bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Competition and coevolution drive the evolution and the diversification of CRISPR immunity

Martin Guillemet[a], Hélène Chabas[b], Antoine Nicot[a], François Gatchich[a], Enrique Ortega-Abboud[a], Cornelia Buus[c], Lotte Hindhede[c], Geneviève M. Rousseau[d,e,f], Thomas Bataillon[c], Sylvain Moineau[d,e,f], Sylvain Gandon[a]

[a] CEFE, CNRS, Univ Montpellier, EPHE, IRD, Montpellier, France
[b] Institute of Integrative Biology, Department for Environmental System Science, ETH Zurich, 8092 Zurich, Switzerland
[c] Bioinformatics Research Centre, Aarhus University, Aarhus C, Denmark
[d] Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec City, G1V 0A6, Canada
[e] Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec City, G1V 0A6, Canada
[f] Félix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Québec City, G1V 0A6, Canada

November 12, 2021

Abstract

The diversity of resistance fuels host adaptation to infectious diseases and challenges the ability of pathogens to exploit host populations [1–3]. Yet, how this host diversity evolves over time remains unclear because it depends on the interplay between intraspecific competition and coevolution with pathogens. Here we study the effect of a coevolving phage population on the diversification of bacterial CRISPR immunity across space and time. We demonstrate that the negative-frequency-dependent selection generated by coevolution is a powerful force that maintains host resistance diversity and selects for new resistance mutations in the host. We also find that host evolution is driven by asymmetries in competitive abilities among different host genotypes. Even if the fittest host genotypes are targeted preferentially by the evolving phages they often escape extinctions through the acquisition of new CRISPR immunity. Together, these fluctuating selective pressures maintain diversity, but not by preserving the pre-existing host composition. Instead, we repeatedly observe the introduction of new resistance genotypes stemming from the fittest hosts in each population. These results highlight the importance of competition on the transient dynamics of host-pathogen coevolution.

1 Introduction

Coevolution is thought to be a powerful evolutionary force at the origin of biological diversity [4–6]. The negative-frequency-dependent selection generated by coevolution can promote the emergence and the maintenance of genetic diversity in interacting species [5, 7, 8]. On the other hand, genetic polymorphism is also affected by intrinsic differences in competitive abilities among genotypes. If this asymmetric competition is strong it can lead to the exclusion of less competitive genotypes and a drop in diversity. The interplay between coevolution and competition has been explored theoretically with models based on the "kill-the-winner" hypothesis which explicitly accounts for the influence of phage predation on diverse host communities [9–11]. This framework, however, is meant to describe the ecological dynamics of interacting species but does not consider the evolutionary dynamics occuring within species. Indeed, studying the interplay between competition among host genotypes and

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coevolution with pathogens is particularly challenging at the species level because it requires a good understanding of the genetic determinants of the interaction between the host and the pathogen [12].

Here, we track the coevolutionary dynamics of CRISPR immunity of the bacterial species *Streptococcus thermophilus* with its lytic phage 2972. This model system offers unique opportunities to explore the microevolutionary processes driven by competition among different bacteria and antagonistic coevolution between bacteria and their viral pathogens. In *S. thermophilus*, coevolution is mainly driven by two (type II-A) CRISPR-Cas loci (CR1 and CR3) which allow the bacteria to incorporate 30-bp DNA sequences (spacers) from the genome of an infecting phage in the CRISPR array [13–15]. After transcription, each spacer RNA is used as a guide by Cas9 to target and cleave the corresponding target sequence in the phage genome (i.e. the protospacer), thereby halting virus replication and reducing its titer. Phages can escape this immunity with mutations in the protospacers which avert recognition by the Cas complex. These mutations have been shown to be particularly effective at escaping immunity when they are located at specific positions in the protospacers like the PAM (protospacer-adjacent motif) or the seed [16]. Crucially, the sequencing of the CRISPR array of the populations of bacteria and the whole-genome sequencing of the populations of phages allowed us to fully characterise the infection network, without any phenotypic assays. Here we focus on the CRISPR array of the CR1 locus which has been shown to be the most active of the CRISPR loci of *S. thermophilus* against phage 2972 [14].

To show how host diversity affects the dynamics of CRISPR immunity, we started cultures with a mix of 17 different bacterial strains in equal frequencies: one strain was fully sensitive to the wild-type lytic phage 2972 and the remaining sixteen strains carried each a distinct single-spacer resistance in the CRISPR array at the CR1 locus (Table S1). These sixteen different strains are resistant to the wild-type phage 2972 but mutations in the phage protospacers may allow the virus to escape CRISPR immunity [17]. We also know from a previous study that there is a significant variation of durability in resistance among these sixteen resistant strains [18] likely due to differences in rate of acquisition of escape mutations in the phage. We designed a short-term coevolution experiment where we transferred 1% of each replicate culture to a fresh medium for 4 consecutive days and we followed the evolution of CRISPR immunity in the presence or in the absence of phages (Figure 1). In the absence of phage (treatment A), the change in the relative frequency of the different host genotypes informed us about the competitive abilities of the 17 different strains of bacteria. The two other treatments allowed us to follow the interplay between competition and phage predation on the evolution of the bacteria. At the beginning of each transfer we added 10^5 phages from a monomorphic or a polymorphic phage populations (treatments B and C, respectively). The monomorphic phage population was obtained from the amplification of the wild-type phage 2972 which infects only the sensitive host strain (about 6% of the host population at the onset of our experiment). In the polymorphic phage population, we used a mix of sixteen escape phage variants (phage cocktail) that were previously selected to escape each of the sixteen CRISPR CR1 resistances of the polymorphic population of bacteria [18] (Table S2). This recurrent inoculation of phages at each transfer was used to maintain a minimal amount of phage predation in all the treatments with phages. As pointed out below, this immigration of phages did not prevent coevolution and the adaptation of phages.

To monitor the demography and evolution of bacteria we used spacers as barcodes and sequenced the 5'-end of the CRISPR array of the CR1 locus of the bacteria (see Methods). This sequencing strategy allowed us to identify the emergence and the spread of additional resistance strains with new spacers in the CRISPR array [19]. To monitor the evolution of the phage populations we used whole genome sequencing in the two experimental treatments exposed to the virus to identify new mutations and estimate their frequencies.

2 Results and Discussion

2.1 Phage diversity drives infection dynamics

The experimental treatments had major effects on both the bacteria and the phage densities (Figure 2). The monomorphic phage treatment had a limited impact on bacterial growth the first day but led to a massive phage epidemic on the second day, marked by a drop in host density and an increase in the viral pathogen density. In contrast, the polymorphic phage treatment immediately led to a viral outbreak on the first day. Yet, under all phage treatments the bacterial populations eventually

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

recovered and by day 4 they reached a density close to the no-phage treatment (Figure 2a).

2.2 Evolution and diversification of CRISPR immunity

To monitor the evolutionary dynamics of bacteria, we tracked the diversity of CRISPR immunity at the CRISPR CR1 locus and estimated the frequency h_i of each resistance genotype i in the population. We used these frequencies to compute the effective number of genotypes [20] across time for each replicate (Figure 3). Strikingly, the effective number of genotypes dropped very fast in the treatment without phage and remained very low until the end of the experiment. Exposition to a monomorphic phage population initially led to a faster drop in diversity after 1 day (effective number of genotype with 95% CI: 6.91 ± 0.17), but the exposition to a polymorphic phage treatment maintained a high level of diversity (14.88 ± 0.24). Both phage treatments led to the maintenance of some diversity at the end of the experiment (monomorphic phage: 3.66 ± 0.87; polymorphic phage: 5.33 ± 1.60). The maintenance of host diversity in the treatment exposed to phages supports the idea that coevolution can drive the diversification of host populations [2, 4, 10, 21]. The variation in the dynamics of diversity among replicate populations exposed to phages, illustrates the impact of demographic stochasticity on this coevolutionary dynamics, particularly after demographic bottlenecks caused by viral epidemics.

Next, to better understand what drives the dynamics of CRISPR diversity we examined the competition between the different bacterial strains using modified Muller plots that provide a description of both the change in density and in the genetic composition for each replicate population of bacteria (Figure 4). The colors used in Figure 4 refer to the competitive ability of each resistant strain obtained from the change in strain frequency at the end of day 1 in the absence of phage (Figure S1). All the replicates followed very similar dynamics in the treatment without phage (Figure 4): one of the strain (indicated in red, strain 31725) outcompeted the other bacteria and nearly reached fixation by day 2, but another strain (indicated in green, strain 16236) increased in frequency towards the end of the experiment. These results indicate major differences in competitive abilities among strains. Interestingly, the fitter strain (in red) is not the phage-sensitive wild-type strain but one of the sixteen resistant strains (Figure S1). Whole genome sequencing of the seventeen strains used at the beginning of the experiment revealed the existence of numerous other mutations across the bacterial genome outside the CRISPR locus (Table S3). These mutations were acquired during the selection process that led to the natural acquisition of a new spacer on the CR1 locus [22]. For instance, the "red" strain has nine unique non-synonymous mutations. In contrast, the sequencing of the "green" strain revealed only two unique, but synonymous mutations. The competitive ability of the "green" strain is also more puzzling because this strain was initially less fit and only increased in frequency towards the end of the experiment. A more detailed analysis of the contribution of each of these mutations on the competitive ability of the strains falls beyond the scope of this study. But these highly consistent measures of fitness among replicates in the treatment without phage allowed us to study how competition affects the coevolutionary dynamics in the populations exposed to phage predation.

Figure 4 shows how phages affect both the density of bacteria and the evolution of CRISPR immunity. As expected from Figure 3, phage predation maintains a higher number of strains. More specifically, we observe the emergence of several new resistant strains that carry up to three additional spacers in the CRISPR array, which are indicated by dark colors in Figure 4. Interestingly, in all treatments exposed to phage predation, almost all the bacterial populations end up being dominated by lineages that are descendants of the two most competitive strains identified in the absence of phages (Figure S2). In other words, the increase in diversity observed at the end of the experiment in the populations exposed to phages (Figure 3) is not due to the initial diversity being restored, but to new resistance genotypes that arose via the acquisition of new spacers in the CRISPR array of the winners of the competition among bacterial strains (Figure S3). Indeed, as these bacteria were the most abundant, they were also more likely to acquire new spacers.

Importantly, the comparison among replicate populations revealed very different dynamics in the presence of phages. To study this variation, we measured the amount of genetic differentiation among replicate populations within each treatment (Figure S4). Complementary measures of host differentiation (F_{ST} and D) allowed us to quantify the changes in population composition due to drift and selection among replicates (see Methods). As expected, differentiation remained very low in the treatment without phages because all replicates displayed very similar dynamics. In contrast, exposition to phages led to the acquisition of distinct spacers in different replicates, which led to a rapid increase in differentiation among host populations. This is particularly true right after the massive demo-

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

graphic bottleneck that took place after the first day in the treatment exposed to a polymorphic phage population.

Another way to demonstrate the influence of phages on bacterial evolution is to detect the presence of negative-frequency-dependent selection (NFDS). As expected, in the absence of phages the change in strain frequency between time t and $t+1$ is independent of strain frequency at time t (Figure 5a). Exposition to phage predation, however, yields a strongly negative relationship between these two quantities (phage predation has a highly significant effect on the slope of the regression line in both the monomorphic and the polymorphic phage treatments, see Methods), which indicates that more frequent strains tend to be selected against because they are preferentially targeted by phages (Figures 5b and 5c). All these results confirm the expected impact of pathogens on the diversification of host resistance [21, 23] and highlight the relevance of the "kill-the-winner" hypothesis [9, 10].

2.3 Phage coevolution across space and time

The sequencing of the phage populations revealed the emergence and the spread of many mutations across the phage genome (Figure S5). Most of these mutations were located in the protospacer regions targeted by CRISPR immunity and particularly in the PAM or the seed of protospacers (Figure S6). These mutations are expected to be strongly beneficial as they allow the virus to escape CRISPR immunity [17]. Knowing the genetic specificity of CRISPR immunity allows us to assign phenotypic effects to these mutations without any additional experimental measures. We combined sequencing data from the bacteria and the phages to compute the mean fitness of each phage population using:

$$\overline{w} = \sum_{i=1}^{n} h_i p_i \tag{1}$$

where n is the total number of host strains, h_i is the frequency of host strain i and p_i is the frequency of phage variants that can infect strain i. Here, the mean fitness measures the mean fraction of the host population available to a randomly sampled virus in the phage population. This *in silico* measure of phage mean fitness provides a powerful way to estimate phage adaptation to contemporaneous host populations (when phage and bacteria frequencies are sampled in the same replicate and at the same point in time) but also across space and time [6, 21].

Measures of phage adaptations across all time points revealed a striking pattern where levels of phage adaptation are maximal against host populations from the recent past (Figure 6). In contrast, the degree of phage adaptation drops very rapidly against bacteria from the future in both phage treatments. This pattern is precisely the one expected under the rapid coevolution dynamics that are predicted to emerge in coevolutionary models [6, 24–26]. The particularly rapid drop of phage mean fitness when matched against bacteria from the future shows how quickly bacteria are able to develop new resistance to the phages. This is consistent with the intrinsic asymmetry inherent to CRISPR specificity: bacteria have access to hundreds of different protospacers from the phage genome [17] allowing them to raise a diverse and distributed immune defense to the phage population at once [27]. In contrast, only mutations in the targeted viral genomic region (Figure S6) can provide an effective way to escape CRISPR immunity, and only against one resistance (one spacer) at a time.

Measures of mean fitness across space allowed us to compute phage local adaptation to determine if the phage is more adapted to sympatric (same replicate) than to allopatric (different replicate) host populations (see Methods) [24, 28]. Figure 6 shows the buildup of local adaptation across time in the two phage treatments. Interestingly, local adaptation remains very low in the treatment with the monomorphic phage population. In contrast, we detect a strong pattern of phage local adaptation in the treatment with a polymorphic phage population. In particular, phage local adaptation is extremely strong ($0.53[\pm 0.18]$) at day 2 which coincides with the time at which host differentiation is maximal. Indeed, phage local adaptation can only occur when the composition of sympatric host populations differ substantially from allopatric host populations.

The dynamics of differentiation varied between the two phage treatments. Even if we detect significant differentiation among replicate populations in the treatment with the monomorphic phage population, the Q_{ST} which measures phenotypic differentiation (see Methods) remains very low because most escape mutations occur on the same protospacer (Figure S7). In contrast, in the treatment with a polymorphic phage population, both F_{ST} and Q_{ST} are increasing on day 2 after the divergence of bacterial populations. Note, however, that the speed of phage adaptation seems too low to catch up

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with the build up of CRISPR immunity. Initial diversity in the polymorphic treatment yields faster adaptation but the acquisition of new mutations in protospacers stops by day 1 (Figure S8). In the monomorphic treatment, the saturation in the acquisition of new mutations in protospacers is delayed but it also stops by the end of the experiment (Figure S8). The drop in local adaptation with time is consistent with the overall drop in phage density we observed in most phage populations (Figure 2b). This suggests that the phages are losing the coevolutionary arms race with their hosts which is in line with previous studies showing that CRISPR immunity often yields phage extinction in this system [29–31]. Besides, we found some evidence that evolution of new immunities may also be due to the second active CRISPR-Cas system (CR3) in this host in which we detected spacer acquisition starting at day 3 or 4 (Table S4). Accounting for evolution at the CR3 locus when estimating phage fitness magnifies the drop of mean fitness of phage populations (Figure S9).

2.4 Host competition governs the coevolution-driven diversification

We can track the dynamics of phage adaptation across space and time but can we predict the speed at which the phage escapes the phage-resistant strains? The speed of adaptation is governed (i) by the rate of mutation, which has been shown to vary among protospacers in a previous experiment [18], (ii) by the strength of selection associated with the ability to escape CRISPR immunity against a specific protospacer and (iii) by the fitness cost of these escape mutations. Because the fitness cost of these mutations has been shown to be a poor predictor of the durability of CRISPR-Cas immunity [18] we focus on the first two points. In the treatment with a polymorphic phage population, the rate of mutation is not limiting because the mutations against the 16 original spacers are pre-existing. In this phage treatment, as expected, we do not find a correlation between the speed of phage adaptation and the rate of escape mutation for different protospacers (Pearson's $r = -0.26$, $P = 0.41$) (Figure S10). In contrast, the speed of adaptation is governed by the competitive ability of the different resistant strains (Pearson's $r = 0.83$, $P = 7.5\text{e-}5$). Indeed, this competitive ability is a good predictor of the abundance of each resistant strain and, consequently, a good predictor of the fitness benefit associated with the ability to exploit these resistant strains. Interestingly, we obtain a very similar pattern in the monomorphic phage treatment (no correlation with the mutation rate: Pearson's $r = -0.02$, $P = 0.41$; strong correlation with competitive ability: Pearson's $r = 0.95$, $P = 3.4\text{e-}8$). These results indicate that phage mutation is not limiting and phage adaptation is mostly driven by the more abundant (i.e. the more competitive) phage-resistant strains of bacteria.

3 Conclusion

Our short-term evolution experiment demonstrates that the coevolutionary battle taking place between bacteria and phages is a potent evolutionary force driving the rapid diversification of interacting populations. The presence of phages generates strong negative-frequency-dependent selection, which prevents the loss of diversity of CRISPR immunity. This is consistent with the "kill-the-winner" hypothesis [9, 10] which states that predation by viruses can maintain the diversity in host populations. But our works also reveals the joint influence of competition and coevolution on this diversity. Indeed, we see that in all but one replicate population exposed to phages, the bacterial population at day 4 is dominated by strains that descend from the most competitive strains (the "winners") identified in the control (Figures 3 and S9). To understand these results it is important to recall that host adaptation results from both the selection imposed by phages at the CRISPR locus and the selection imposed on the rest of the bacterial genome. The recurrent bottlenecks imposed by phage predation may lead to a faster fixation of new mutations. Even if these additional mutations are expected to be often deleterious [32], their effects on fitness will vary and introduce variation in the competitive abilities of different strains [32, 33]. In the absence of phages, fitter host genotypes outcompete other strains. In the presence of the phage, viral adaptation targets preferentially more abundant and competitive strains. But the evolution of CRISPR immunity allows the winners of the intraspecific competition to strike back after phage adaptation. Ultimately, this explains why diversity is generally maintained among the descendants of the winners in populations exposed to phages. This feedback of competition on coevolutionary dynamics can also be discussed in the light of the recent "royal family model" [34]. In a classic version of the "kill-the-winner" framework, the most frequent host strain is preferentially targeted by the evolving population of pathogens and is driven to low frequency. Next,

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

another host strain rises to high frequency and the cycle repeats. In the "royal family model" intrinsic asymmetries in competitive abilities imply that the newly rising host genotypes are likely to descend from the previously dominating genotypes. Our experiment squarely fits within this framework as we can readily identify a royal family in the bacteria population which often derives from the more competitive strains (the red and green strains in Figure 4). Note, however, that our experiment does capture the rise of a new royal family in one population after a particularly strong demographic bottleneck (replicate C3 in Figure 4). As expected from the "royal family model" this evolutionary dynamics within the population of bacteria implies that there is also a royal family of phages, which is particularly adapted to the royal family of bacteria.

Our work demonstrates that ecological and evolutionary processes can take place on a similar time scale. A better understanding of the coevolution between CRISPR immunity and the phage requires a more comprehensive theoretical framework considering the mutations involved in the interaction as well as in the rest of the genome. Current models of host-parasite coevolution neglect possible asymmetries in competitive abilities among host genotypes carrying the same number of resistance genes. However, the accumulation of mutations in loci not involved in interactions with the phages leads to a drop in the immune diversity after a local extinction of the phage population. This drop in resistance diversity is likely to facilitate the evolutionary emergence of the phages when new viruses are introduced in the population [2, 3] and this succession of local extinction and recolonisation could ensure the long-term coexistence of bacteria and phages in spatially structured environments.

4 Materials and Methods

4.1 Bacteria and bacteriophage strains

S. thermophilus DGCC 7710 and phage 2972 [35] were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (www.phage.ulaval.ca). Sixteen derivative phage-resistant strains, each with an unique CRISPR spacer were generated previously [18] and sequenced to look for mutations outside of the CRISPR loci (Table S1). Similarly, sixteen phages carrying mutation to escape the resistance of these individual spacers were isolated after selection on each resistant bacteria (see list of protospacer sequences in Table S2) [18].

4.2 Experimental procedure

Prior to the experiment, the 17 bacterial strains were mixed and grown during 6 hours in LM17+$CaCl_2$ (37 g/l of M17 (Oxoid) supplemented with 5 g/l of lactose and 10 mM of sterile $CaCl_2$). Then, the bacterial mix was transferred 1:100 into 10 ml of fresh LM17+$CaCl_2$ (no phage treatment, 7 replicates), infected with 10^5 wild-type 2972 phages (monomorphic phage treatment, 8 replicates) or infected with the mix of 10^5 phages (polymorphic phage treatment, 8 replicates), then incubated at 42°C. Every day (after 18 hours of incubation), 1% of the cultures were transferred into 10 ml of LM17+$CaCl_2$, and 10^5 phages were inoculated from the same population of phage (monomorphic or polymorphic) used at the beginning of the experiment. Following each transfer, the bacteria and phages from each replicate were separated by filtration (0.2 µm) and titrated as described in [18]. To guarantee that there was enough DNA for sequencing, the phages were reamplified once on the host bacteria overnight, then DNA was extracted using the ZYMO Quick-DNA Miniprep plus kit.

4.3 Bacteria sequencing

The CRISPR-Cas CR1 locus was amplified through PCR (primers 5'-3': AGTAAGGATTGACAAG-GACAGT; CCAATAGCTCCTCGTCATT) and sequenced with Illumina MiSeq. The spacers were extracted from the sequences by searching for the flanking repeats allowing for a maximum of one mismatch. The spacers were then matched with their protospacers on the phage genome using Blast version 2.8.1 [36] and the protospacer database presented in the next section. After these steps, an average sequencing depth of around 95700 was obtained. A minimum identical word size of 10, and a 70% identity threshold was used. The top result of the search, if any, was used to replace the name of the spacer by the middle position of the protospacer in the phage genome. A frequency cutoff of 1% was used to optimize the quality of our dataset. We found that in the treatment with the monomorphic phage population there has been significantly more acquisition of spacers that were

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

already present in the original 16 bacteria strains than the other 677 potential spacers (Chi-square test: chi2 obs=12.17, df =1, P =4.8e-4). This means that the spacers already present in the mix were acquired preferentially which may be due to DNA transfer among bacteria. The CRISPR-Cas CR3 locus was amplified through PCR (primers 5'-3': GGTGACAGTCACATCTTGTCTAAAACG; GCTGGATATTCGTATAACATGTC) and migrated on 1.5% agarose gel to check for spacer acquisition. The samples with additional bands indicating the acquisition of an additional spacer are given in Table S4.

4.4 Phage sequencing

The phage DNA samples were sequenced (Illumina MiSeq) with 150-bp paired-end reads. Trimmomatic [37] was used to clean and trim the sequencing reads yielding an average sequencing depth of around 650, before mapping them on the reference genome using Bowtie2 [38]. The software FreeBayes [39] was then used to detect single-nucleotide polymorphism and the phage reference genome [35] was updated to include the SNPs with frequency > 0.45 in the initial mix, to distinguish these pre-existing mutations from the ones that arose during the experiment. The read mapping and the SNP detection were done a second time using this updated genome as reference. To detect the protospacers in the phage genome, we looked for the CR1 specific PAM sequence 'GGAA' or 'AGAA' in both strands of this reference genome and found 693 occurrences (281 and 412 respectively for the two PAMs).

4.5 Fitness and adaptation estimates

We computed the mean phage fitness in a certain host population with equation (1). Our short-read sequencing data for the phages does not give linkage information between mutations so we need a linkage hypothesis to compute p_i from the frequencies of escape mutation frequencies derived from whole-genome sequencing of phage populations. When the host resistance genotype i carried more than a single spacer we assumed that the genotype frequency of the phage variant able to infect host resistance genotype i was the product of the frequencies of the mutations on all the protospacers targeted by this set of spacers (i.e., we assumed linkage equilibrium among these mutations). To check the robustness of our results we computed phage fitness under the alternative assumption that escape mutations are fully linked (by setting to the frequency of phage mutations providing escape to the last spacer in the CRISPR locus). We observed less than 2.7% difference between the measures of mean fitness of the phage in sympatric (same replicate) and contemporaneous (same time point) host populations under the two alternative assumptions for linkage. Hence, since linkage seems to have a limited effect in our analysis, all the results computed are derived under the assumption of no linkage.

Phage local adaptation was obtained for each replicate r at time t by computing the mean fitness of the phage on contemporaneous hosts (same time point t) from the same replicate r and by subtracting the mean fitness of the phage on contemporaneous hosts (same time point t) from all other replicates:

$$LA(r,t) = \sum_{i=1}^{n} h_i(r,t)\, p_i(r,t) - \frac{1}{n_r - 1} \sum_{j \neq r} \sum_{i=1}^{n} h_i(j,t)\, p_i(j,t) \tag{2}$$

where $h_i(r,t)$ and $p_i(r,t)$ are the frequencies of host and phage genotypes in replicate r at time t and n_r is the number of replicates per treatment. In Figure 6a we present the phage local adaptation for different values of t after averaging over the $n_r = 8$ replicates for the monomorphic and the polymorphic phage treatments. The shaded areas present the 95% confidence interval after boostraping over replicates.

Phage temporal adaptation was obtained for each replicate r at time t by computing the mean fitness of the phage on hosts from the same replicate r but at a different time point in the past or in the future (τ measures the time delay between bacteria and the phage: when $\tau > 0$ bacteria come from the future, when $\tau < 0$ bacteria come from the past). This measure was averaged over time t which yields:

$$TA(\tau,r) = \frac{1}{n_t - |\tau|} \sum_{t=max(0,-\tau)}^{min(n_t,n_t-\tau)} \sum_{i=1}^{n} h_i(r,t+\tau)\, p_i(r,t) \tag{3}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where n_t is the number of time points in the experiments, here $n_t = 5$ (i.e., 0 to 4). Note that when we average over t we have to account for the fact that the number of elements we use for this calculation varies with τ. For instance, if $\tau = 0$ there are $n_t = 5$ points we can use (i.e., the diagonal in Figure S9). In contrast, if $\tau = -4$ there is only one point (i.e., the lower right corner in Figure S9). Hence, the number of elements in the sum over time in (3) is equal to $n_t - |\tau|$. In Figure 6b we present the phage temporal adaptation for different values of τ after averaging over the $n_r = 8$ replicates for the monomorphic and the polymorphic phage treatments. The shaded areas present the 95% confidence interval after boostraping over replicates.

4.6 Differentiation measures

Jost's D for bacteria was computed on the CRISPR-Cas CR1 locus according to Jost [40] with equation:

$$D = \frac{H_T - H_S}{1 - H_S} \frac{n_r}{n_r - 1} \tag{4}$$

with n_r the number of replicates, H_T the mean heterozygosity of the pooled replicates and H_S the mean within-replicate heterozygosity, considering each different set of spacers a different genotype. Phage F_{ST}, Q_{ST} and bacteria F_{ST} was computed according to Weir and Cockerham [41] to take into account the variance in sample size. For the Q_{ST} which measures phenotypic rather than genetic differentiation, we treated phage mutations that led to the same phenotype, for example two mutations in the same protospacer, as a single phenotype.

4.7 Statistical analysis

The 95% confidence intervals displayed on the figures 6, S4, S7 and S8 were computed using a bootstrap approach, by resampling the data from the different replicates within a treatment 1000 times.

The linear regressions and the associated statistics for Figures 5 and S10 were computed using the SciPy [42] and statsmodel [43] Python packages. In Figure 5, the statistical significance of the results was assessed by comparing separately each phage treatment to the treatment without phages. For each phage treatment, we built the following linear model: $\Delta h_i(t) \sim h_i(t) * treatment$, with $\Delta h_i(t) = h_i(t+1) - h_i(t)$, including the data from that treatment and the treatment without phages. To demonstrate the presence of NFDS we tested the significativity of the interaction term in the ANOVA table of the linear model (i.e., this interaction measures the effect of phage predation on the effect of $h_i(t)$ on $\Delta h_i(t)$). This analysis confirmed the presence of NFDS induced by phage predation: the p-values associated with the interaction term were 1e-192 and 3e-267 for the monomorphic and polymorphic phage treatments, respectively.

For all differentiation estimates (Figures S4 and S8), the replicate measures necessary for confidence interval were generated with the Jackknife approach. This was done by computing the measures n_r times, each time leaving a different replicate out of the calculation [44]. The analysis and plotting was carried out using R 3.6.3 [45] and Python 3.8.5 [46].

5 Acknowledgments

Sequencing data were obtained through the genotyping and sequencing facilities of ISEM (Institut des Sciences de l'Evolution-Montpellier) and Labex Centre Méditerranéen Environnement Biodiversité. We thank Gabrielle Pageau for technical assistance. S.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada (Discovery program). S.M. holds a T1 Canada Research Chair in Bacteriophages. S.G. acknowledges support from a grant on "Phylodynamics for experimentally evolving viruses" funded by the CNRS-MITI (Mission pour les Initiatives Transverses et Interdisciplinaires) and from the grant ANR-17-CE35-0012 from the ANR (Agence National de la Recherche).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

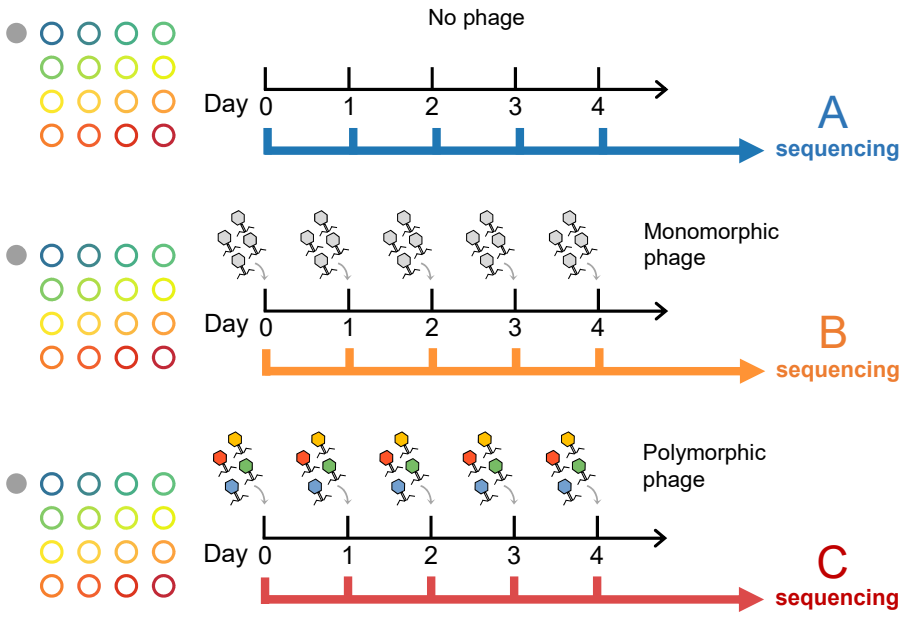

Figure 1: The three treatments of our coevolutionary experiment. Bacterial cultures were inoculated with a mix of 17 different strains in equal frequencies: one strain (grey circle) fully sensitive to the wild-type lytic phage 2972 and the remaining sixteen strains (colored circles) carrying a distinct single-spacer resistance in the CRISPR 1 (CR1) locus. (1) daily transfers of 1% of the bacterial culture with no exposition to phages, (2) daily transfers of 1% of the bacterial culture with inoculation of 10^5 phages at each transfer sampled from a monomorphic population of the wild-type phage, (3) daily transfers of 1% of the bacterial culture with inoculation of 10^5 phages at each transfer sampled from a polymorphic phage population. This polymorphic phage population is a mix of sixteen escape variants that were previously selected to escape each of the sixteen CRISPR CR1 resistance of the polymorphic population of bacteria.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

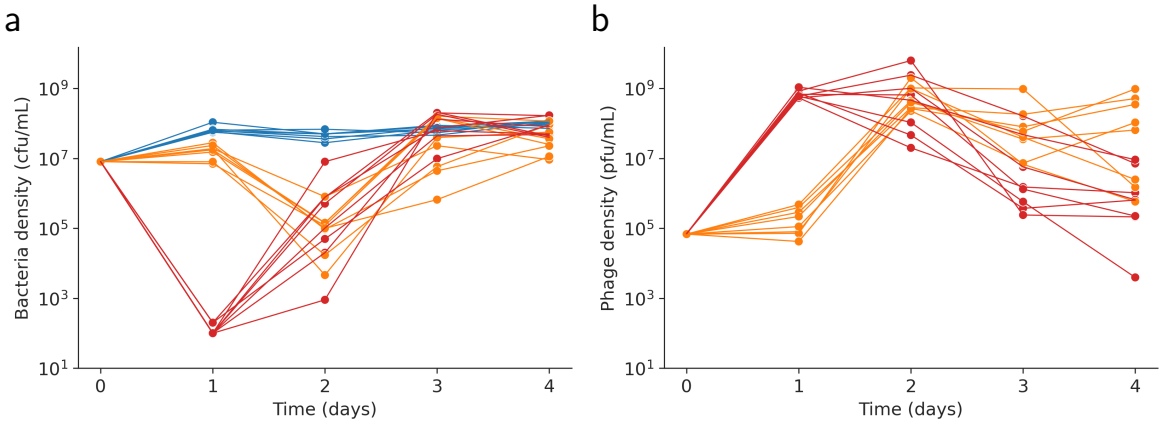

Figure 2: Demography of bacteria and phages. Titre of bacteria (a) and phages (b) in the three treatments. All replicates are shown, 7 for the control and 8 for the two phage treatments. Blue points show the data in the absence of phages, while orange and red respectively show data for the monomorphic and polymorphic phage treatments.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

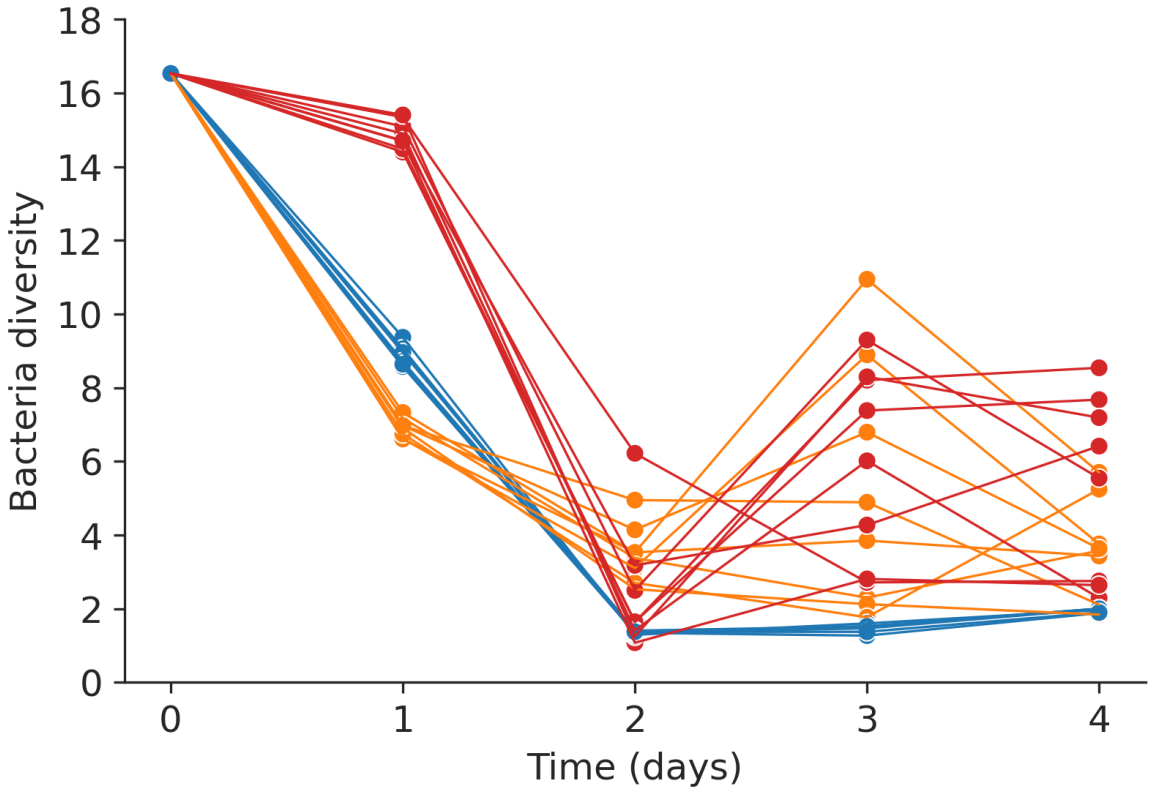

Figure 3: Dynamics of the diversity of CRISPR immunity. Dynamics of CRISPR locus diversity computed with the effective number of genotypes. Blue points show the data in the absence of phages, orange and red respectively show data for the monomorphic and polymorphic phage treatments.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

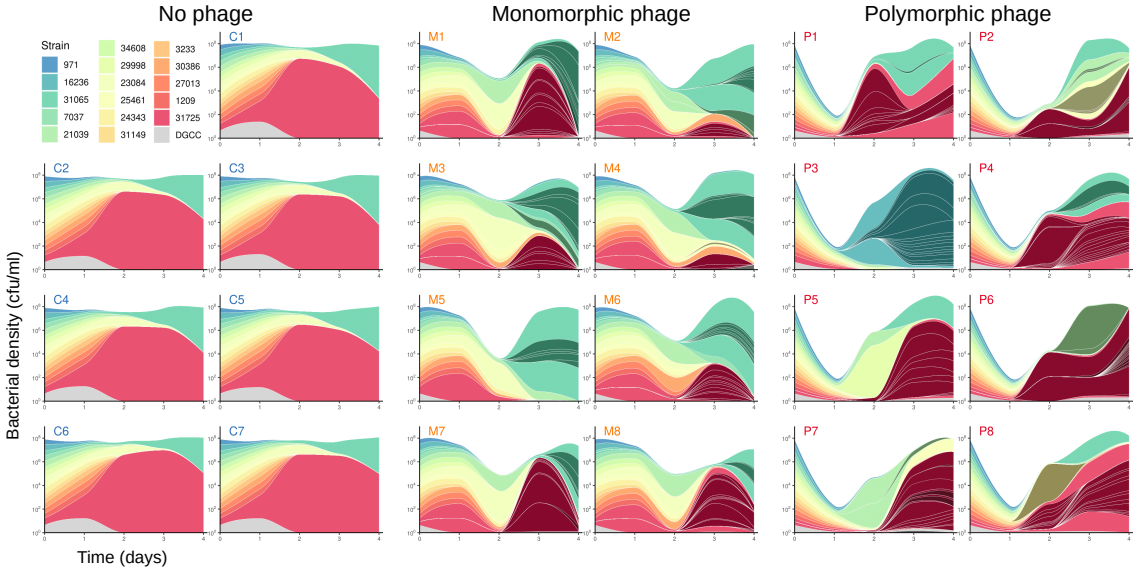

Figure 4: Host populations resist phages through the diversification of the CR1 locus. Modified Muller plots show the bacterial populations in each replicate as indicated above each graph ('A' for the no phage control, 'B' for the monomorphic phage treatment and 'C' for the polymorphic phage treatment). The total height for each day shows the bacterial density (in cfu/ml) on a log scale, and the different colors show the proportion of the strains at each time point on a linear scale. The 17 strains that were added on day 0 (including the phage sensitive in grey) are shown in the legend. The blue to red color scale used ranks the strains according their density in the control at day 1. A darker colored strain seen in later days stemming from one of the 17 original strains shows the acquisition of a new spacer. An even darker color represents strains with 2 additional spacers (2 new spacers is the maximum represented here). The lines are smoothed between each day.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

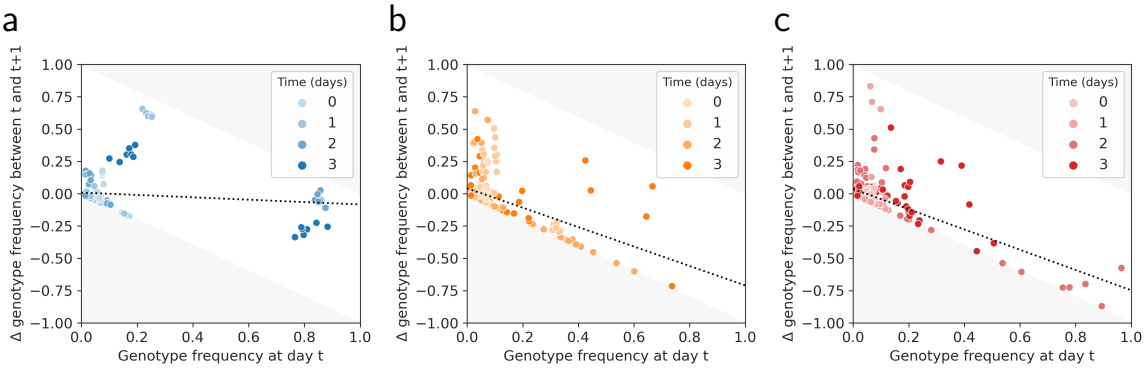

Figure 5: Phages induce negative frequency dependent selection. Evidence of Negative Frequency Dependent Selection (NFDS) in the two phage treatments (b,c) compared to the control (a). In every replicate, the change in frequency of a genotype between day t and $t+1$ is plotted according to its frequency at day t. A linear regression is plotted in each panel to highlight the NFDS (or lack thereof in the control). Indeed a significantly negative slope indicates that more frequent genotypes are counter-selected and tend to decrease in frequency the following day (see Methods). The light grey area refers to unfeasible change in frequency. Blue points show the data in the absence of phages, orange and red respectively show data for the monomorphic and polymorphic phage treatments.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

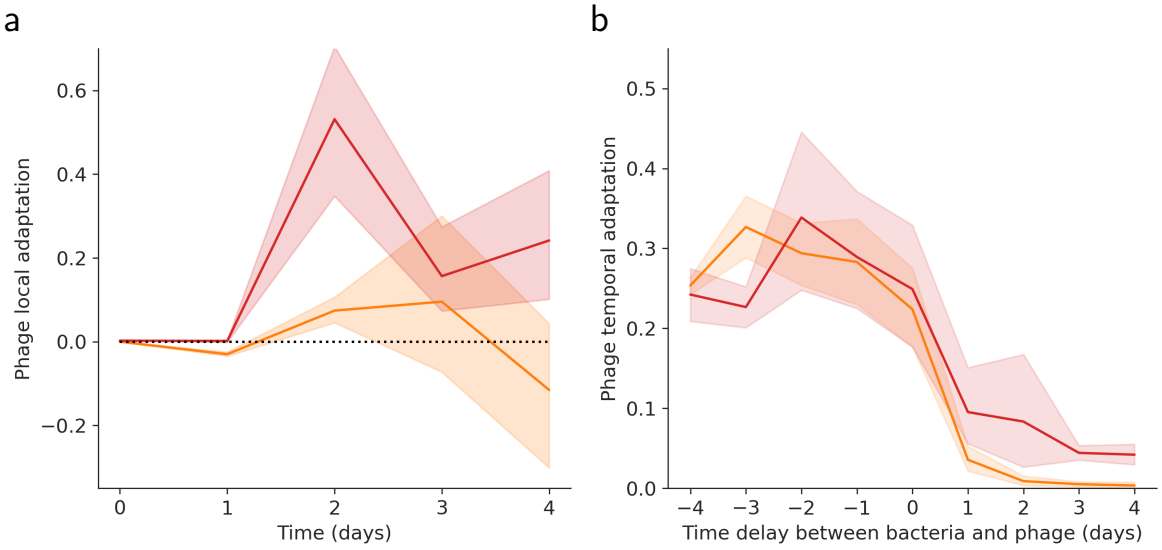

Figure 6: Phage adaptation across space and time. Evidence of local (a) and temporal adaptation (b) of the CRISPR immunity in the two phage treatments. Local adaptation is shown among the replicates of a given treatment, according to the time. Temporal adaptation shows, among a given replicate, the mean fitness of the phage populations infecting past, contemporary or future bacteria according to time difference separating phage and bacteria populations. Orange and red respectively show data for the monomorphic and polymorphic phage treatments. The colored areas show the 95% confidence interval after boostraping over the replicates (see Methods).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

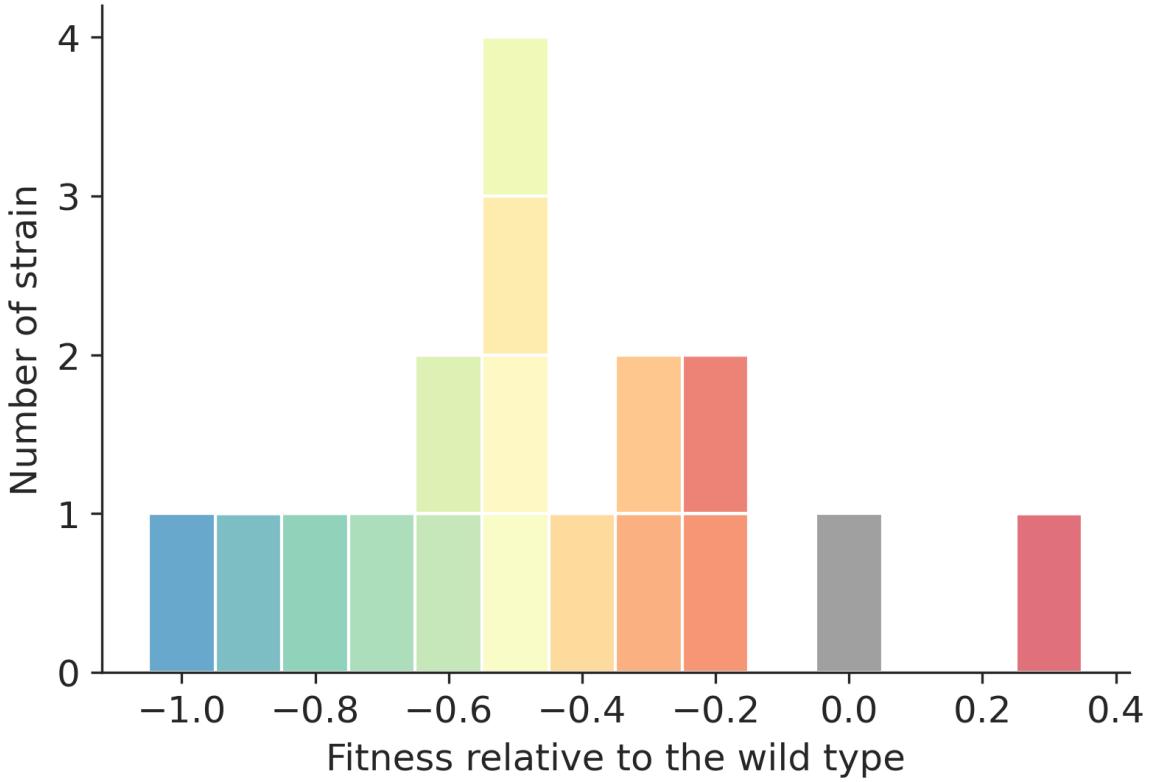

Figure S1: Fitness distribution of the 16 resistant and the wild-type bacteria strains in the absence of the phage. The wild-type bacteria is shown in grey and the colors indicate the relative fitness of each of the 16 resistant strains. The fitness of strain i (relative to the wild-type) is computed with $W_i = log_{10}\frac{f_1(1-f_0)}{f_0(1-f_1)} - W_{wild-type}$ with f_0 the frequency of the strain at day 0 et f_1 the frequency at day 1.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S1: Summary of the CR1 locus of the host strains. The strains are named according to the middle position of the protospacer which they target on the phage genome.

Host strain name	Protospacer sequence (5'-3')	Strand	Position of the protospacer in the phage genome
971	AGGAGGTGGACATATTGGGCTAAATCAACG	+	954 to 983
1209	GCTCTACGACTTCTTCCACGAGTTCCTGCC	-	1199 to 1228
3233	CCATCTCGTTGTCCTTACGACGACCAGACT	-	3223 to 3252
7037	AGATATTGATTATGGTGTTAAAGCAGACCA	+	7020 to 7049
16236	TTATCTGATTTTTTCCCCTTGATTTCGGGG	-	16226 to 16255
21039	TAAGGCAAACGAGACCGAGAGAGCTGCAGC	+	21022 to 21051
23084	TTGACGATTGGGAACCGTGGAAGGAATTTG	+	23067 to 23096
24343	AACACAGATGTTTTAGACCATGCGCAGAAG	+	24326 to 24355
25461	TATTTGTACGTGAGTGGAAGTGCTTAGACT	+	25444 to 25473
27013	TTTCATCGTCAATTTCCATGTTATAAATCT	-	27003 to 27032
29998	TCGTTTTCAGTCATTGGTGGTTTGTCAGCG	-	29988 to 30017
30386	AGAAGCACCTCTTGCGTTGATAAAAGTATT	+	30369 to 30398
31065	ATATTCATATTCCCTGCTCATGTTTGATAG	-	31055 to 31084
31149	CTTTATACTCGTTAAGAATGGCATCTACGA	+	31132 to 31161
31725	CACATATCGACGTATCGTGATTATCCCATT	+	31709 to 31737
34608	AGCCTAGATAGCGAAGTTGATCGTATCTAT	+	34587 to 34616

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S2: Summary of the phages used in the polymorphic phage treatment and their escape mutations. The escape mutations are shown in red in the protospacer sequence.

Host strain name	Phage mutant name	Phage protospacer sequence	Gene containing the protospacer	Annotation
971	2972RsA9-A	AGGAGGTGGACATATTGGGCTAAATCGACGACAGAA	non coding	-
1209	2972RsA8-C	GCTCTACGACTTCTTCCACGAGTTCCTTCCTCAGAA	AAW27924.1	Terminase small subunit
3233	2972RsC7-A	CCATCTCGTTGTCCTTACGACGACCATACTTGAGAA	AAW27927.1	Portal protein
7037	2972RsB7-A	AGATATTGATTATGGTGTTAAAGCAGAGCATAAGAA	AAW27931.1	Head protein
16236	2972RsD5-A	TTATCTGATTTTTTCCCCTTGATTTCGCGGATAGAA	AAW27941.1	Tail protein
21039	2972RsC4-B	TAAGGCAAACGAGACCGAGAGAGCTGCAGCCGAGAC	AAW27942.1	Antireceptor
23084	2972RsD1-D	TTGACGATTGGGAACCGTGGAAGGAATTTGCAAAAA	AAW27943.1	Structural protein
24343	2972RsC6-C	AACACAGATGTTTTAGACCATGCGCAGA-GGGAGAA	AAW27946.1 and non coding	hypothetical protein
25461	2972RsA10-A	TATTTGTACGTGAGTGGAAGTGCTTAGACTTTAAAA	AAW27949.1 and non coding	hypothetical protein
27013	2972RsB8-B	TTTCATCGTCAATTTCCATGTTATAAATCTCTTGAA	AAW27955.1	hypothetical protein
29998	2972RsC8-A	TCGTTTTCAGTCATTGGTGGTTTGTCAGCGAAAGAG	AAW27959.1	Replication protein
30386	2972RsD6-B	AGAAGCACCTCTTGCGTTGATAAAAGCATTGCAGAA	AAW27959.1	Replication protein
31065	2972RsC2-B	ATATTCATATTCCCTGCTCATGTTTGTTAGCAAGAA	AAW27960.1	Primase
31149	2972RsA7-A	CTTTATACTCGTTAAGAATGGCATCTTCGACAAGAA	AAW27960.1	Primase
31725	2972RsC3-E	ACATATCGACGTATCGTGATTATGCCATTCAAGAA	AAW27960.1	Primase
34608	2972RsB9-B	AGCCTAGATAGCGAAGTTGATCGTATCTGTTTAGAA	AAW27966.1	hypothetical protein

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S3: Summary of mutations in the genome of the 16 starting host strains.

Host strain name	Mutations			
	In gene			Intergenic
	Non-synonymous		Synonymous	
	Missense	Nonsense		
31725	8	3	0	1
1209	2	0	0	0
27013	4	0	1	1
30386	2	0	0	0
3233	9	1	2	1
25461	3	0	1	0
24343	6	2	2	0
29998	6	2	2	1
31149	2	0	0	1
23084	2	0	0	1
21039	2	1	0	0
34608	2	0	2	1
7037	1	0	0	0
16236	2	0	2	0
31065	2	0	1	0
971	7	0	0	1

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S4: The black shading indicates the time at which we detected an additional spacers in the CRISPR-Cas CR3 locus by PCR (see Methods) for each replicate of both (a) the monomorphic and (b) the polymorphic phage treatments.

(a)

Replicate	Time (days)				
	0	1	2	3	4
B1					
B2					
B3					
B4					
B5					
B6					
B7					
B8					

(b)

Replicate	Time (days)				
	0	1	2	3	4
C1					
C2					
C3					
C4					
C5					
C6					
C7					
C8					

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

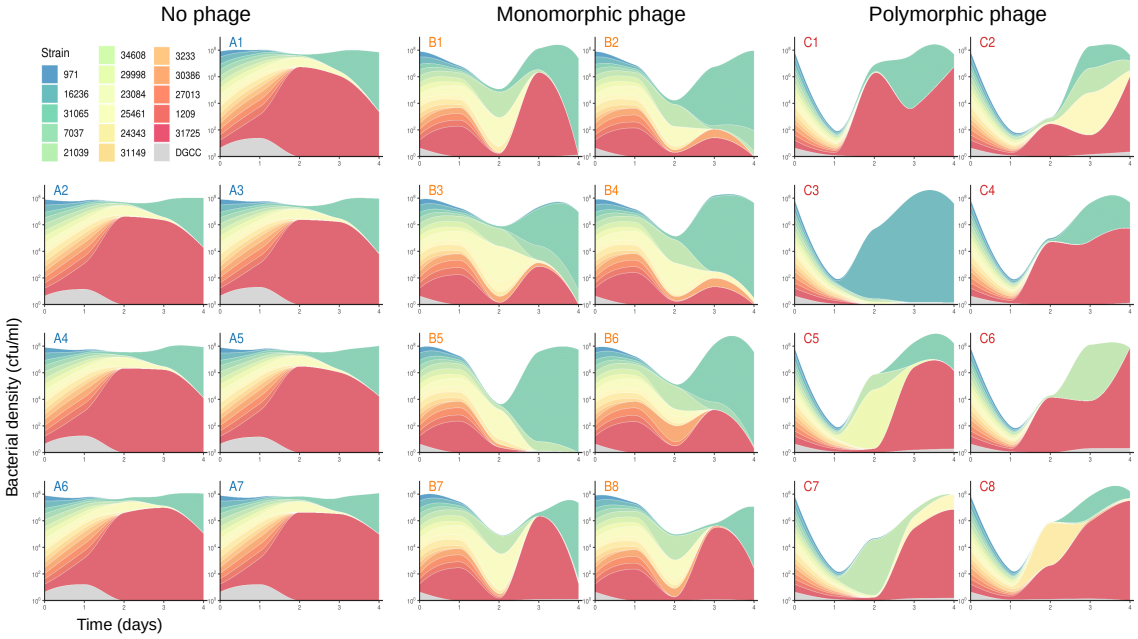

Figure S2: Modified Muller plots of the bacterial populations based on the first spacer. Above each graph is the name of the replicate ('A' for the no phage control, 'B' for the monomorphic phage treatment and 'C' for the polymorphic phage treatment). The total height for each day shows the bacterial density (in cfu/ml) on a log scale, and the different colors show the proportion of the strains at each time point on a linear scale. The 17 strains that were added on day 0 (including the phage sensitive strain in grey) are shown in the legend. The blue to red color scale is used to rank the strains according their density in the control at day 1. The lines are smoothed between each day.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

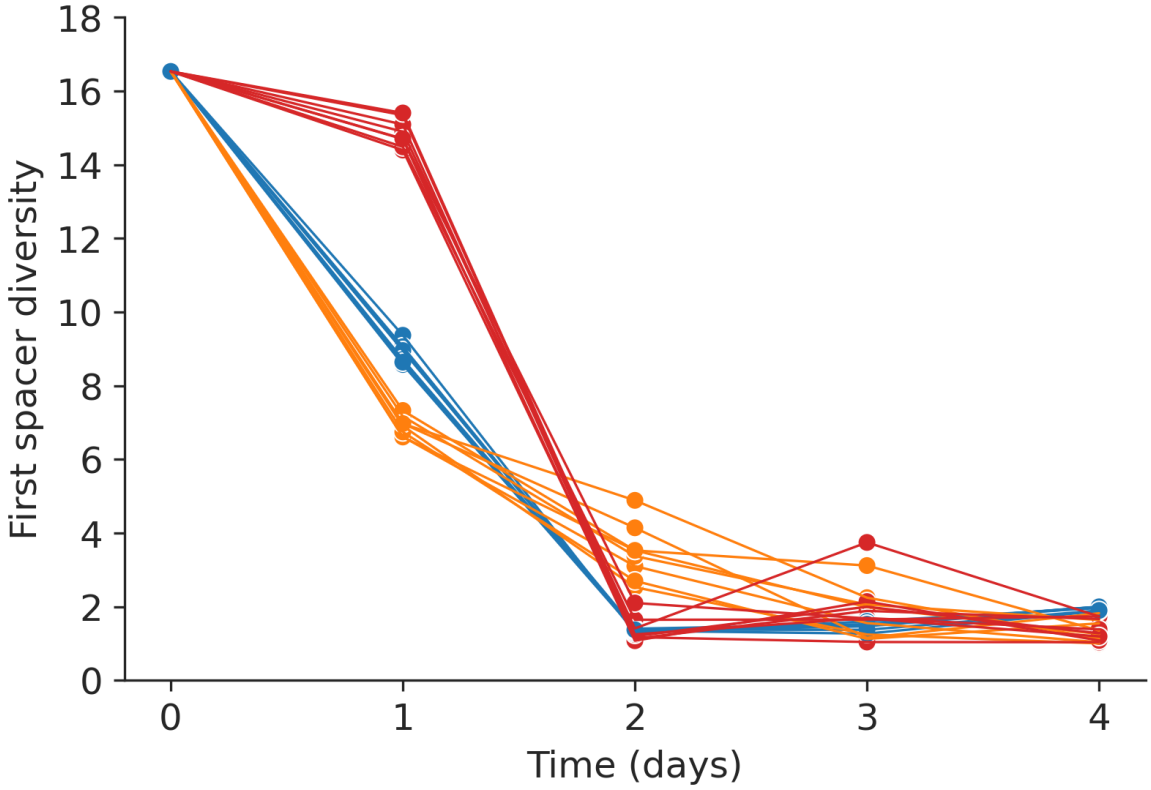

Figure S3: Diversity of the first spacer of resistance in the bacterial population. The diversity is shown as the effective number of genotypes computed only on the first spacer from each CRISPR locus. Blue points show the data in the absence of phages, orange and red respectively show data for the monomorphic and polymorphic phage treatments.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

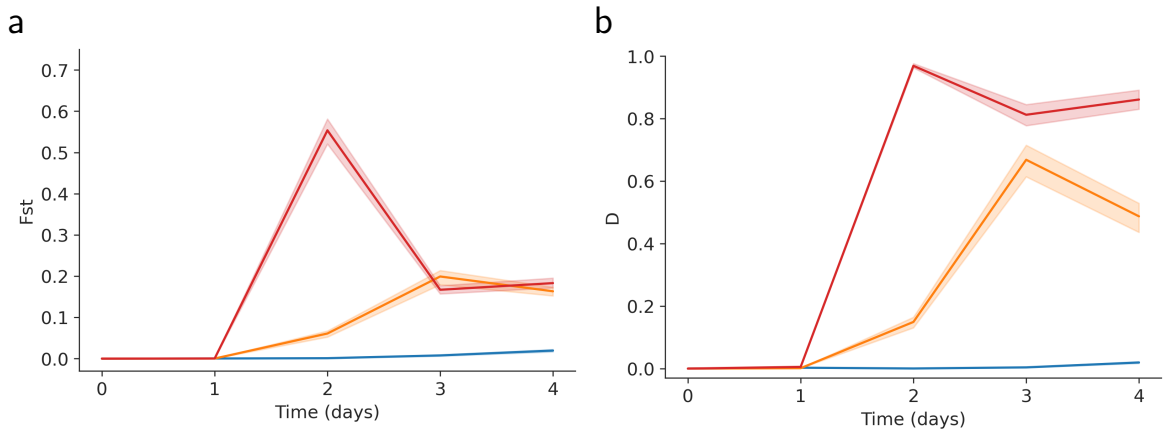

Figure S4: Measure of the differentiation of bacterial population between replicates of the same treatment with F_{ST} (a) and Jost's D (b). The F_{ST} was computed for each treatment with the method from (Weir Cockerham 1983) and Jost's D with the heterozygosities [40]. Blue points show the data in the absence of phages, orange and red respectively show data for the monomorphic and polymorphic phage treatments. The colored areas show the bootstrap 95% confidence interval.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

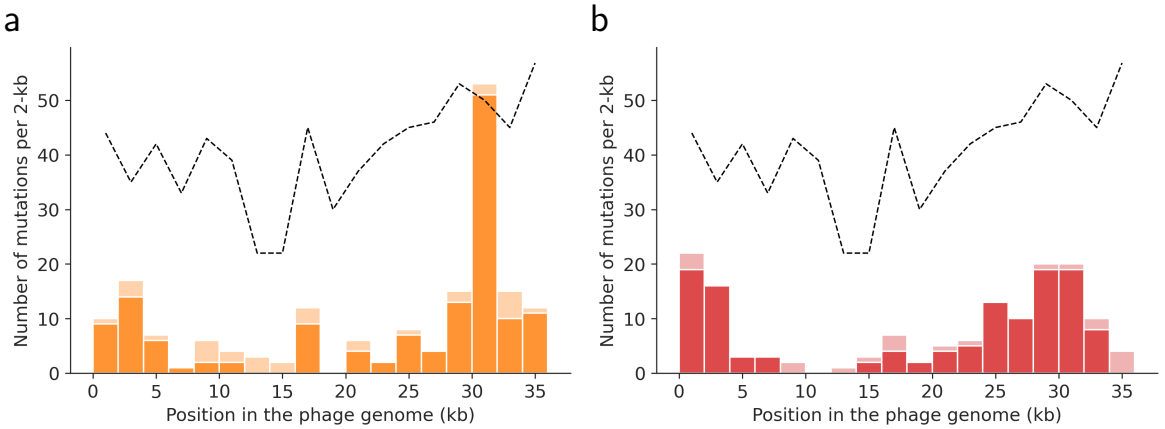

Figure S5: Mutations in the phage genome in the monomorphic (a) and polymorphic (b) phage treatments. The histogram shows the number of mutations per region of 2-kb in the phage genome. The light colors show mutations that are not located in a protospacer. The black dashed line shows the density of PAM in the genome.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

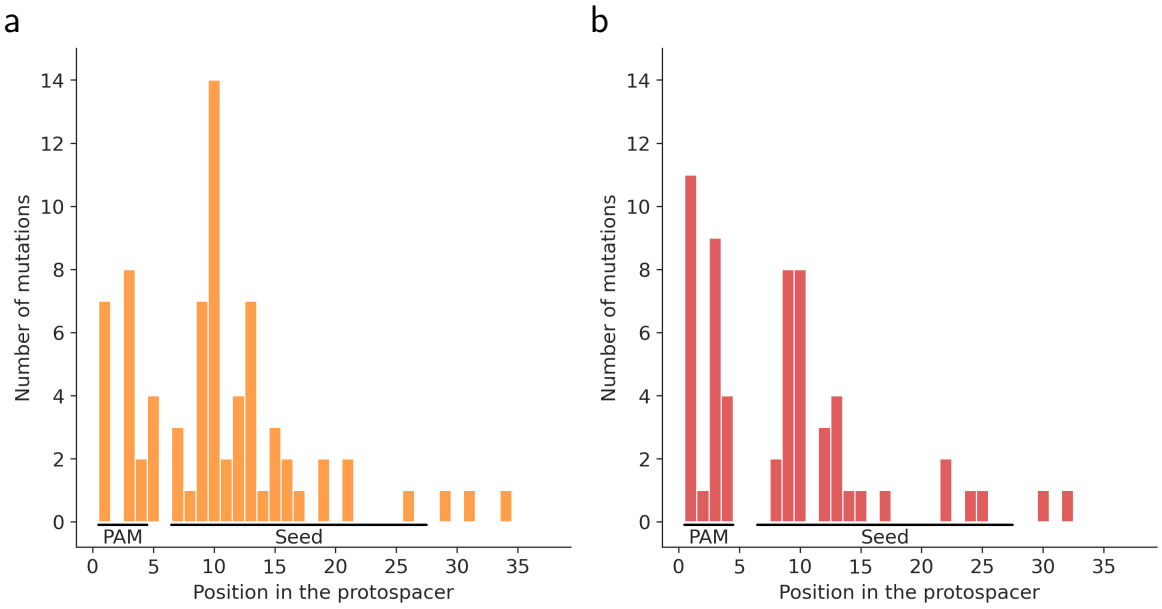

Figure S6: Position of the phage mutation in the protospacer in the monomorphic (a) and polymorphic (b) phage treatments. The mutations falling into two overlapping protospacers were discarded. The PAM and the seed region of the protospacer are shown.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

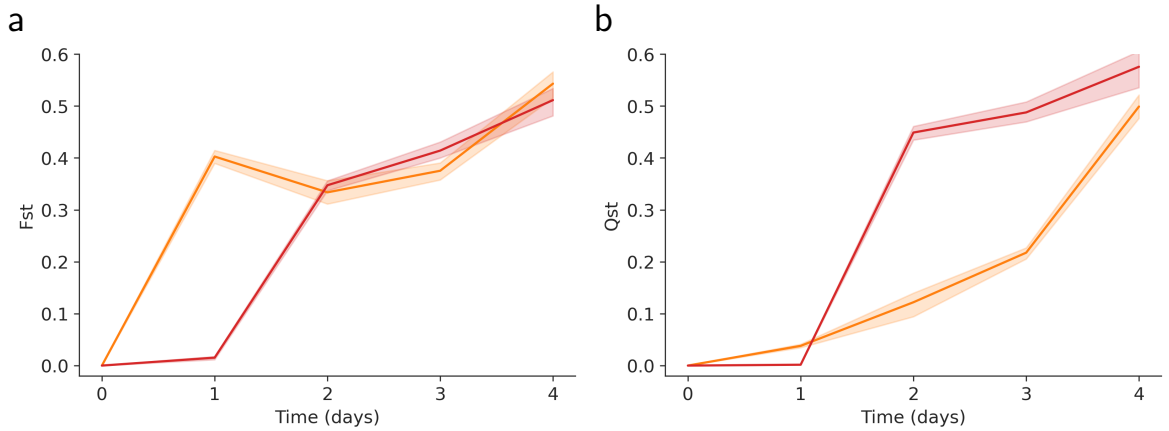

Figure S7: Measure of the differentiation of phage population between replicates of a same treatment with the F_{ST} (a) and Q_{ST} (b). The F_{ST} was computed for each treatment with the method from (Weir Cockerham 1983) [41]. The Q_{ST} is computed like the F_{ST}, but the mutations granting escape to a spacer are regrouped. Orange and red respectively show data for the monomorphic and polymorphic phage treatments. The colored areas show the bootstrap 95% confidence interval.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

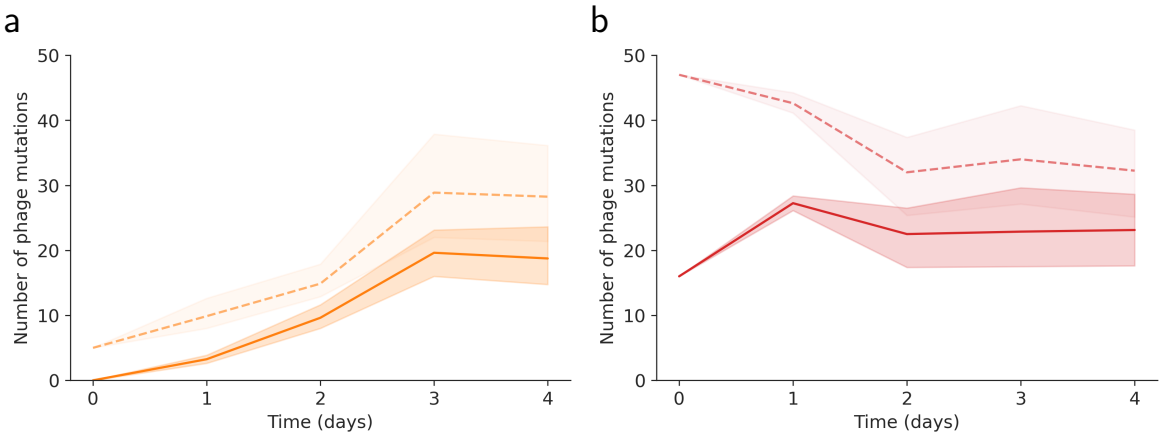

Figure S8: Number of phage mutations through time in the monomorphic (a) and polymorphic (b) phage treatments. The plain line shows the mutation in protospacers, the dashed line shows all of the mutations. Only mutations with frequencies over 0.025 are kept. The colored areas show the bootstrap 95% confidence interval.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

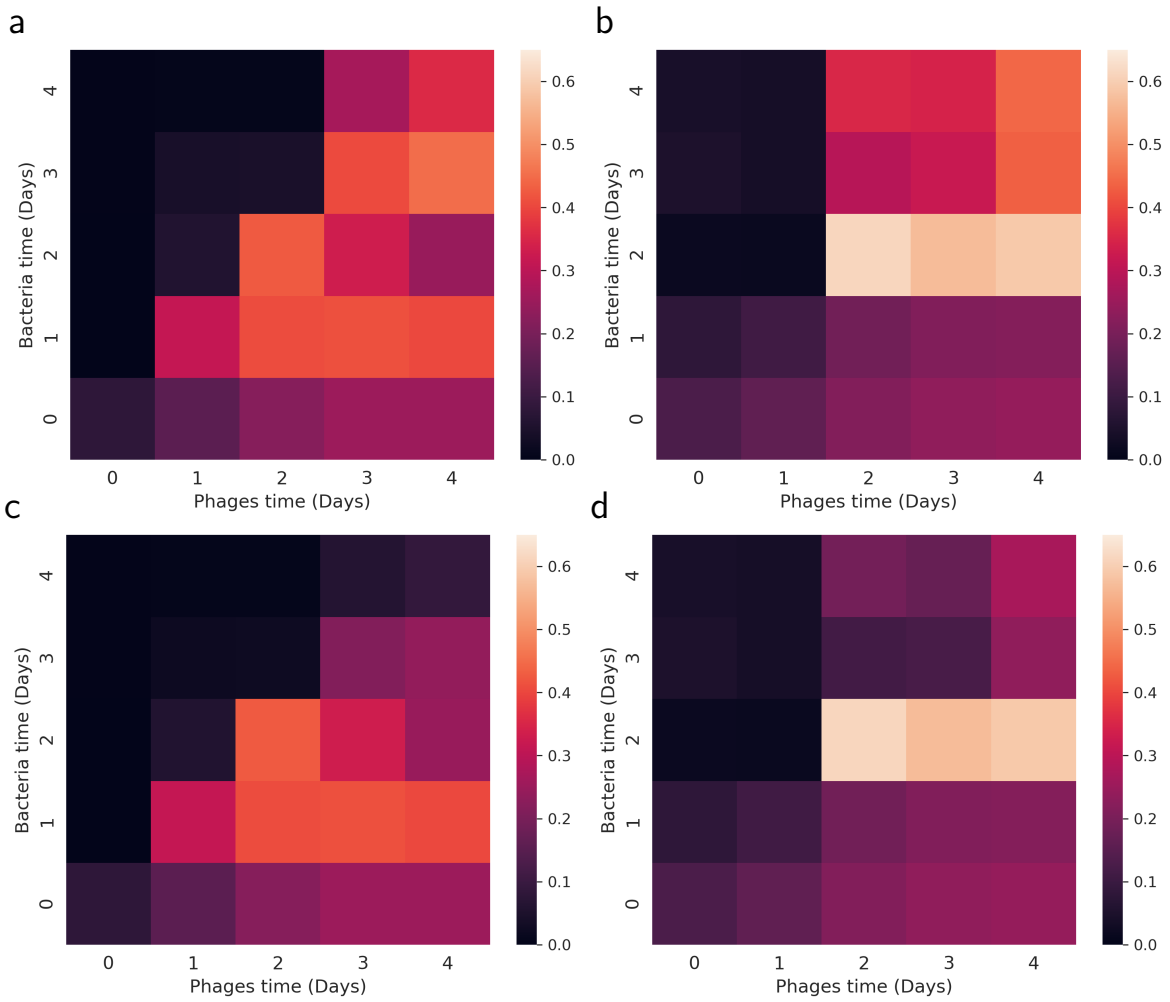

Figure S9: Phage fitness when confronting *in silico* phages and bacteria of each time points from the same replicate in the monomorphic (a,c) and polymorphic (b,d) phage treatments. The fitness was computed using equation (1). In panels c and d we try to correct the signal from the CR3 locus. For this we selected all bacterial strains i that stay at a frequency over 0.1 while the corresponding escape mutation i in the phage is at a frequency higher than 0.5. In this case, these host strains keep growing when they should be susceptible to most phages so we assume that these host strains are actually resistant and do not contribute to the fitness of the phage, thus putting their respective contribution $h_i p_i = 0$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

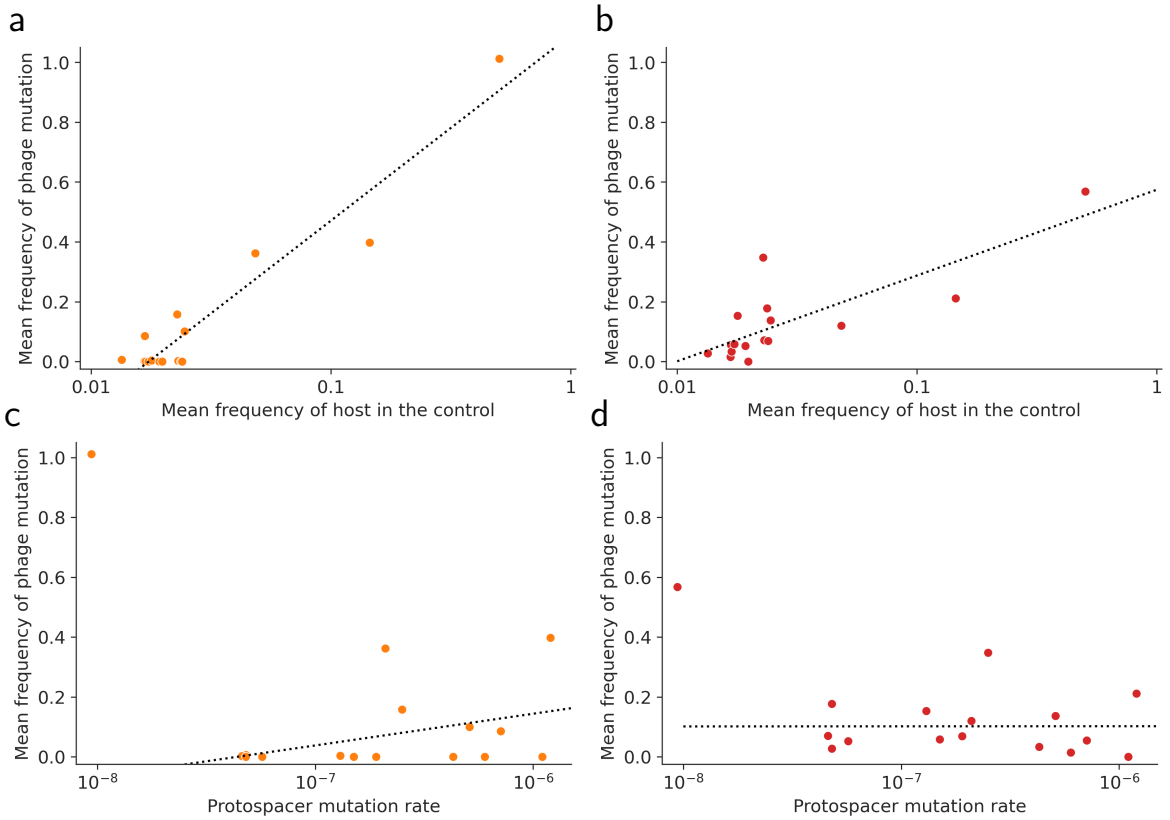

Figure S10: Phage mutation frequencies correlate with host frequency in the control. The phage escape mutation frequency against the 16 original spacers is shown according to the corresponding host frequency in the control (a,b) or according to the protospacer mutation rate (c,d). The results are shown for the monomorphic phage treatment (a,c) and the polymorphic phage treatment (b,d). One point represents one protospacer corresponding to one of the original 16 spacers, and the frequency data for each point is the mean over 4 days of the experiment, excluding day 0. A log-linear relationship was plotted in a dashed line to illustrate the relationship. In panels (c,d), the point on the upper left side was left out of the linear regression as it can be seen as an extreme point. This point is not left out of the Pearson's r calculation given in the main text.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. King, K. & Lively, C. Does genetic diversity limit disease spread in natural host populations? *Heredity* **109,** 199–203 (2012).
2. Van Houte, S. *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532,** 385–388 (2016).
3. Chabas, H. *et al.* Evolutionary emergence of infectious diseases in heterogeneous host populations. *PLoS biology* **16,** e2006738 (2018).
4. Ehrlich, P. R. & Raven, P. H. Butterflies and plants: a study in coevolution. *Evolution,* 586–608 (1964).
5. Thompson, J. N. *The coevolutionary process* (University of Chicago press, 2009).
6. Koskella, B. & Brockhurst, M. A. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS microbiology reviews* **38,** 916–931 (2014).
7. Frank, S. Models of plant-pathogen coevolution. *Trends in Genetics* **8,** 213–219 (1992).
8. Nuismer, S. *Introduction to coevolutionary theory* (Macmillan Higher Education, 2017).
9. Weinbauer, M. G. Ecology of prokaryotic viruses. *FEMS microbiology reviews* **28,** 127–181 (2004).
10. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* **45,** 1320–1328 (2000).
11. Winter, C., Bouvier, T., Weinbauer, M. G. & Thingstad, T. F. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. *Microbiology and Molecular Biology Reviews* **74,** 42–57 (2010).
12. Brockhurst, M. A. & Koskella, B. Experimental coevolution of species interactions. *Trends in ecology & evolution* **28,** 367–375 (2013).
13. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
14. Horvath, P. *et al.* Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus.* *Journal of bacteriology* **190,** 1401–1412 (2008).
15. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8,** 317–327 (2010).
16. Martel, B. & Moineau, S. CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. *Nucleic acids research* **42,** 9504–9513 (2014).
17. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus.* *Journal of bacteriology* **190,** 1390–1400 (2008).
18. Chabas, H. *et al.* Variability in the durability of CRISPR-Cas immunity. *Philosophical Transactions of the Royal Society B* **374,** 20180097 (2019).
19. Hynes, A. P. *et al.* An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. *Nature microbiology* **2,** 1374–1380 (2017).
20. Nei, M. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70,** 3321–3323 (1973).
21. Betts, A., Gray, C., Zelek, M., MacLean, R. & King, K. High parasite diversity accelerates host adaptation and diversification. *Science* **360,** 907–911 (2018).
22. Barrangou, R. *et al.* Genomic impact of CRISPR immunization against bacteriophages. *Biochemical Society Transactions* **41,** 1383–1391 (2013).
23. Koskella, B. & Lively, C. M. Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution: International Journal of Organic Evolution* **63,** 2213–2221 (2009).
24. Blanquart, F. & Gandon, S. Time-shift experiments and patterns of adaptation across time and space. *Ecology letters* **16,** 31–38 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.12.468349; this version posted November 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

25. Gandon, S., Buckling, A., Decaestecker, E. & Day, T. Host–parasite coevolution and patterns of adaptation across time and space. *Journal of evolutionary biology* **21,** 1861–1866 (2008).

26. Nourmohammad, A., Otwinowski, J. & Plotkin, J. B. Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS genetics* **12,** e1006171 (2016).

27. Childs, L. M., England, W. E., Young, M. J., Weitz, J. S. & Whitaker, R. J. CRISPR-induced distributed immunity in microbial populations. *PloS one* **9,** e101710 (2014).

28. Blanquart, F., Kaltz, O., Nuismer, S. L. & Gandon, S. A practical guide to measuring local adaptation. *Ecology letters* **16,** 1195–1205 (2013).

29. Common, J., Walker-Sünderhauf, D., van Houte, S. & Westra, E. R. Diversity in CRISPR-based immunity protects susceptible genotypes by restricting phage spread and evolution. *Journal of evolutionary biology* **33,** 1097–1108 (2020).

30. Common, J., Morley, D., Westra, E. R. & van Houte, S. CRISPR-Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage. *Philosophical Transactions of the Royal Society B* **374,** 20180098 (2019).

31. Paez-Espino, D. *et al.* CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *MBio* **6,** e00262–15 (2015).

32. Kassen, R. & Bataillon, T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nature genetics* **38,** 484–488 (2006).

33. Bataillon, T., Zhang, T. & Kassen, R. Cost of adaptation and fitness effects of beneficial mutations in Pseudomonas fluorescens. *Genetics* **189,** 939–949 (2011).

34. Breitbart, M., Bonnain, C., Malki, K. & Sawaya, N. A. Phage puppet masters of the marine microbial realm. *Nature Microbiology* **3,** 754–766 (2018).

35. Lévesque, C. *et al.* Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. *Applied and environmental microbiology* **71,** 4057–4068 (2005).

36. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10,** 1–9 (2009).

37. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

38. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9,** 357–359 (2012).

39. Garrison, E. & Marth, G. *Haplotype-based variant detection from short-read sequencing* 2012. arXiv: `1207.3907 [q-bio.GN]`.

40. Jost, L. GST and its relatives do not measure differentiation. *Molecular ecology* **17,** 4015–4026 (2008).

41. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *evolution,* 1358–1370 (1984).

42. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods* **17,** 261–272 (2020).

43. Seabold, S. & Perktold, J. *Statsmodels: Econometric and statistical modeling with python* in *Proceedings of the 9th Python in Science Conference* **57** (2010), 61.

44. Efron, B. in *Breakthroughs in statistics* 569–593 (Springer, 1992).

45. Team, R. C. *et al.* R: A language and environment for statistical computing (2013).

46. Van Rossum, G., Drake, F. L., *et al. Python reference manual* (iUniverse Indiana, 2000).