed using Illumina sequencing technology (Illumina Inc., San Diego,
565 CA). Bacteriophage DNA was extracted from a sterile bacteriophage solution, using DNase and
566 RNase pretreatments followed by a phenol-chloroform extraction, modified from Pickard
567 (Pickard, 2009). For bacteriophage sequence analysis, the quality of Illumina reads was
568 visualised by FastQC v0.10.1 Brabraham Bioinformatics
569 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Bacteriophage assembly was
570 performed using a workflow implemented in Galaxy-Institut Pasteur using clc_assembler v4.4.2
571 and clc_mapper v4.4.2 (CLC Bio, Qiagen). P10 annotation was performed by the RAST v2.0
572 server (Aziz *et al.*, 2008) followed by manual curation. For synteny analysis, we aligned all the
573 contigs of phage genomes using blastn v2.2.18 (Altschul *et al.*, 1990) (with option –F F) against
574 the sequence of P10. Artemis Comparison Tool (Carver *et al.*, 2005) was used to visualize the

synteny and for detection of recombination. No other bacteriophage than P10 was ever purified and sequenced from amplification on strains LF82 or MG1655.

For researching hIGR events in other bacteriophage genomes, we downloaded the annotation of all bacteriophages available on PhAnToMe (http://www.phantome.org/) (September 2016) and looked for homologous regions in all annotated tail fiber sequences within each bacteriophage genome using megablast blastn v2.2.18 (with option –F F). Genome mutations were identified using the BRESEQ Variant Report - v0.26 pipeline (Barrick *et al.*, 2009) and the annotation of identified variants was processed using SnpEff v4.1 (Cingolani *et al.*, 2012).

Bacteriophage recombination

An adenine to guanidine substitution at nucleotide position 55079 of P10 genome was introduced in the *gp91* gene by site-directed mutagenesis using Polymerase Chain Reaction (PCR). The mutated gene was cloned as an *Eco*RI-*Xba*I fragment in plasmid pUC18 (Norrander *et al*., 1983) and transformed into LF82. P10 was added to a culture of LF82, carrying such plasmid (or an empty plasmid as control), in early logarithmic growth phase, at an M.O.I. of 0.1, co-growing with strain MG1655 in equal proportions to enrich the recombinant population. A lysate recovered after 4 hrs incubation at 37°C was plated on a lawn of either strain LF82 or MG1655 to recover ad_P10 plaques. Presence of the mutation was confirmed by PCR in 20 plaques isolated on MG1655 and following three rounds of purification. No plaques on MG1655 were observed for the control group.

Bacteriophage fitness test

The infection efficiency of the recombinant P10 bacteriophage (rP10) carrying the Y284H mutation in the Gp91 protein compared to the wild type P10 bacteriophage was measured as follows. First, rP10 was replicated on the MG1655 strain during four cycles of four hours at multiplicity of infection (M.O.I.) = 0.001. The resulting bacteriophage lysate was named rP10 [MG]. Then, P10, rP10 and rP10 [MG] were added to exponentially growing cultures at OD_{600}=0.2 of either strains LF82 + MG1655, LF82 + MEc1 or MEc1 + MG1655 mixed in equal proportions and at M.O.I. = 0.01. A negative control included the same amount of bacteriophages mixed to growing medium only. Co-cultures were incubated at 37°C statically during 4 hrs and bacteriophages were titered on each of the bacterial strains. Data were normalized as coefficient of multiplication compared to the initial bacteriophage inoculum. Experiments were independently replicated three times and means were compared by unpaired, one-tailed, t-test.

MEc1 isolation and sequencing

Frozen feacal samples from streptomycin-treated control mice (not exposed to P10 or LF82 or MG1655) were serially diluted and plated aerobically onto Drigalski plates to select for possible Enterobacteriaceae. Isolated positive colonies were replicated in liquid cultures to approximate OD_{600}=0.2, and 6 µL of culture were spotted onto agar plates and exposed to 1 x 10^5 pfu of P10 in a 4µL double spot assay (Saussereau *et al.*, 2014). MEc1 was chosen as representative of 50 bacteriophage-sensitive colonies out of 60. Bacteriophage-bacteria co-culture was performed as detailed above in the experimental conditions of coevolution. Genomic sequencing (study

618 accession number PRJEB21810) was performed using Illumina technology (Illumina Inc., San
619 Diego, CA) and analyzed as above.
620
621 **QUANTIFICATION AND STATISTICAL ANALYSIS**
622 Statistical analysis
623 Statistical analyses (Mann Whitney test and Student t-test) were conducted with Prism 5
624 software (GraphPad). $p < 0.05$ was considered statistically significant.
625 Sample size was determined by the "resource equation method" (Mead, 1988).
626
627 **DATA AND SOFTWARE AVAILABILITY**
628 The genomic sequences of bacteriophage P10, ad_P10 and rP10 as well as bacteriophage
629 populations have been deposited in the ENA under study accession number PRJEB18073.
630 The genomic sequence of strain MEc1 has been deposited in the ENA under study accession
631 number PRJEB21810.
632