ISME

ARTICLE

Long-run bacteria-phage coexistence dynamics under natural habitat conditions in an environmental biotechnology system

Leandro D. Guerrero[1] · María V. Pérez[1,2] · Esteban Orellana[1] · Mariana Piuri[3] · Cecilia Quiroga [4] ·
Leonardo Erijman [1,5]

Abstract

Bacterial viruses are widespread and abundant across natural and engineered habitats. They influence ecosystem functioning through interactions with their hosts. Laboratory studies of phage–host pairs have advanced our understanding of phenotypic and genetic diversification in bacteria and phages. However, the dynamics of phage–host interactions have been seldom recorded in complex natural environments. We conducted an observational metagenomic study of the dynamics of interaction between *Gordonia* and their phages using a three-year data series of samples collected from a full-scale wastewater treatment plant. The aim was to obtain a comprehensive picture of the coevolution dynamics in naturally evolving populations at relatively high time resolution. Coevolution was followed by monitoring changes over time in the CRISPR loci of *Gordonia* metagenome-assembled genome, and reciprocal changes in the viral genome. Genome-wide analysis indicated low strain variability of *Gordonia*, and almost clonal conservation of the trailer end of the CRISPR loci. Incorporation of newer spacers gave rise to multiple coexisting bacterial populations. The host population carrying a shorter CRISPR locus that contain only ancestral spacers, which has not acquired newer spacers against the coexisting phages, accounted for more than half of the total host abundance in the majority of samples. Phages genome co-evolved by introducing directional changes, with no preference for mutations within the protospacer and PAM regions. Metagenomic reconstruction of time-resolved variants of host and viral genomes revealed how the complexity at the population level has important consequences for bacteria-phage coexistence.

✉ Leonardo Erijman
erijman@dna.uba.ar

1   Instituto de Investigaciones en Ingeniería Genética y Biología Molecular "Dr Héctor N. Torres" (INGEBI-CONICET), Vuelta de Obligado 2490, C1428ADN Buenos Aires, Argentina

2   Agua y Saneamientos Argentinos S.A., Tucumán 752, C1049APP Buenos Aires, Argentina

3   Departamento de Química Biológica, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, CONICET, Intendente Güiraldes 2160, C1428EGA Buenos Aires, Argentina

4   Instituto de Investigaciones en Microbiología y Parasitología Médica (IMPaM), Facultad de Medicina, Universidad de Buenos Aires, CONICET, Paraguay 2155, C1121ABG Buenos Aires, Argentina

5   Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires, Intendente Güiraldes 2160s, C1428EGA Buenos Aires, Argentina

Introduction

Bacterial viruses (also bacteriophages or phages) are present in large number and are widely distributed in naturally functioning microbial ecosystems, including engineered environments, where they have a recognized key role in controlling microbial abundance and composition [1, 2]. Well-designed experiments examining single host-phage pairs in the laboratory have advanced significantly our understanding of phenotypic and genetic diversification in bacteria and phages [3, 4]. Longitudinal studies revealed that bacterial hosts under phage attack respond by evolving continuously, using a number of mechanisms to evade phage predation. In turn, genome rearrangements, mutations, and antibacterial defense systems allow phages to overcome these barriers, leading to an evolutionary arms race [5–8].

However, the responses of bacterial cultures following phage infection in laboratory settings do not necessarily reflect the more complex interactions experienced in real-world ecosystems, where bacterial hosts can be targeted by

multiple phages [9], phages can infect more than one host [10], and density-dependent mechanisms may affect phage-bacteria coexistence [11] and immunity [12]. In addition, it has been recently demonstrated that bacteria-phage coevolutionary interactions in natural environments are strongly affected by the complexity of biotic networks [13] and by the spatial structure of the environment [14].

Therefore, there is a strong need to complement lab-scale studies with field evaluation of the dynamics of phage–host interactions in naturally occurring microbial communities [15]. Even with the current advancement of metagenomics, monitoring the coevolutionary dynamics of phages and bacteria in the environment is not straightforward. Particularly challenging is the assignment of phages to their microbial hosts [16, 17]. Possibly one of the most suitable tool for the identification of phage–host pairs is the genomic region of clustered regularly interspaced short palindromic repeats (CRISPR), present in about half of bacterial and in most archaeal genomes. Within CRISPR arrays, bacteria and archaea accumulate short sequences of invading DNA, including phages, as a genetic memory of past invasions [18]. The presence of these foreign DNA sequences, named spacers, provides prokaryotes a molecular mechanism of immunity, which allows further defense against new attacks from the same entity [19]. Because a spacer sequence matches a protospacer sequence in the infecting phage, it can be used to pinpoint phage sequences within a complex metagenomic dataset and thus identify phage–host pairs. In addition, since new spacers are added at one end of the locus (the leader end), CRISPR loci provide a reverse-chronological record of previous infections. The usefulness of this concept was demonstrated in a pioneer study, where phage and host-microbial genome sequences were reconstructed from natural acidophilic biofilms [20]. Metagenomics was later successfully applied to address the role of CRISPR–Cas immunity in shaping microbial communities in a variety of natural environments [21–24].

The diversity of CRISPR loci in natural populations is directly linked to environmental conditions and the nature of virus-host interactions. The presence of multiple coexisting CRISPR alleles has been observed in various microbial populations occurring in different habitats [20, 23, 25–27]. In several CRISPR-Cas systems, spacer diversity can be achieved either by naïve acquisition introducing single spacers or by primed acquisition, in which the effective resistance against the invader involves the acquisition of multiple spacers [28, 29]. The independent acquisition of resistance to a dominant phage by multiple host cells within a population may also contribute to their diversity and stability, a condition called population-distributed immunity [30]. The increase in population-level CRISPR diversity may protect the susceptible fraction of the host population in accordance with the epidemiological concept of 'herd immunity' [31, 32]

and dilute susceptible hosts, which may eventually lead to phage extinction, suggesting asymmetry in the coevolution between host and phage [33–35].

There remain many unknowns about the dynamics of populations of phages and bacterial in the environment [36]. A key challenge is to understand how phages and bacteria manage to coexist in complex natural environments, avoiding complete host and virus eradication [37]. To help bridge this gap, we conducted a detailed observational study of the dynamics of phage-bacteria interaction under natural habitat conditions, using a three-year data series of samples collected from a municipal full-scale activated sludge plant [38].

In this study, we focused on a bacterial host belonging to the genus *Gordonia*, which was one of the most abundant taxa and present in the majority of samples of our previous longitudinal study [38]. *Gordonia* spp. are *Actinobacteria* within the suborder *Corynebacterineae* of the order *Actinomycetales*, and thus phylogenetically related to the mycobacteria [39]. *Gordonia* species are environmental aerobic heterotrophs, generally non-pathogenic, commonly found in diverse aquatic and soil habitats. They possess a remarkable repertoire of metabolic pathways for the synthesis, transformation, and degradation of a large variety of chemical compounds, thus offering a considerable potential for environmental and industrial application [40, 41]. Members of this genus allegedly have a prevailing role in biological foaming, a common operational problem affecting wastewater treatment plants (WWTPs) worldwide [42, 43].

To date, more than 2000 individual phages infecting *Gordonia* have been isolated, mainly using the species *G. terrae* as a host. The vast majority of *Gordonia* phages are temperate. The Actinobacteriophage Database includes 460 *Gordonia* phage genome sequences that vary widely in genome length (14.6–152.4 kb) and GC-content (47.0–70.6%) [44]. A recent comparative analysis of 79 *Gordonia* phages revealed a high degree of mosaicism and a broad range of genetic relatedness [45]. They were grouped into 37 clusters according to their nucleotide sequence similarities and the sharing of at least 35% of their genes [44, 45]. Thirteen of the clusters were further divided into subclusters. Another 15 phage genomes were singletons that did not cluster with other phages. With very few exceptions, phages in all clusters infect exclusively *Gordonia* [46].

We used metagenomic data to reconstruct CRISPR-Cas loci of *Gordonia* populations in activated sludge. We then searched for corresponding protospacers in the metavirome, which allowed us to match phages to their specific bacterial host. The dynamics of phage–host interaction was followed by monitoring changes over time in the CRISPR loci of *Gordonia* populations, and the reciprocal changes in the viral genomes. The aim of the study was to obtain a comprehensive picture of phage–host coevolution in naturally

evolving populations within a complex environment at relatively high time resolution.

Methods

Wastewater treatment plant

Characteristics of the wastewater treatment plant, and full details of the sampling, nucleic acid extraction, and metagenomic analysis are provided in our previous article [38]. Briefly, samples were obtained from a full-scale municipal WWTP that provides sewage treatment for a population of ca. 600,000 residents of the metropolitan area of Buenos Aires (Argentina). A total of 60 samples of activated sludge collected biweekly over a period of 3 years from one aeration basin were stored at −80 °C until analysis.

DNA extraction and sequencing

DNA from sludge samples was extracted by a direct lysis procedure involving physical disruption of cells and a CTAB (cetyltrimethylammonium bromide) purification step [38]. The DNA extracted from the 60 sludge samples were sent to INDEAR (Rosario, Argentina) for Nextera DNA library preparation and sequencing. A rapid-run sequencing on two lanes was performed in a HiSeq 1500 Illumina, generating paired-end (PE) reads of 250 bp.

Gordonia MAG assembly and characterization

Reads were recovered from the *Gordonia* MAG assembled in Perez 2019 [38]. We confirmed that all scaffolds belonged to the same MAG by calculating the abundance and tetranucleotide frequency correlation of each scaffold in the MAG. Relatedness of the MAG to known *Gordonia* species was evaluated on the basis of whole-genome pairwise average nucleotide identity (ANI) [47].

The number and abundance of strains in the core *Gordonia* MAG that were present over the whole sampling period was inferred using DESMAN [48]. DESMAN pipeline was applied over marker genes extracted from checkM and filtered by COG annotation (only annotated genes), protein minimum length (330 aa) and total coverage (range 280X to 375X).

Targeted metagenomic CRISPR array reconstruction

CRISPR-Cas system identification was done by detection and analysis of the cas operons using CRISPRCasFinder [49] and CRISPRDetect [50]. The leader regions were defined as the sequences spanning from the last cas gene to the first repeat.

After identified the repetitive sequence of each CRISPR, *bbduck* (bbmap software), was used to extract all the reads from the metagenomic samples containing that sequence allowing one mismatch. These reads were used to reconstruct all the detectable variants of that particular CRISPR array as follow:

For each read from the previous step, the position of the repeat sequence was identified and used to trim the read into small sequences compose of seqL-repeat-seqR (s-r-s), where seqL and seqR were the left and right flanking sequences of the repeat. These s-r-s sequences were used as building blocks to reconstruct the CRISPR array network by pairwise alignment of the flanking sequences (s part of the sequence). If two s-r-s aligned by one of their extremes, both sequences were contiguous (s_1-r-s_2 + s_3-r-s_{4+} ==> s_1-r-$s_{2=3}$-r-s_4). Therefore, if a sequence was flanked by repeats it had to be a spacer. The leader sequence was identified as the sequence that more often appeared as a "leaf" in the network. Its presence was later checked in the CRISPR-Cas locus of the *Gordonia* MAG. As a result, we obtained a network of all possible spacers (nodes) connected by repeats (edges). A series of parameters for nodes and edges, such as % mismatch, alignment length, coverage, sequence quality, were calculated and used to manually curate the network. For the analysis, a series of python scripts were created and used in combination with other available scripts from Mothur [51] (Supplementary Fig. S1).

Phage genomes assembly and annotation

The reconstructed CRISPR were used to predict phage–host associations [17]. We used the sequence of the CRISPR spacers to match phage protospacers in the metagenome obtained from the same longitudinal study and phage genomes available at the Actinobacteriophage Database [44]. Matching contigs with identical temporal abundance profile were extracted and reassembled using Spades [52]. Based on this approach, two complete or nearly complete phage genomes were reconstructed. These genomes were used as templates to estimate their relative abundance mapping reads using bowtie2 [53] and the JGI script from Metabat2 [54]. When possible, reads from individual samples were assembled independently to identify potential genomics rearrangement.

Phages annotation was performed using DNAMaster [55] and curated manually.

Protospacers were identified in the phage genomes sequences using blastn, allowing for up to five mismatches, which were subsequently inspected manually. To identify the PAM sequence, 6 bp at both end of the protospacers were extracted, aligned, and visualized using WebLogo [56].

In order to find other possible bacterial host for the two identified *Gordonia* phages, Crass [57] and metaCRT [58]

were used to detect and reconstruct other CRISPR arrays from the metagenomic contigs and raw reads, respectively. After CRISPR identification, spacers from all CRISPR arrays were extracted and blasted (blastn) against the two *Gordonia* phage genomes.

Analysis of sequence variants in phage DC-56 genome

Phage SNVs were identified in samples with a minimum coverage of 10X. The tool *Samtools mpileup* [59] was used to assess the proportion of each variant per position in the phage genome using the read mapping information (BAM files). Each nucleotide position was considered as an SNV when a mismatch to the consensus nucleotide at that position was supported for at less 5 reads. To reduce the possibility of including false-positive SNV due to unspecific recruitment of reads, positions with coverage in excess of two-fold greater than the average coverage of the genome were excluded.

Phylogenetic signal analysis

Differences in phage SNV composition among samples were identified using *bcftools gtcheck* [59] and represented as an error rates matrix. Phylogenetic signal analysis was performed with the R package phylosignal [60], using a cluster of samples based on the number of discordant SNV (error rate matrix) and time as trait. Random generated values and sample time randomization were used as control traits.

Results

Virus assembly and annotation

Draft genomes of two phages, designated phage DC-56 and phage DS-92, were assembled each into single contigs of 56,274 bp and 92,107 bp, respectively (Supplementary Fig. S2). Structural variants between genomes assembled from individual samples were not detected. Therefore, reconstructed phage genomes were composite assemblies from all reads collected from samples in which phages were present.

GC-content of phage DC-56 and phage DS-92 were 65.2 and 52.4, respectively. Both phages belonged to the *Siphoviridae* family. Phage DC-56 was classified into cluster DC and phage DS-92 into cluster DS, according to the similarity with other *Gordonia* phages and members of the corresponding mycobacteriophage clusters [44, 45].

Phage DC-56 has 67 putative open reading frames (ORFs), 64 of which are transcribed in one direction and only three in the opposite direction (Supplementary Fig. S2A). Like other *Gordonia* phages, phage DC-56 encodes two integrase genes, a tyrosine integrase (Int-Y, gene 46) and a serine integrase (Int-S, gene 47) near the center of the genome. Phage DS-92 encodes 111 potential ORFs, of which 46 are transcribed rightwards and 65 leftwards. We were not able to identify integrase genes in this genome.

Gordonia metagenome-assembled genome

The *Gordonia* metagenome-assembled genome (MAG) was one of the 173 MAGs obtained in a previous study using data from a time-series analysis of a municipal full-scale activated sludge plant [38]. The refined *Gordonia* sp. MAG comprised 59 contigs, with a size of 5.1 Mbp, N50 of 158040 and a GC-content of 66.35%. The analysis using single-copy marker genes [61] indicated that the assembly was 98.6% complete with less than 0.3% contamination. The MAG contained a total of 4512 gene-coding regions, 58 tRNAs, and the complete sequences of 5S, 16S and 23S ribosomal RNAs. Therefore, the MAG complied with all standards of a high-quality draft [62].

Scores of genome pairwise average nucleotide identity computed with 87 genomes of recognized species of the genus *Gordonia* deposited in the NCBI database were <86%, well below the value of 95% that is considered the threshold for novel species definition (Supplementary Table S1).

Since *Gordonia* MAG was co-assembled using multiple samples across the time-series, it was important to assess the strain heterogeneity of the resulting assembly. Single-nucleotide variants (SNV) in 216 marker genes and co-occurrence across samples from the longitudinal study were used to investigate strain-level variations in the *Gordonia* MAG. Only 178 significant SNVs were distributed across the 5.1 MB *Gordonia* MAG, allowing the differentiation into four coexisting strains, with one strain dominating throughout the whole investigated period (Supplementary Fig. S3). The very low genetic diversity between the four strains suggests that they derived from the same ancestor and most likely did not differ in their functional and genetic traits. Therefore, the *Gordonia* MAG, assembled as a composite genome using reads from the four closely related strains, is considered hereafter as a single lineage.

CRISPR loci

Gordonia genome contained 2 CRISPR-Cas loci, classified as type I-E and type III-B-like systems, based on the architecture of the Cas proteins (Supplementary Fig. S4). Automatic prediction of CRISPR arrays from MAGs may underestimate the diversity of spacers because of the limited
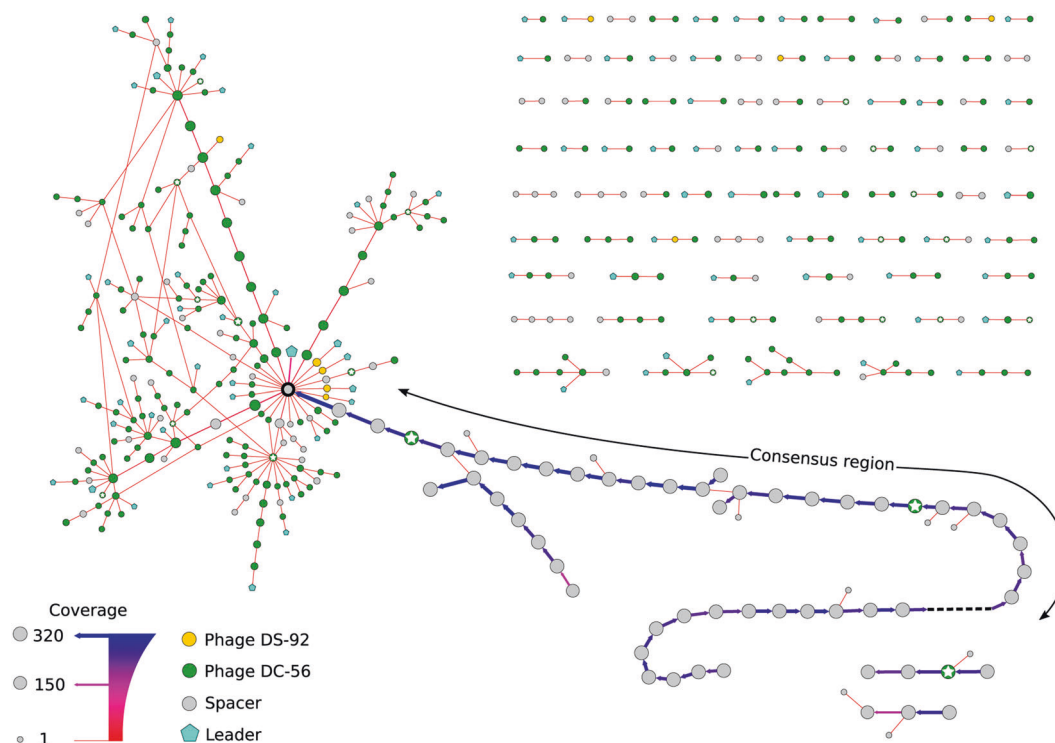
Fig. 1 Reconstruction of the spacer arrangement of *Gordonia's* CRISPR-1 across the complete set of metagenomes from Perez et al. [38]. Each circle represents a spacer. Leader sequence is shown as pentagons. The lines connecting each spacer represent their position relative to other spacers, with arrows pointing to the direction of the leader sequence. Every combination of nodes connected by edges pointing in the same direction represents a possible CRISPR variant in the population. Nodes with more than one incoming connection (arrows) represent spacers that may be linked to one or more variants. Coverage is indicated by the diameter of the nodes and the width and color of the lines. Spacers targeting phage DS-92 and phage DC-56 are colored with yellow and green, respectively. Spacer with mismatches with phages' protospacers are marked with a star. Unconnected spacers in the upper right portion of the figure are probably new acquired spacers, which is suggested by their low coverage and the general presence of the leader sequence. Unconnected nodes of high abundance in the lower right portion of the figure have the same coverage as the rest of *Gordonia* MAG but could not be identified as part of other contigs of the genome.

capacity to resolve the multiple CRISPR variants present in a mixed population. To overcome this problem we used a specific pipeline to reconstruct all observed CRISPR variants, based on the overlapping of every possible connection between spacers identified in raw reads across the complete time-resolved metagenomic dataset (see Methods) [57, 63, 64]. The reconstruction of the type I-E CRISPR array, designated CRISPR-1, can be visualized in the network shown in Fig. 1. The spacer arrangement of CRISPR-1 showed one particular spacer (node circled with a thick line in Fig. 1) acting as a breaking point, after which the diversity of CRISPR variants increased markedly in the direction of the leader end. Thus, spacers were highly conserved in the direction of the trailer end of the CRISPR locus, whereas new spacer acquisition led to a considerable increase in the CRISPR population.

The low-coverage, fan-like-structure at the leader end (>=1X, Fig. 1) confirmed the active incorporation of new spacers. The majority of these newly acquired spacers perfectly matched sequences of one of the coexisting phages (DC-56). Few spacers also matched sequences of the

phage DS-92 genome, notably in the short-branched variants, suggesting simultaneous adaptation to multiple phages. The proximal end of most CRISPR-1 variants contained leader sequences, which are essential for the acquisition of new spacers. Spacers within the consensus region did not match phage DC-56 genome, with the exception of three spacers that had 1 or 2 mismatches, and no leader sequences could be detected downstream of the breaking point node.

We searched for evidences of previous encounters between *Gordonia* and other phages in the system but found that no other contig from the entire metagenomic dataset matched spacers in the CRISPR loci. A blast search of the complete set of *Gordonia* MAG spacers did not match any phage genome sequence in the actinobacteriophage database [44] indicating that the bacteria did not acquire spacers from any other sequenced actinobacteriophage.

The average coverage of spacers within the consensus was similar to the average coverage of contigs from the complete *Gordonia* MAG, suggesting that the majority of *Gordonia* cells carried the CRISPR-1 array, and therefore it
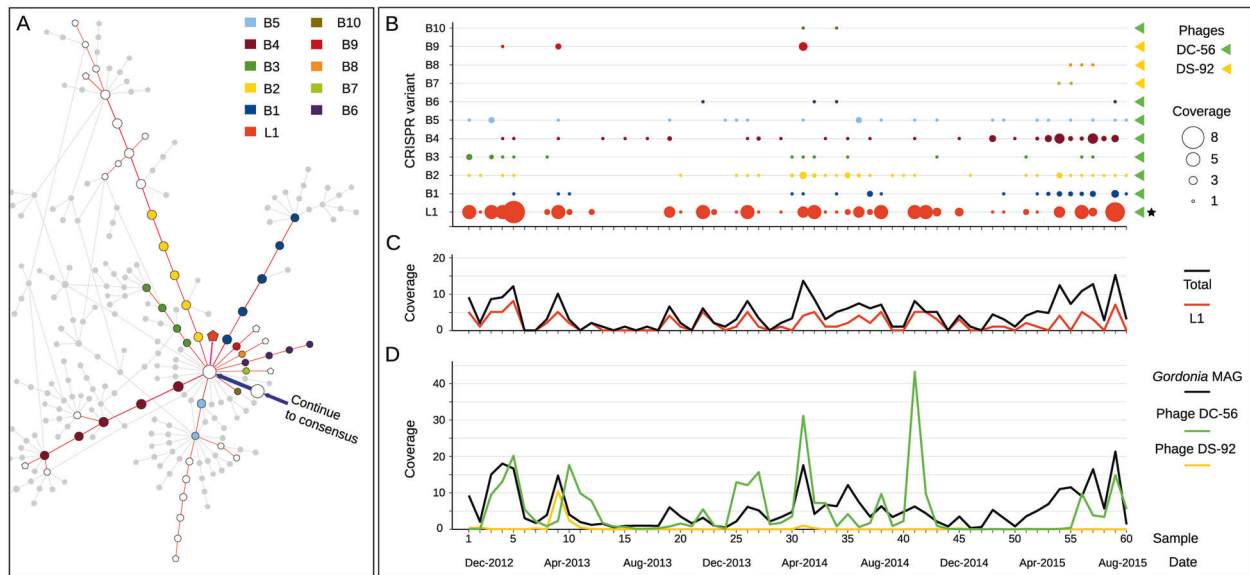
Fig. 2 Abundance of the main CRISPR-1 variants over time. A Simplified representation of Fig. 1 showing nodes with coverage >1 of the main variants. Nodes used to estimate the average abundance of each variant are identified by colors. **B** Abundance of each variant per sample. Colored triangles indicate to which phage matched each allele. The star indicates that spacers in variant L1 had at least one mismatch to the phage DC-56 genome. **C** Sum of all variants in panel B compared to variant L1. **D** Time course of the abundance of the *Gordonia* MAG and phages DC-56 and DS-92.

was not lost over the time series (Supplementary Fig. S5). Upon filtering the low-coverage nodes, 11 main variants of the CRISPR-1 were identified upstream the breaking point node (Fig. 2A). Most of the spacers within these main branches matched to the coexisting phage DC-56 (Fig. 1). The abundance distribution of the main variants over time is shown in Fig. 2B. The CRISPR-1 variant with a leader sequence immediately upstream the consensus region (named L1 in Fig. 2) accounted for more than half of the total abundance in the majority of samples (Fig. 2C).

The reconstructed type III CRISPR locus, designated CRISPR-2, contained only eight spacers (Supplementary Fig. S4, Supplementary Fig. S6A). The fact that the coverage of this locus matched the coverage of the *Gordonia* MAG (Supplementary Fig. S6B) indicates that it was likely also present in most members of the *Gordonia* population. Nevertheless, the lack of spacers matching coexisting phages and the absence of "branching" suggested that this system was either inactive or at least was not actively incorporating new spacers during the time span of this study.

## Protospacer adjacent motif sequences

The protospacer adjacent motif (PAM) sequence was searched by alignment of short stretches of sequences (up to 6 bp) immediately flanking the protospacer regions of phage DC-56 sequence. The trinucleotide GTT was highly conserved immediately at the 5′ end of those protospacers that matched spacers in CRISPR-1 (Supplementary Fig. S7),

which suggests that this is the PAM sequence required for incorporation of new spacers [65].

Phage–host dynamics

Time course of phages and *Gordonia* coexistence is shown in Fig. 2D. Abundance profiles over the investigated time interval suggest that phage abundance is sustained by lytic, rather than lysogenic cycles. In addition, there is no supportive genomic evidence that phage integration has occurred at any time point. As in any comparative analysis, abundance shifts are susceptible to the compositionality effects that arise from relative abundance measures, which may be especially conspicuous at low coverage. However, we note that at each time point bacteria and phage abundance are subject to the same relative change, whereas mixed liquor-suspended solids (MLSS) concentration, a proxy for microbial biomass, remained relatively stable across the time scale of the study [38]. Accordingly, there was a considerable overlap in peaks of bacteria and phage DC-56 over the time of the study (Pearson correlation $r = 0.42$, $p = 0.0008$; Supplementary Table S2). On the bacterial side, coexistence requires a way to avoid or resist infection. Nevertheless, it is inferred that phage replication and survival were allowed by the presence of populations of susceptible bacterial cells. This is supported by the fact that peaks of phage DC-56 coincided especially with time points of higher abundance of *Gordonia* bacterial cells lacking spacers against the virus (L1, Fig. 2 Pearson correlation $r = 0.50$; $p = 4.10^{-5}$; Supplementary Table S2). Phage DS-92
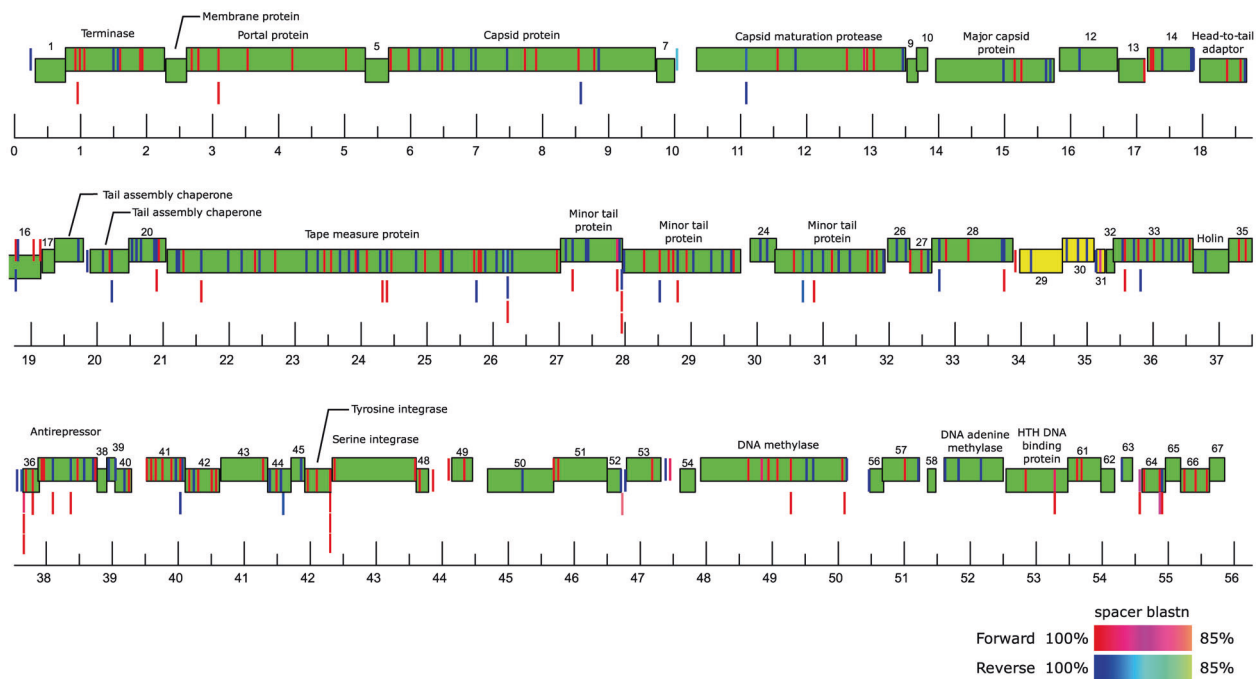
Fig. 3 Phage DC-56 genome map highlighting the position of protospacers. Genes are indicated by boxes, colored according to their forward (green) or reverse (yellow) orientation. When known, the name of the genes was added. The location of protospacers targeted by CRISPR-1 was indicated by verticals bars. Bar colors indicate the orientation and the similarity to the corresponding spacer sequence in the *Gordonia* MAG.

peaked with different abundance at two different times, in which *Gordonia* was also present in high abundance (Fig. 2D).

Adaptation of phage DC-56

To analyze the viral response to CRISPR-1 diversification, we focused on the sequence of phage DC-56 genome, whose abundance mirrored *Gordonia's* MAG. The 285 spacers that targeted the phage DC-56 genome were distributed across the entire phage genome (Fig. 3). There is an indication of a moderate spacer acquisition bias, although we found no evidence that the location and strand specificity of protospacers were influenced by previously acquired spacers, as it occurs during primed acquisition [66]. We also could not relate genomic regions with lower density of protospacers with the presence of higher frequency of the canonical *E. coli* Chi sequence [67]. Nevertheless, with the available data we cannot ruled out that low protospacer density regions are protected by a different Chi sequence that is recognized by Gordonia's RecBCD system, or the involvement of other processes such as interference-driven or primed spacer acquisition [66].

Changes in phage DC-56 genome over three years were investigated through viral SNV analysis. Each nucleotide position was marked as an SNV when a mismatch to the consensus nucleotide at that position had a minimum coverage of 5. A total of 610 SNVs satisfied this criterion. To test for differences in the efficacy of variants selection, we examined the ratios of nonsynonymous to synonymous polymorphisms in protospacers and elsewhere in the phage genome and found that the ratio was not significantly different (two-tailed Fisher's exact test $P = 0.66$). Eighty-two SNVs were located within 54 protospacers and 5 PAM regions, (Fig. 4A, B). Fisher's exact test revealed that SNVs did not occur in protospacers more frequently than in the rest of the genome ($p = 0.11$). In addition, no mutational bias towards positions in seed or PAM regions were observed.

Figure 4C shows that most spacers appeared in a limited number of samples and had very low coverage. Notwithstanding, it can be observed that changes in SNV patterns across time did not appear to be determined by the presence of a spacer targeting the corresponding sequence, casting doubt that SNV patterns are the result of immune selection from CRISPR alleles.

Phylogenetic signal was used to investigate the pattern of longitudinal population dynamics of phage DC-56. We looked at the relationship between DC-56 sequence in each sample and time with an analysis based on the occurrence of SNVs within the phage DC-56 genome [68]. Samples clustered largely according to time point, i.e. adjacent samples were, with few exceptions, more similar to each other than to more distant samples (Fig. 5, Supplementary Fig. S8).
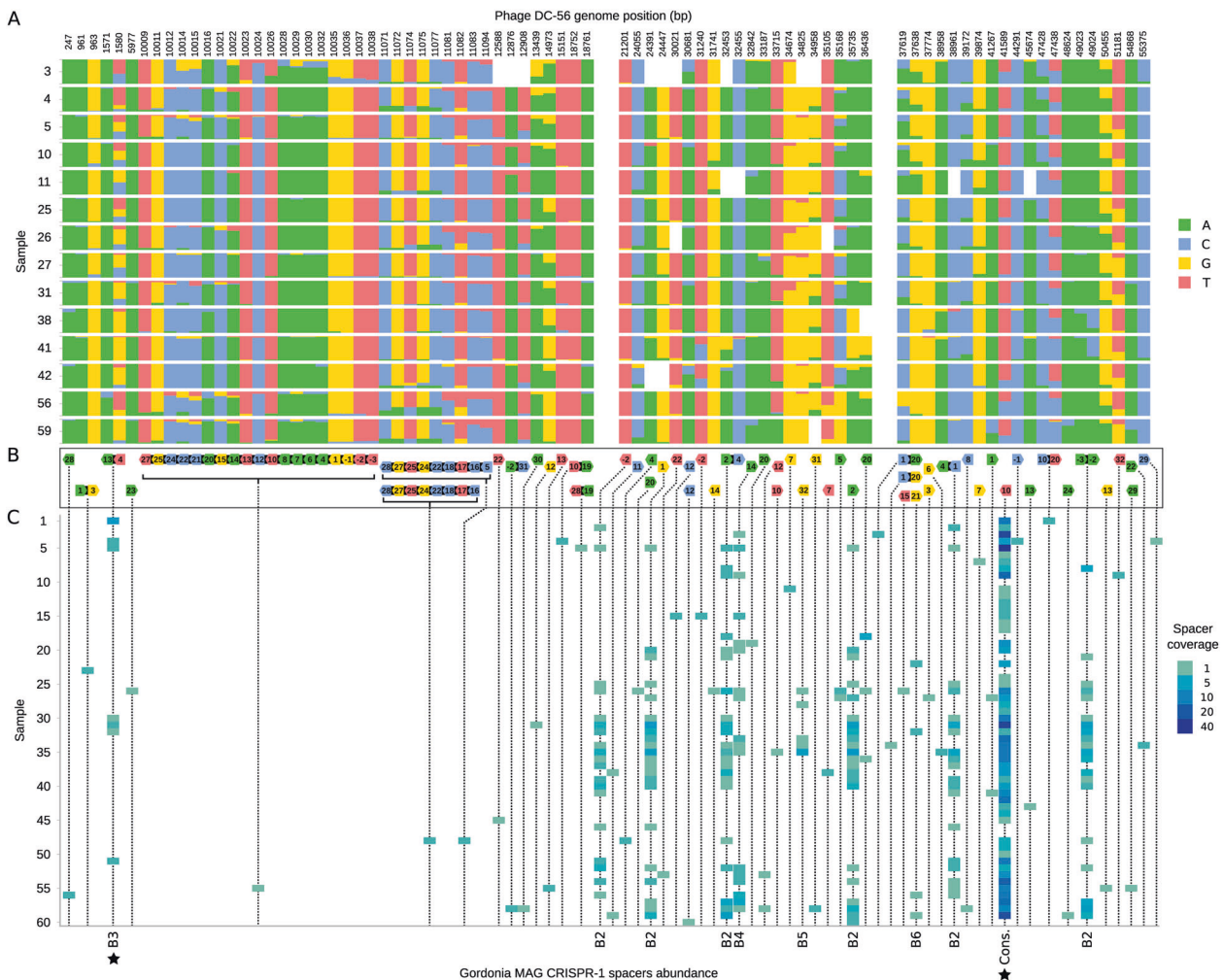
Fig. 4 Occurrence of phage DC-56 SNVs and abundance of matching spacers over time. A The 82 single-nucleotide variants (SNV) of phage DC-56 affecting protospacers and PAM regions. Only samples with a total coverage >=10 and a minimum coverage of 5 reads per sample were analyzed. White boxes represent nucleotide positions with insufficient coverage. **B** Nucleotides in the *Gordonia* CRISPR spacers corresponding to the SNV in the viral genome, drawn as arrows to show orientation. Numbers inside symbols indicate the position of the base within the spacer. Bases that match PAM sequences have negative numbers. Bases targeted by the same spacer are connected by a black bar. Bases in spacers located in the reverse orientation were colored according to the corresponding forward sequence. **C** Abundance per sample of spacers in **B**. Labels for the CRISPR alleles are shown at the bottom of the corresponding spacers (only for the main CRISPR variants and consensus). Stars indicate spacers that have other nucleotide mismatches with the DC-56 protospacer in addition to the SNV site. Sample number labels on the *Y*-axis correspond to the same time intervals as in Fig. 2.

The significant temporal autocorrelation suggests directional changes over time in phage DC-56 populations.

Discussion

This work provides a comprehensive picture of the complex dynamics of naturally evolving populations of a bacterial host and its infecting phage within an environmental biotechnology system. Metagenomic reconstruction of time-resolved variants of host and virus genomes revealed how selection operates at the population level. Diversification of the bacterial CRISPR locus, which occurs as a result of phage infection, gives rise to multiple coexisting populations of a *Gordonia* strain. The majority of newly acquired spacers matched sequences of one of the coexisting phages. These observations coincide with simulations [30], experiments performed under laboratory settings [33, 69] and observation from natural environments [24], confirming that CRISPR array evolution is driven by co-occurring phages and suggesting that the synergistic effect of spacer diversity established at the host population level aim at reducing the probability of phages to escape mutations [33, 69].

Phage–host dynamics

Under the arms race dynamics, phages would co-evolve to escape CRISPR by preferentially selecting variants with
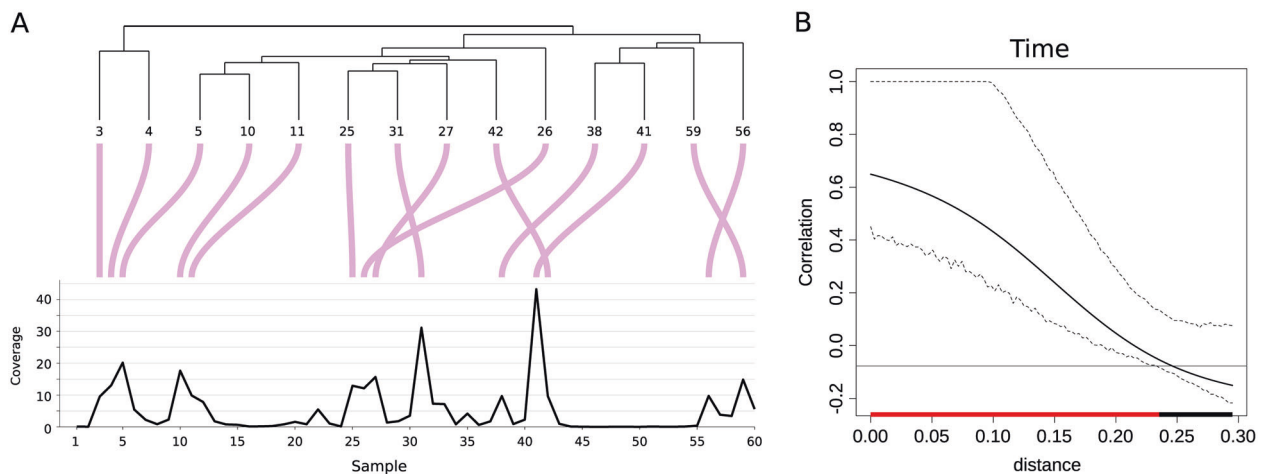
Fig. 5 Phylogenetic signal analysis of reconstructed phage DC-56. A Cluster of phage DC-56 genome based on shared SNV in samples with coverage >=10 and their correspondence to phage abundance profile (coverage) over time. **B** Correlogram of the distance between samples using time as trait. Red bar indicates positive correlation.

polymorphic positions at the seed or the PAM regions. For example, directional changes in the phage genome were observed upon long-term coincubation of *S. thermophilus* DGCC7710 and phage 2972 in milk medium [7]. This effect, which has also been observed in other phages, helps them to overcome sequence-specific CRISPR immunity [70, 71]. In our study we observed directional changes (Fig. 5); however, the replacements of phage DC-56 population were not necessarily associated with a strategy to evade CRISPR-Cas immunity, as accumulation of mutations in the PAM and protospacer regions was not favored (Fig. 4). This is not surprising, given that phages that escape bacterial immunity by mutating their target protospacers within one host will not be successful against coexisting bacterial populations that possess different spacers. Long-term virus-host coexistence through high diversity of CRISPR alleles against phage DC-56 is consistent with a model that predicts that CRISPR-Cas immunity stabilizes the virus-host system for intermediate values of the viral mutation rate [72], and it is expected under the concept of population-distributed immunity, in which a large diversity of CRISPR loci affords more resilient immunity across the entire population [30]. As long as no individual bacterial population achieves a much higher density than their competitors, phages would still find excess of susceptible hosts, reducing the selection pressure to mutate their targeted protospacers or PAM regions. As a matter of fact, the most abundant *Gordonia* population observed throughout this study contained a CRISPR array with only ancestral spacers, with none having a perfect match in the phage DC-56 genome (L1 in Fig. 2). The continuous presence of a large population of susceptible host may act as a propagating host for the phage DC-56, granting its survival [8, 34]. Nevertheless, we cannot tell

whether this is a non-immune population, associated with the fitness cost of maintaining a fully active CRISPR-Cas system [73] or that the population maintains active different defense mechanisms or it is protected from phage exposure by the spatial heterogeneity of the environment [74].

It follows that in the long run neither the host nor the virus are driven extinct, in opposition to the results of laboratory coevolution experiments, where spacer diversity of CRISPR-mediated resistance caused rapid phage extinction [33]. Whereas our data show directional changes, which are suggestive of arms race dynamics (ARD), there is no clear evidence that mutations were selected for increasing bacterial resistance and viral escape over time. Fluctuating selection dynamics (FSD) is an alternative evolutionary model that explains the antagonistic coexistence of several host-parasite systems, where host abundance can fluctuate over time without having evolved genetic resistance [75, 76]. FSD has been reported to occur in environments with low resource availability, such as soil [77] and the ocean [78], where the costs of phage resistance are expected to be high [79–81]. Interestingly, *Gordonia* thrives in activated sludge, where bacteria grow under carbon limitation conditions, with typical loading rates in the range of 0.05–0.4 kg Biochemical Oxygen Demand per kg dry biomass and per day [38, 82, 83]. Yet, our data do not allow to determine whether fluctuations of phage and bacterial genotypes are associated to a negative frequency-dependent selection of virus and hosts with different infectivity and resistance profiles [15]. This hypothesis can be tested in further investigation on phage infectivity and host range and bacterial resistance [4, 24], as well as on the protection effect due to spatial refuges for susceptible bacteria [84].

CRISPR-Cas systems are one among several mechanisms that have evolved in bacteria to fight phages [80, 85].

In its natural habitat, *Gordonia* would be protected by several lines of defense, including the protecting effect of EPS and spatial heterogeneity of the matrix. However, the fact that phage DC-56 appears to evolve relatively slowly in this environment suggests that it is the continuous presence of large number of susceptible cells that makes it possible to maintain the stable phage–host coexistence.

A potential limitation of this study is that the fluctuations in abundance of *Gordonia* and its phages were interpreted mainly in terms of their direct interaction. However, the targeting a competitor species by one of the phages or any other indirect interaction may also affect bacterial and phage abundance. Although we cannot completely rule out the existence of additional hosts, *Gordonia* phages are generally genus specific [46] whereas no evidence was found for the presence of other *Gordonia* species in the system. In addition, analysis of reconstructed CRISPR alleles from the complete set of unassembled metagenomic reads revealed that the only spacers matching sequences in phage DC-56 genome belonged to the *Gordonia* MAG. In any case, the dynamics of *Gordonia*-DC-56 phage coexistence shows a pattern in which the most abundant phages replicate actively as long as the host populations remain susceptible to their attack, as predicted by the "killing the winner" model [86]. The appearance of peaks of infection by a second phage are in turn consistent with the seed bank model, which postulates that minor virus populations are maintained in a bank fraction and can potentially become active when the environment changes [10].

Conservation of trailer-end spacers

This work confirms earlier observations that spacers at the trailer end of the CRISPR locus is highly conserved despite the fact that they would not provide immunity against coexisting viruses [21, 27, 87]. Simulations have predicted that trailer-ends may become clonal across time due to selective sweeps of highly immune CRISPR lineages [88]. More generally, the purge of genetic heterogeneity at the trailer end of the CRISPR locus appears to conform to the genome-wide sweep predicted by the ecotype model [89], where a single population (ecotype) acquires an advantageous trait that helps it to survive, reducing overall population diversity. The occurrence of genome-wide selective sweeps has already been revealed in freshwater lake microbial communities, where with few exceptions, overall genetic heterogeneity of most bacterial populations was relatively low over a period of several years [90]. Nonetheless, the low bacterial strain diversity and the existence of a conserved trailer end is puzzling because it suggests that the entire *Gordonia* population observed over 3 years appears to share a common ancestry, despite the fact that *Gordonia* are widespread in the environment and activated sludge is an open ecosystem, in which immigration plays an important role in community assembly [91, 92].

Implications for the application of phage therapy in environmental processes

Gordonia belongs to a group of Actinobacteria that have relatively high content of mycolic acids on their cell surface, called thereafter Mycolata. The highly hydrophobic cell surface of Mycolata in combination with their ability to excrete extracellular polymers, which may act as biosurfactants, favor the formation and stabilization of foams in wastewater treatment processes [93, 94]. Despite major efforts to find universal strategies to prevent or to control the unwanted proliferation of these bacteria, foaming remains a common operational problem in activated sludge. In recent years, considerable research has been directed towards the use of bacteriophages as an alternative approach to chemicals to reduce the levels of filamentous bacteria in wastewater treatment, including foam forming *Gordonia* [94, 95]. The results obtained in this work suggest that the presence of an active lytic phage might not be sufficient to control *Gordonia* abundance. Successful application of phage therapy at the full-scale will still require a better understanding of virus-host interactions in the complex real-world context.

Conclusions

This work provides a high-resolution temporal profiling of a bacterium and two of its phages in a complex engineered environment such as activated sludge. Long-term host-virus coexistence was established through diversification of the CRISPR locus of an almost clonal *Gordonia* population and directional changes in the phage genome, which however was not clearly associated with a CRISPR–Cas targeting escape strategy.

By detecting and tracking viral SNVs and bacterial CRISPR spacers, we offered a glimpse into the complexity of virus-host interaction at the population level in a real-world setting. More than fifteen years ago, there was a call proposing wastewater treatment as model systems to address fundamental ecological questions in realistic environmental conditions [96, 97]. We renew this call, proposing that wastewater treatment, in conjunction with metagenomic time-series studies, are suitable means to gain further insight into bacteria-phage dynamics and interactions.

Data availability

Sequencing data are available at NCBI BioProject under accession no. PRJNA484416. Genome sequences and

CRISPR reconstruction scripts are available at https://github.com/GuerreroCRISPR/Gordonia-CRISPR.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Bouvier T, del Giorgio PA. Key role of selective viral-induced mortality in determining marine bacterial community composition. Environ Microbiol. 2007;9:287–97.
2. Brown MR, Baptista JC, Lunn M, Swan DL, Smith SJ, Davenport RJ, et al. Coupled virus - bacteria interactions and ecosystem function in an engineered microbial system. Water Res. 2019;152:264–73.
3. Shapiro OH, Kushmaro A, Brenner A. Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. ISME J. 2010;4:327–36.
4. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. Proc R Soc B Biol Sci. 2002;269:931–6.
5. Stern A, Sorek R. The phage-host arms race: shaping the evolution of microbes. BioEssays. 2011;33:43–51.
6. Rostøl JT, Marraffini L. (Ph)ighting phages: how bacteria resist their parasites. Cell Host Microbe. 2019;25:184–94.
7. Paez-Espino D, Sharon I, Morovic W, Stahl B, Thomas BC, Barrangou R, et al. CRISPR immunity drives rapid phage genome evolution in streptococcus thermophilus. MBio. 2015;6:1–9.
8. Common J, Morley D, Westra ER, Van Houte S. CRISPR-Cas immunity leads to a coevolutionary arms race between Streptococcus thermophilus and lytic phage. Philos Trans R Soc B Biol Sci. 2019;374:20180098.
9. Flores CO, Valverde S, Weitz JS. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. ISME J. 2013;7:520–32.
10. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 2005;13:278–84.
11. Heilmann S, Sneppen K, Krishna S. Coexistence of phage and bacteria on the boundary of self-organized refuges. Proc Natl Acad Sci USA. 2012;109:12828–33.
12. Patterson AG, Jackson SA, Taylor C, Evans GB, Salmond GPC, Przybilski R, et al. Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems. Mol Cell. 2016;64:1102–8.
13. Alseth EO, Pursey E, Luján AM, McLeod I, Rollie C, Westra ER. Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. Nature. 2019;574:549–52.
14. Sousa JAM, de, Rocha EPC. Environmental structure drives resistance to phages and antibiotics during phage therapy and to invading lysogens during colonisation. Sci Rep. 2019;9:1–13.
15. Scanlan PD. Bacteria–bacteriophage coevolution in the human gut: implications for microbial diversity and functionality. Trends Microbiol. 2017;25:614–23.
16. Dang VT, Sullivan MB. Emerging methods to study bacteriophage infection at the single-cell level. Front Microbiol. 2014;5:724.
17. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev. 2016;40:258–72.
18. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709–12.
19. McGinn J, Marraffini LA. Molecular mechanisms of CRISPR–Cas spacer acquisition. Nat Rev Microbiol. 2019;17:7–12.
20. Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. Science. 2008;320:1047–50.
21. Sun CL, Thomas BC, Barrangou R, Banfield JF. Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. ISME J. 2016;10:858–70.
22. Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. Genome Res. 2012;22:1985–94.
23. Emerson JB, Andrade K, Thomas BC, Norman A, Allen EE, Heidelberg KB, et al. Virus-host and CRISPR dynamics in archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. Archaea. 2013;2013:370871.
24. Laanto E, Hoikkala V, Ravantti J, Sundberg LR. Long-term genomic coevolution of host-parasite interaction in the natural environment. Nat Commun. 2017;8:111.
25. Held NL, Herrera A, Quiroz HC, Whitaker RJ. CRISPR associated diversity within a population of *Sulfolobus islandicus*. PLoS ONE. 2010;5:e12988.
26. Tomida J, Morita Y, Shibayama K, Kikuchi K, Sawa T, Akaike T, et al. Diversity and microevolution of CRISPR loci in *Helicobacter cinaedi*. PLoS ONE. 2017;12:e0186241.
27. Pauly MD, Bautista MA, Black JA, Whitaker RJ. Diversified local CRISPR-Cas immunity to viruses of *Sulfolobus islandicus*. Philos Trans R Soc B Biol Sci. 2019;374:20180093.
28. Richter C, Dy RL, McKenzie RE, Watson BNJ, Taylor C, Chang JT, et al. Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res. 2014;42:8516–26.
29. Watson BNJ, Vercoe RB, Salmond GPC, Westra ER, Staals RHJ, Fineran PC. Type I-F CRISPR-Cas resistance against virulent phages results in abortive infection and provides population-level immunity. Nat Commun. 2019;10:1–8.
30. Childs LM, England WE, Young MJ, Weitz JS, Whitaker RJ. CRISPR-induced distributed immunity in microbial populations. PLoS ONE. 2014;9:e101710.
31. Common J, Walker-Sünderhauf D, van Houte S, Westra ER. Diversity in CRISPR-based immunity protects susceptible genotypes by restricting phage spread and evolution. J Evol Biol. 2020;33:1097–108.
32. Payne P, Geyrhofer L, Barton NH, Bollback JP. CRISPR-based herd immunity can limit phage epidemics in bacterial populations. Elife. 2018;7:e32035.
33. Van Houte S, Ekroth AKE, Broniewski JM, Chabas H, Ashby B, Bondy-Denomy J, et al. The diversity-generating benefits of a prokaryotic adaptive immune system. Nature. 2016;532:385–8.
34. Chabas H, Lion S, Nicot A, Meaden S, van Houte S, Moineau S, et al. Evolutionary emergence of infectious diseases in heterogeneous host populations. PLoS Biol. 2018;16:e2006738.
35. Morley D, Broniewski JM, Westra ER, Buckling A, van Houte S. Host diversity limits the evolution of parasite local adaptation. Mol Ecol. 2017;26:1756–63.
36. Westra ER, Van Houte S, Gandon S, Whitaker R. The ecology and evolution of microbial CRISPR-Cas adaptive immune systems. Philos Trans R Soc B Biol Sci. 2019;374:20190101.
37. Fernández L, Rodríguez A, García P. Phage or foe: an insight into the impact of viral predation on microbial communities. ISME J. 2018;12:1171–9.

38. Pérez MV, Guerrero LD, Orellana E, Figuerola EL, Erijman L. Time Series Genome-Centric Analysis Unveils Bacterial Response to Operational Disturbance in Activated Sludge. mSystems. 2019;4: e00169–19.

39. Goodfellow M, Kumar Y, Maldonado LA. Bergey's manual of systematics of archaea and bacteria. Hoboken, New Jersey: John Wiley & Sons, Inc; 2015.

40. Drzyzga O. The strengths and weaknesses of Gordonia: a review of an emerging genus with increasing biotechnological potential. Crit Rev Microbiol. 2012;38:300–16.

41. Arenskötter M, Bröker D, Steinbüchel A. Biology of the metabolically diverse genus Gordonia. Appl Environ Microbiol. 2004; 70:3195–204.

42. de los Reyes MF, de los Reyes FL, Hernandez M, Raskin L. Quantification of Gordona amarae strains in foaming activated sludge and anaerobic digester systems with oligonucleotide hybridization probes. Appl Environ Microbiol. 1998;64:2503–12.

43. Kragelund C, Remesova Z, Nielsen JL, Thomsen TR, Eales K, Seviour R, et al. Ecophysiology of mycolic acid-containing Actinobacteria (Mycolata) in activated sludge foams. FEMS Microbiol Ecol. 2007;61:174–84.

44. Russell DA, Hatfull GF. PhagesDB: The actinobacteriophage database. Bioinformatics. 2017;33:784–6.

45. Pope WH, Mavrich TN, Garlena RA, Guerrero-Bustamante CA, Jacobs-Sera D, Montgomery MT, et al. Bacteriophages of Gordonia spp. Display a spectrum of diversity and genetic relationships. mBio. 2017;8:e01069–17.

46. Montgomery MT, Guerrero Bustamante CA, Dedrick RM, Jacobs-Sera D, Hatfull GF. Yet more evidence of collusion: a new viral defense system encoded by Gordonia phage CarolAnn. mBio. 2019;10:e02417–18.

47. Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial species, culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. Microbe. 2014;9:111–7.

48. Quince C, Delmont TO, Raguideau S, Alneberg J, Darling AE, Collins G, et al. DESMAN: a new tool for de novo extraction of strains from metagenomes. Genome Biol. 2017;18:181.

49. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246–51.

50. Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC Genomics. 2016;17:356.

51. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–41.

52. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.

53. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

54. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.

55. Pope WH, Jacobs-Sera D. Annotation of bacteriophage genome sequences using DNA master: an overview. Methods Mol Biol. 2018;1681:217–29.

56. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.

57. Skennerton CT, Imelfort M, Tyson GW. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. Nucleic Acids Res. 2013;41:e105.

58. Rho M, Wu YW, Tang H, Doak TG, Ye Y. Diverse CRISPRs evolving in human microbiomes. PLoS Genet. 2012;8:e1002441.

59. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.

60. Keck F, Rimet F, Bouchez A, Franc A. Phylosignal: an R package to measure, test, and explore the phylogenetic signal. Ecol Evol. 2016;6:2774–80.

61. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25:1043–55.

62. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35:725–31.

63. Lei J, Sun Y. Assemble CRISPRs from metagenomic sequencing data. Bioinformatics. 2016;32:i520–8.

64. Lam TJ, Ye Y. Long reads reveal the diversification and dynamics of CRISPR reservoir in microbiomes. BMC Genomics. 2019;20:567.

65. Leenay RT, Maksimchuk KR, Slotkowski RA, Agrawal RN, Gomaa AA, Briner AE, et al. Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. Mol Cell. 2016;62:137–47.

66. Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJJ. CRISPR-Cas: adapting to change. Science. 2017;356:eaal5056.

67. Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature. 2015;520:505–10.

68. Ellington MJ, Heinz E, Wailan AM, Dorman MJ, de Goffau M, Cain AK, et al. Contrasting patterns of longitudinal population dynamics and antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single center. Genome Biol. 2019;20:184.

69. Jackson SA, Birkholz N, Malone LM, Fineran PC. Imprecise spacer acquisition generates CRISPR-Cas immune diversity through primed adaptation. Cell Host Microbe. 2019;25:250–60.

70. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol. 2008;190:1390–400.

71. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc Natl Acad Sci USA. 2011;108:10098–103.

72. Iranzo J, Lobkovsky AE, Wolf YI, Koonin EV. Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context. J Bacteriol. 2013;195:3834–44.

73. Lenski RE. Coevolution of bacteria and phage: Are there endless cycles of bacterial defenses and phage counterdefenses? J Theor Biol. 1984;108:319–25.

74. Bull JJ, Christensen KA, Scott C, Jack BR, Crandall CJ, Krone SM. Phage-bacterial dynamics with spatial structure: Self organization around phage sinks can promote increased cell densities. Antibiotics. 2018;7:8.

75. Hall AR, Scanlan PD, Morgan AD, Buckling A. Host-parasite coevolutionary arms races give way to fluctuating selection. Ecol Lett. 2011;14:635–42.

76. Best A, Ashby B, White A, Bowers R, Buckling A, Koskella B, et al. Host-parasite fluctuating selection in the absence of specificity. Proc R Soc B Biol Sci. 2017;284:20171615.

77. Gómez P, Buckling A. Bacteria-phage antagonistic coevolution in soil. Science. 2011;332:106–9.

78. Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. Long-term stability and Red Queen-like strain dynamics in marine viruses. Nat Microbiol. 2020;5:265–71.

79. Lopez Pascua L, Hall AR, Best A, Morgan AD, Boots M, Buckling A. Higher resources decrease fluctuating selection during host-parasite coevolution. Ecol Lett. 2014;17:1380–8.

80. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their phage foes. Nature. 2020;577:327–36.

81. Brockhurst MA, Chapman T, King KC, Mank JE, Paterson S, Hurst GDD. Running with the Red Queen: The role of biotic conflicts in evolution. Proc R Soc B Biol Sci. 2014;281:20141382.

82. Vuono DC, Benecke J, Henkel J, Navidi WC, Cath TY, Munakata-Marr J, et al. Disturbance and temporal partitioning of the activated sludge metacommunity. ISME J. 2015;9:425–35.

83. Kuai L, Verstraete W. Ammonium removal by the oxygen-limited autotrophic nitrification- denitrification system. Appl Environ Microbiol. 1998;64:4500–6.

84. Hosseinidoust Z, Tufenkji N, van de Ven TGM. Formation of biofilms under phage predation: considerations concerning a biofilm increase. Biofouling. 2013;29:457–68.

85. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of antiphage defense systems in the microbial pangenome. Science. 2018;359:eaar4120.

86. Winter C, Bouvier T, Weinbauer MG, Thingstad TF. Trade-Offs between competition and defense specialists among unicellular planktonic organisms: the 'Killing the Winner' hypothesis revisited. Microbiol Mol Biol Rev. 2010;74:42–57.

87. Tyson GW, Banfield JF. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol. 2008;10:200–7.

88. Weinberger AD, Wolf YI, Lobkovsky AE, Gilmore MS, Koonin EV. Viral diversity threshold for adaptive immunity in prokaryotes. mBio. 2012;3:e00456–12.

89. Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial diversity. Curr Biol. 2007;17:R373–86.

90. Bendall ML, Stevens SLR, Chan LK, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 2016;10:1589–601.

91. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ Microbiol. 2006;8:732–40.

92. Ayarza JM, Erijman L. Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. Micro Ecol. 2011;61:486–95.

93. de los Reyes FL. Foam in wastewater treatment facilities. In: Handbook of hydrocarbon and lipid microbiology. Berlin Heidelberg: Springer; 2010. pp. 2401–11.

94. Petrovski S, Seviour RJ, Tillett D. Prevention of Gordonia and Nocardia stabilized foam formation by using bacteriophage GTE7. Appl Environ Microbiol. 2011;77:7864–7.

95. Liu M, Gill JJ, Young R, Summer EJ. Bacteriophages of wastewater foaming-associated filamentous Gordonia reduce host levels in raw activated sludge. Sci Rep. 2015;5:13754.

96. Curtis TP, Head IM, Graham DW. Peer reviewed: theoretical ecology for engineering biology. Environ Sci Technol. 2003;37:64A–70A.

97. Daims H, Taylor MW, Wagner M. Wastewater treatment: a model system for microbial ecology. Trends Biotechnol. 2006;24:483–9.