

# CRISPR Immunity Drives Rapid Phage Genome Evolution in *Streptococcus thermophilus*

David Paez-Espino,<sup>a</sup> Itai Sharon,<sup>a</sup> Wesley Morovic,<sup>b</sup> Buffy Stahl,<sup>b</sup> Brian C. Thomas,<sup>a</sup> Rodolphe Barrangou,<sup>c</sup> Jillian F. Banfield<sup>a</sup>

University of California, Berkeley, California, USA<sup>a</sup>; DuPont Nutrition and Health, Madison, Wisconsin, USA<sup>b</sup>; Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, North Carolina, USA<sup>c</sup>

**ABSTRACT** Many bacteria rely on CRISPR-Cas systems to provide adaptive immunity against phages, predation by which can shape the ecology and functioning of microbial communities. To characterize the impact of CRISPR immunization on phage genome evolution, we performed long-term bacterium-phage (*Streptococcus thermophilus*-phage 2972) coevolution experiments. We found that in this species, CRISPR immunity drives fixation of single nucleotide polymorphisms that accumulate exclusively in phage genome regions targeted by CRISPR. Mutation rates in phage genomes highly exceed those of the host. The presence of multiple phages increased phage persistence by enabling recombination-based formation of chimeric phage genomes in which sequences heavily targeted by CRISPR were replaced. Collectively, our results establish CRISPR-Cas adaptive immunity as a key driver of phage genome evolution under the conditions studied and highlight the importance of multiple co-existing phages for persistence in natural systems.

**IMPORTANCE** Phages remain an enigmatic part of the biosphere. As predators, they challenge the survival of host bacteria and archaea and set off an “arms race” involving host immunization countered by phage mutation. The CRISPR-Cas system is adaptive: by capturing fragments of a phage genome upon exposure, the host is positioned to counteract future infections. To investigate this process, we initiated massive deep-sequencing experiments with a host and infective phage and tracked the coevolution of both populations over hundreds of days. In the present study, we found that CRISPR immunity drives the accumulation of phage genome rearrangements (which enable longer phage survival) and escape mutations, establishing CRISPR as one of the fundamental drivers of phage evolution.

Received 17 February 2015 Accepted 17 March 2015 Published 21 April 2015

**Citation** Paez-Espino D, Sharon I, Morovic W, Stahl B, Thomas BC, Barrangou R, Banfield JF. 2015. CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. mBio 6(2):e00262-15. doi:10.1128/mBio.00262-15.

**Editor** Janet K. Jansson, Pacific Northwest National Laboratory

**Copyright** © 2015 Paez-Espino et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Rodolphe Barrangou, rbarran@ncsu.edu, or Jillian F. Banfield, jbanfield@berkeley.edu.

Bacteria-phage interactions influence microbial community structure, ecology, and evolution and are therefore important determinants of the overall functioning of natural environments (1–3). Phages rely on specific hosts for replication and are a major cause of bacterial cell death; thus, phages impact carbon and other nutrient cycles, as well as microbial population levels, turnover, and diversity (4). In the face of this pressure, bacteria have evolved sophisticated phage defenses, such as the CRISPR-Cas adaptive immune system (5), which provides sequence-specific interference against invasive nucleic acids (6, 7).

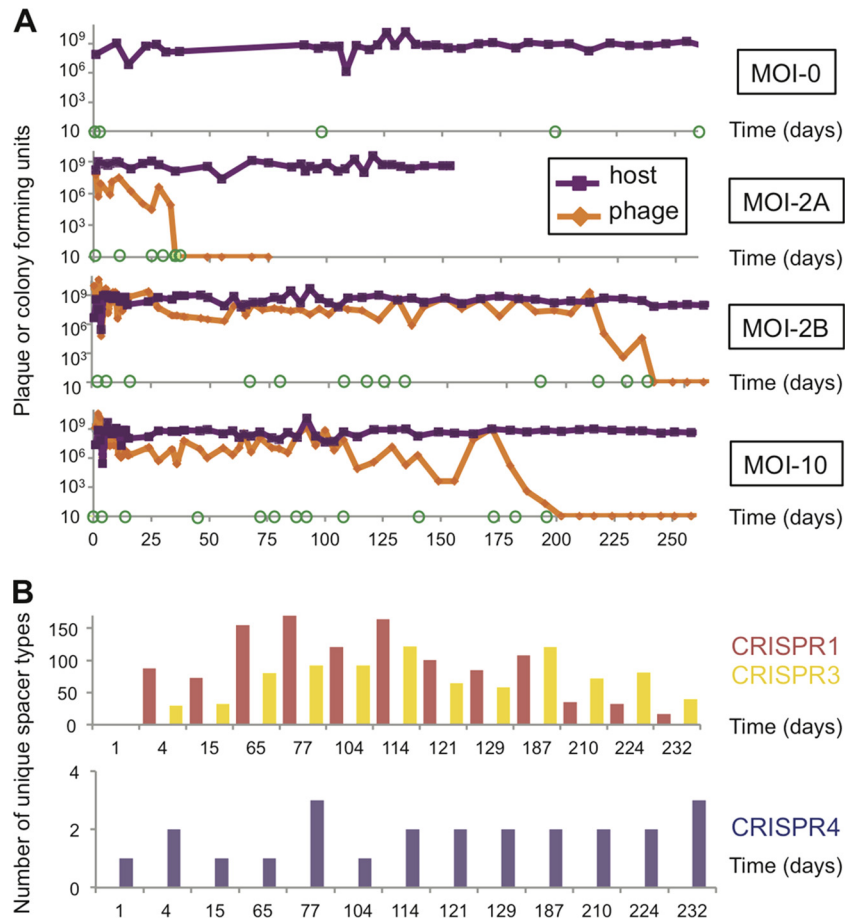
Previous studies showed that upon exposure to phage, the lactic acid bacterium model strain *Streptococcus thermophilus* DGCC7710 readily acquired spacers in two of its four CRISPR loci (CRISPR1 and CRISPR3, both classified as type IIA) to build adaptive immunity against lytic phage (8, 9).

Here, we investigate the genomic response of a phage to rapid diversification of the CRISPR spacer immunity in the host population by exposing *S. thermophilus* DGCC7710 grown in milk medium to infective lytic phage 2972 at multiplicities of infection (MOI) that trigger a novel CRISPR spacer acquisition in this system. Samples from three experimental series, and from a control series (*S. thermophilus* culture that was not inoculated with

phage), were collected periodically over time until the loss of the phage; the sampling period ranged up to 232 days. The combination of the unprecedented large amount of samples and data processed in this work shows that, in a closed system, this bacterium rapidly acquires a highly diverse set of CRISPR spacers that drive the evolution of the phage over extended time periods to extinction.

## RESULTS

**Deep sequencing of host-phage co-culture in a time series.** DNA was extracted from a subset of four experimental series. We used lytic phage 2972 MOIs that trigger novel CRISPR spacer acquisition in *S. thermophilus* and are commonly used to generate bacteriophage-insensitive mutants in dairy starter cultures (9). Thus, we used a phage-to-host abundance ratio of 2 (MOI-2) in two replicate experiments (MOI-2A and MOI-2B) and also a phage-to-host abundance ratio of 10 (MOI-10). Samples from these three experimental series, and from an *S. thermophilus* culture that was not inoculated with phage (MOI-0), were selected based on dynamically interesting growth curve fluctuations of culture/co-culture experiments distributed across time, covering a wide range of host:virus ratios, until the loss of the phage



**FIG 1** Host-phage dynamics and CRISPR spacer acquisition. (A) Fluctuations in the numbers of host bacteria (in CFU per milliliter) and phage (in PFU per milliliter) in the serial transfer experiment, over time (days), for four experimental series with different MOIs (host:phage ratios of 1:2 [MOI-2A and MOI-2B], 1:10 [MOI-10], and a no-phage control [MOI-0]). Samples were transferred daily as a 1% (vol/vol) inoculum. Time points for which samples were subjected to deep-sequencing analyses are indicated by green dots. (B) Number of unique CRISPR spacer types (defined by sequence) in the CRISPR1 and CRISPR3 systems, oth type IIA systems, and CRISPR4 locus, classified as a type IE system.

(Fig. 1A). MOI-0, MOI-2A, MOI-2B, and MOI-10 samples (Fig. 1) were subjected to Illumina metagenomic sequencing (165 Gb of data from 36 samples) (see Table S1 in the supplemental material). Host and phage counts were monitored over the course of the experiments (see Materials and Methods). After 5 weeks of co-culture (day 35), we were unable to detect phage particles in the MOI-2A experiment (Fig. 1A). In contrast, the phage and host coexisted for 232 days in the MOI-2B experiment (Fig. 1A) and for 195 days in the MOI-10 experiment (Fig. 1A).

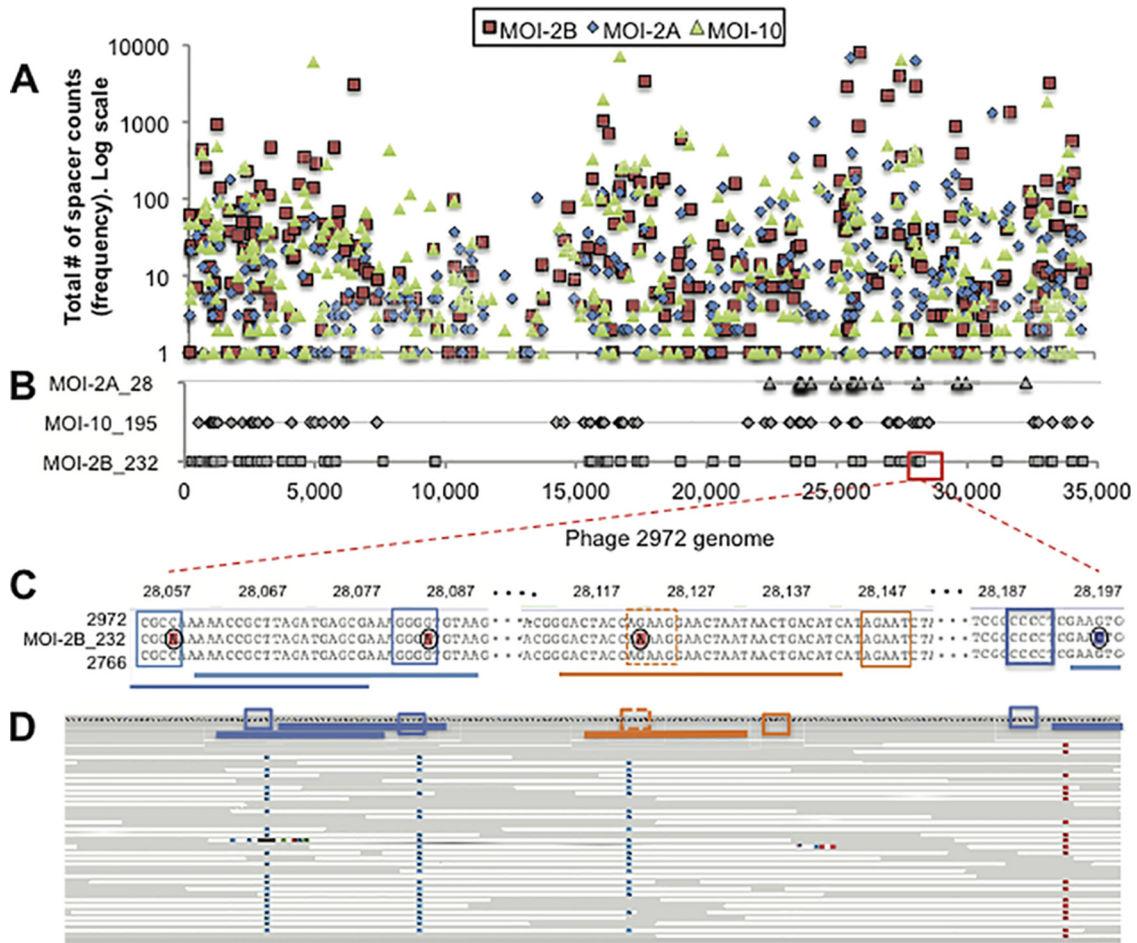
***S. thermophilus* DGCC7710 CRISPR spacer analysis.** A time series metagenomic analysis was used to document the inventory of spacer sequences added to each locus across the experimental series. There was underrepresentation or no representation of some potential spacer sequences predicted from the phage genome (proto-spacers), consistent with findings of a prior study (10). Interestingly, we detected acquisition of spacers into the type IE CRISPR4 locus, as well the type IIA CRISPR1 and CRISPR3 loci (Fig. 1B). We observed that spacers mapped unevenly over the phage genome (Fig. 2A).

Although CRISPR4 was recently shown to be biochemically active *in vitro* (11, 12), and the associated Cas proteins have been induced in the presence of phages (13), spacer acquisition has not

been reported previously. Intriguingly, none of the 71 new spacer sequences added to CRISPR4 targeted phage 2972. However, one newly acquired CRISPR4 spacer targeted other *S. thermophilus* phages, hinting at the presence of several additional low-abundance phages in the experimental system. The majority of the CRISPR4 spacers targeted unidentified DNA.

Interestingly, in the control series (MOI-0), both CRISPR1 and CRISPR3 acquired one unique new spacer, although the culture was not inoculated with phage. These spacers target host chromosomal regions. Host chromosomal targeting was also detected in the three other experimental series. Overall, 0.01%, 0.03%, and 0.02% of newly acquired spacers in series MOI-2A, MOI-2B, and MOI-10, respectively, targeted the host genome (consistent with a previous report of a frequency of  $\sim 4 \times 10^{-4}$  self-targeting immunization events [10]), and each spacer was only detected at one time point.

Using information about host genome sampling depth and the number of spacers sequenced, we estimated the average length of each CRISPR array at each genomically sampled time point. Results showed that over the course of the experiments, locus length fluctuated, with selection of variants with up to  $\sim 11$  new spacers in CRISPR1. Typically, CRISPR3 variants were estimated to have



**FIG 2** Phage genome sampling and mutations in response to CRISPR immunization. (A) The frequency of CRISPR spacer incorporation events from the phage 2972 genome, across the three experimental series, measured by the sequencing of novel spacers acquired in the host active CRISPR loci. Note that the x axis is the same for panels A and B. (B) Mutations observed in the surviving phage population at the last time point, sampled in the three experimental series. (C) Examples of mutations in the surviving phage population that enable the phage to evade CRISPR immunization. SNPs are localized in the proto-spacer (underlined sequences) and the PAM (boxed sequences). Perfect PAMs are represented in boxes with solid lines. An imperfect hypothetical PAM is shown in a box with dashed lines. Examples of fixed SNPs circumventing targeting by CRISPR1 (orange) or CRISPR3 (blue) are shown. Proto-spacers that extend off the ends of the figure were truncated. (D) Screen capture of Strainer program (31) of assembled scaffolds (white) mapped to the reference 2972 phage genome. The region displayed corresponds to the same shown in panel C with the same color code. Colored bases indicate differences in the nucleotide sequence from that of 2972 (SNPs).

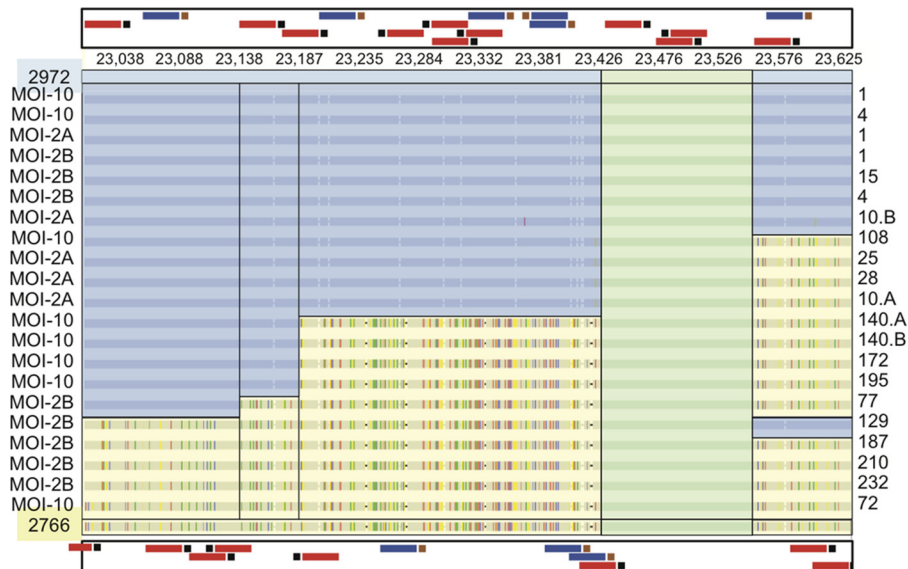
had one or two newly acquired spacers, although evidence pointed to selection for variants with up to five new spacers at some time points. In contrast, the type IIIA CRISPR2 and CRISPR4 typically lacked newly acquired spacers (see Table S2 in the supplemental material). The four CRISPR loci did not exhibit wild-type parental CRISPR spacer loss, as previously documented (14, 15).

**Sensitivity of the approach.** A subset of spacers acquired in CRISPR1 and CRISPR3 in the MOI-2B and MOI-10 series, representing 13.5% of the total set of 160,000, did not target phage 2972, but instead had sequence identity to phage 2766, a closely related phage family used in other experiments in the same laboratory (see Table S3 in the supplemental material). Thus, we inferred that, although careful procedures were followed, a phage closely related to 2766 had migrated into these experiments, as is often the case in such studies.

**Phage mutations and mutation rates.** *In silico* analyses revealed the presence of many fixed (100%) or nearly fixed (70 to 99%) single nucleotide polymorphisms (SNPs) in the phage ge-

nome (e.g., ~0.3% of sites over ~1,600 bacterial generations in the MOI-2B experiment). Notably, fixed SNPs were detected only in regions targeted by CRISPR spacers (proto-spacers) or in their associated proto-spacer adjacent motifs (PAMs) (Fig. 2B). Prior lower-scale experimental studies (16–18) had established that mutations in these regions allow the phage to evade the CRISPR-Cas machinery that is responsible for recognition and cleavage of the cognate phage sequence. The localization of fixed SNPs to exclusively proto-spacers and PAMs over the extended periods of time surveyed in this study demonstrates a strong and direct link between CRISPR targeting and phage genome evolution.

We calculated the rates of accumulation of CRISPR escape mutations in the phage genome in the MOI-2A, MOI-2B, and MOI-10 series experiments to be  $4 \times 10^{-6}$ ,  $1 \times 10^{-6}$ , and  $2 \times 10^{-6}$  substitutions per nucleotide per host generation, respectively. In comparison, the rate of SNP fixation in the host genome, calculated by comparing the genome sequence at the last time point to that of the reference genome draft, was  $4 \times 10^{-9}$  and  $8 \times$



**FIG 3** Phage genome rearrangements to escape CRISPR targeting. Blocks of hypervariability in the phage genome reflect recombination events that have eliminated wild-type sequences heavily targeted by active CRISPR1 and CRISPR3 loci in the host. (Top) Proto-spacers targeted by CRISPR loci in the 23.0- to 23.6-kb region of the reference phage 2972 genome. (Bottom) Proto-spacers targeted by CRISPR loci in the 23.0- to 23.6-kb region of the phage 2766 genome. PAMs and proto-spacers are colored according to loci (brown square and blue line for CRISPR1, and black square and red line for CRISPR3). The regions in blue correspond to the wild-type 2972 phage (top), and the regions in yellow correspond to the variant phage 2766 (bottom), from which blocks were acquired through recombination events. Sequence that is identical in both genotypes is shown in green. (Left) Experimental series from which the chimeric phages were isolated. (Right) Time points (days) at which the observed chimera were sequenced and assembled. Blocks of conserved sequences are shown in similar colors.

$10^{-10}$  fixed substitutions per nucleotide per generation for the MOI-2B and MOI-10 series, respectively. Intriguingly, not a single fixed SNP was observed in the host genome in the MOI-0 or the MOI-2A series (in which the phage was lost by day 35).

Previously, we noted that spacers that were anomalously highly represented in the CRISPR-Cas locus of *S. thermophilus* tended to be clustered on the phage genome (10). Here, we report a correlation between such regions and elevated rates of SNP fixation in the phage genome (Fig. 2C and D). Mutations that could enable phage to escape CRISPR silencing could occur in the PAM (19), the ~7-bp proto-spacer “seed” region adjacent to the PAM, which drives target sequence recognition and cleavage for the interference step (16, 19–21), or elsewhere in the proto-spacer region. In the large data set generated by this experiment, we noted that fixed SNPs statistically significantly accumulated in the PAMs relative to the nonseed proto-spacer region by factors of 1.86, 11.65, and 3.50 in the MOI-2A, MOI-2B, and MOI-10 experiments, respectively. SNPs were also preferentially fixed in the seed regions relative to the nonseed proto-spacer region by factors of 2.06, 7.05, and 2.22 in the MOI-2A, MOI-2B, and MOI-10 experiments, respectively (see Table S4 in the supplemental material).

**Reconstruction of phage genomes and recombination events.** A second mechanism of escape of CRISPR immunity was identified by comparing phage genomes reconstructed from metagenomic data with the parental viral genomes (the original 2972 phage and later-introduced phage 2766) (Fig. 3). At most time points, the samples were surprisingly dominated by phages with genomes that were chimeras of the 2972 and 2766 genotypes (Fig. 3 and 4). Strikingly, at least 16 different recombination boundaries were identified (see Table S5 in the supplemental material). Notably, the genome sequence blocks from phage 2766 replaced blocks of phage 2972 that were specifically highly tar-

geted by the set of novel CRISPR spacers acquired in the host genome at that time point (Fig. 2 and 4).

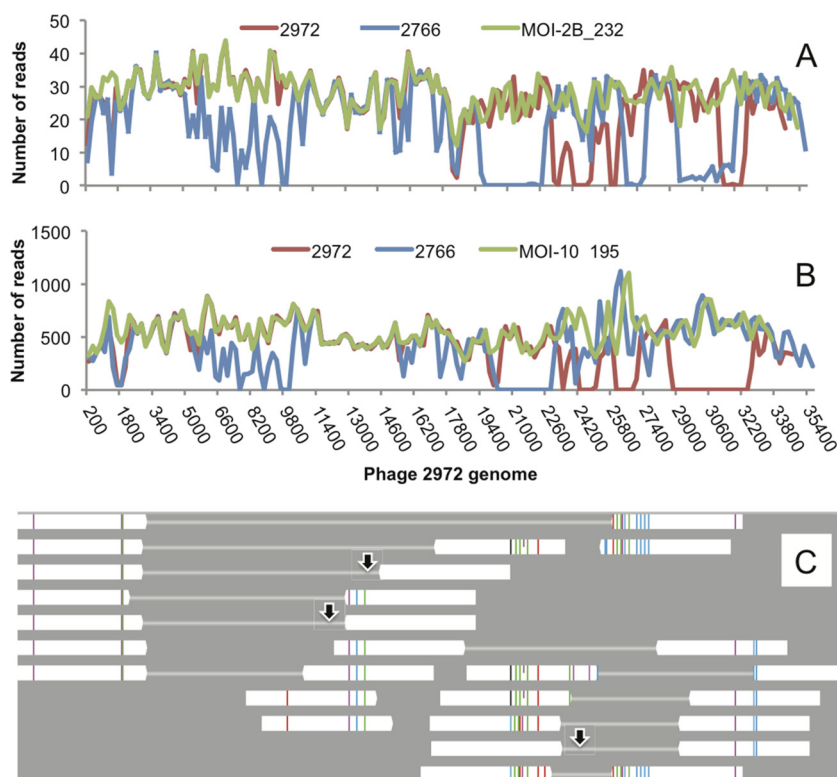
A phylogenetic tree built using all the reconstructed phage genomes demonstrated the dynamic nature of the phage genomes over the time scales of the MOI-2A, MOI-2B, and MOI-10 experiments (Fig. 5) and allowed us to detect the presence of the second phage (2766) from early stages of the sampling for MOI-2B and MOI-10 (around day 4 and day 15, respectively).

**Phage loss and phage “rescue.”** Because the assessment of the hypothesis that sustainable phage coexistence with a single host with active CRISPR-Cas systems requires the presence of multiple phage types, we conducted two additional experiments. First, we tested the hypothesis that phage loss was predictable in the single-phage MOI-2A experiment. To do this, we reinoculated samples archived from time points preceding phage loss and repeatedly observed the disappearance of the phage population (Fig. 6). In a second series of experiments, we added a second phage (recovered from the series MOI-2B\_41) to a sample archived from a time point before the loss of the phage in the MOI-2A experiments to determine whether the second phage population could “rescue” the first. Results showed that an increase in phage genotypic diversity could extend the phage-versus-host “arms race” (see Fig. S1 in the supplemental material).

## DISCUSSION

Although the CRISPR-Cas system has been extensively described as a resistance mechanism against phages, it has remained unclear how phage genomes evolve in response to CRISPR immunity. We measured phage evolutionary rates by applying metagenomics time-series analyses to samples from long-term experiments that included *S. thermophilus* DGCC7710 and phage 2972. Interestingly, we identified a migrant phage in two of the experimental





**FIG 4** Phage genome re-arrangements to escape CRISPR targeting and recombination events between phages. (A) The surviving phage genome at transfer 232 in the MOI-2B series. The red line reflects coverage consensual with the WT 2972 phage, whereas the blue line reflects coverage consensual with the 2766 phage, from which sequences were acquired through recombination events to generate the chimeric phage (green). (B) Assembly of the surviving phage genome at transfer 195 in the MOI-10 series. Color codes are as used in panel (A). Baseline coverage reflects boundaries of recombination events (i.e. 28k-33k in the lower panel). (C) Screen capture of Strainer program 29 of sequencing reads (in white) mapped to the reference phage genome. Arrows mark sequences that are hybrid between the reference genome and the mutated phage genomes. These patterns suggest that recombination events introduced polymorphic blocks. Colored tick marks indicate differences in nucleotide sequence from the reference. Blue, red, purple, and green ticks indicate substitutions for the bases A, C, T and G respectively. Ticks of half height indicate extra bases in a read sequence, and missing bases are colored black.

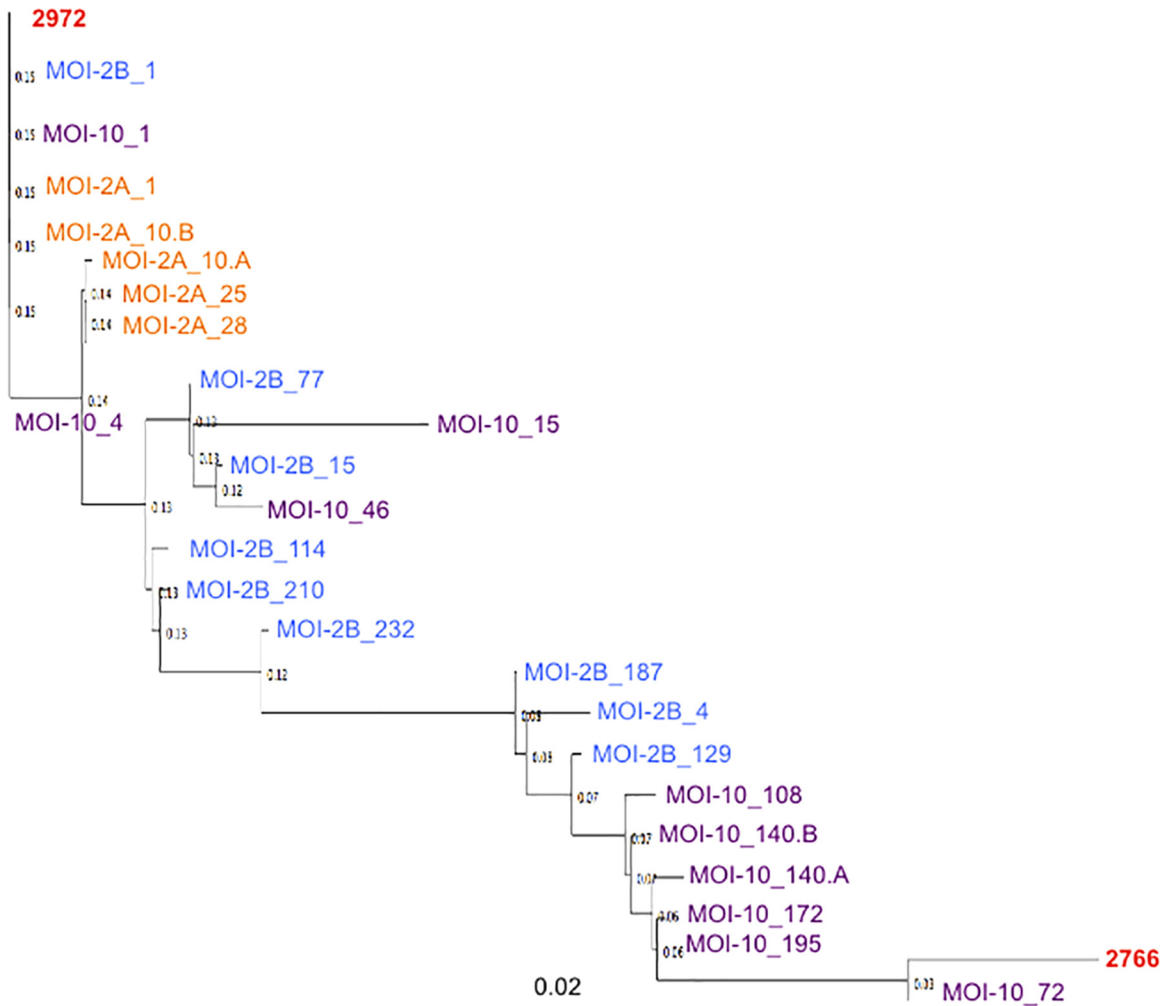
series. Such events probably occur commonly in long-term experiments but would normally be undetectable without long-term monitoring and deep sequencing. By implication, metagenomic analysis of host CRISPR spacer complements provided a sensitive method by which potentially very rare phages in the environment of a bacterial culture can be detected. This finding has implications for bacterial source tracking applications.

Within the inventory of over 160,000 spacers recovered across all the samples, we detected extensive novel spacer acquisition in the CRISPR1 and CRISPR3 loci, as well as rare acquisition events for the CRISPR4 locus, a novel type I system for which we observed activity for the first time. The fastest locus expansion rate occurred in CRISPR1, with addition of one spacer added almost every generation early in the experiment. Only one spacer every ~7 generations was added to CRISPR3 over the same time period. Both CRISPR2 and CRISPR4 typically lacked any newly acquired spacers, although 71 CRISPR4 spacers were detected across all the experiments. We attribute the unusual detection of spacer acquisition in CRISPR4 to the long experimental timescale and the deep sequencing used in this study. The bulk of the CRISPR4 spacers targeted unidentified DNA (probably trace milk-associated sequences). Spacers targeting the bacterial genome were also identified, but they were both rare and lethal (10, 22, 23).

We measured the rate of SNP fixation in the host genome by

comparing the bacterial genome sequence at the last time point to that of the reference draft genome. The numbers of fixed substitutions per nucleotide per generation for the MOI-2B and MOI-10 series were very low ( $4 \times 10^{-9}$  and  $8 \times 10^{-10}$ , respectively). Thus, the host genome evolves much more slowly than the phage genome counterpart. Intriguingly, not even one fixed SNP was observed in the host genome in the MOI-0 or the MOI-2A series (from which the phage was lost by day 35 and no immigrant phage 2766 was present). This observation hints that the presence of phages may accelerate host genome evolution.

To ensure persistence, we postulated that phage genomes must circumvent spacer targeting at rates that are comparable to those of spacer acquisition. We noted that SNPs are preferentially fixed in the PAM relative to the nonseed proto-spacer region. This is consistent with previous findings (9, 19) and the observation that Cas9 first binds a PAM within the phage DNA prior to interrogation of adjacent phage DNA by the CRISPR spacer sequence (24). SNPs are also preferentially fixed in the seed regions relative to the nonseed proto-spacer regions. Given the role of the PAM in spacer sampling from the phage genome and in subsequent phage genome targeting and cleavage by Cas9, and also the significant role of the seed sequence for targeting (16, 20), this result further elaborates the strong link between CRISPR immunity and phage genome evolution demonstrated in the present study. The apparent



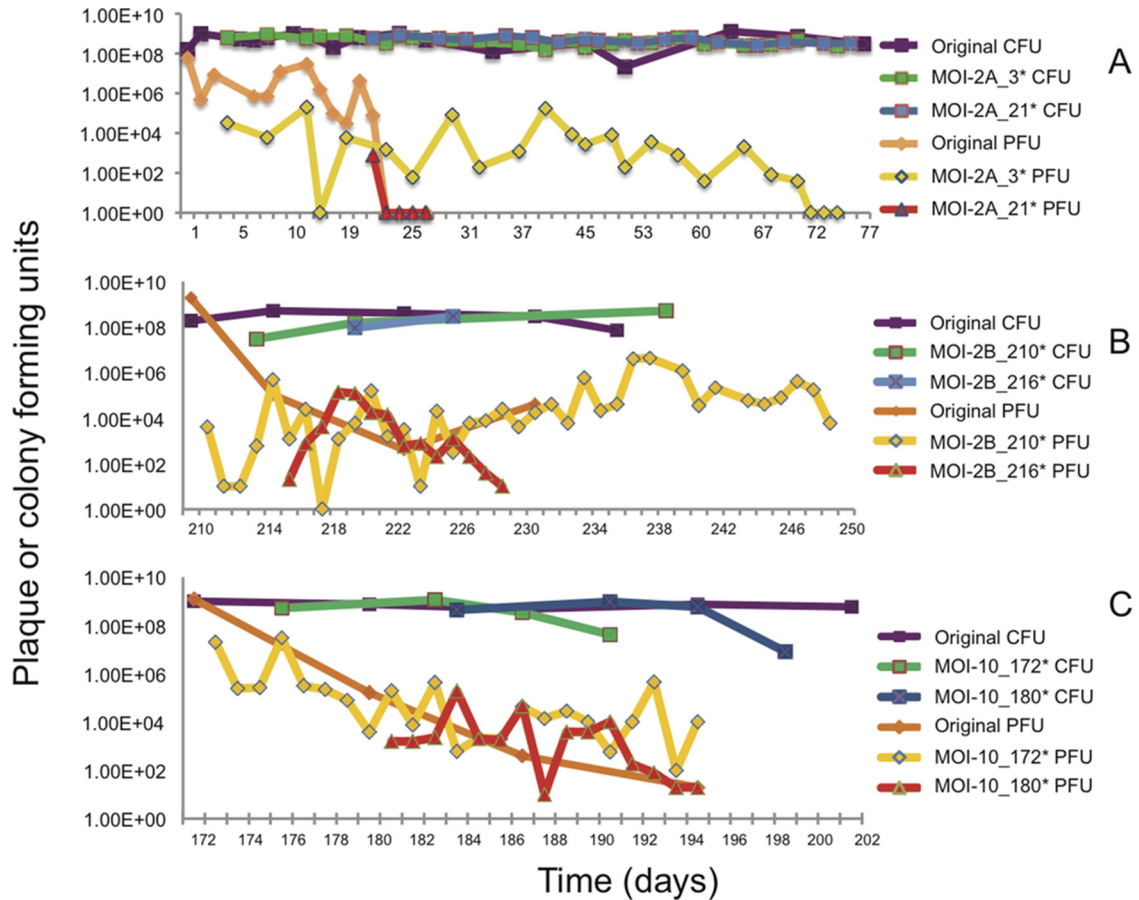
**FIG 5** Phylogenetic tree of the phage genotypes. Whole-genome phylogenetic tree (using PhymI over multiple Muscle Alignment, with phage 2972 as the reference) showing the relatedness of the various dominant chimeric phages across the experimental series over time, and their similarity to the WT 2972 and 2766 phage (both in red). In orange, MOI-2A series; Blue: MOI-2B series; Purple: MOI-10 series. The last digit for each sample indicates the time point from which dominant phage genomes were assembled (in case there were 2 dominant phage in a time point, they will be shown as A or B). Noteworthy, all genomes from the first time point in each series cluster with the WT 2972 phage, whereas the phage reconstructed from the last time point in each series (MOI-2A\_28; MOI-2B\_232; MOI-10\_195) are less phylogenetically related.

selection for phage genomes with PAM compared to “seed” mutations may arise because these mutations can prevent both spacer acquisition and interference (whereas seed sequence mutations only impact the latter).

The phage persistence in the two experiments that contained multiple phages, but not in the experiment with a single phage (MOI-2A) (Fig. 1), may not be a coincidence. We identified a second mechanism by which phage escape CRISPR immunity that involves formation of chimeric genomes (e.g., involving the 2972 and 2766 genotypes in experiments MOI-2B and MOI-10). Phage genome mosaicism has been widely reported (25), and a link between recombination and phage evasion of CRISPR immunity was suggested previously, based on metagenomic analysis of natural phage population sequence information (26). The findings of the current study directly constrain the rate at which selection for (potentially rare) recombination events can occur and experimentally demonstrate recombination as a mechanism by which phages specifically circumvent CRISPR-conferred immunity.

Phage loss was predictable and reproducible in experiments that contained only a single phage (MOI-2A experiment). The “tipping point” probably occurs when a sufficiently high level of host immune diversity is achieved. Addition of a second, distinct phage type (from the series MOI-2B\_41) extended phage persistence and, thus, the time period of phage-host coevolution. The increased phage predation pressure may dilute the ability of *S. thermophilus* to assemble sufficient immunity to eliminate any single phage population. Because all these experiments eventually ended with phage loss (Fig. 1 and 6; see also Fig. S1 in the supplemental material), continual immigration of new phage populations may be important to extend the bacteria-phage arms race indefinitely in natural systems.

By directly and quantitatively linking phage genome modification to the selective pressure associated with CRISPR-Cas immunity, this study establishes CRISPR adaptive immunity of *S. thermophilus* as a fundamental driver of phage genome evolution. When present and active, CRISPR-mediated immunity influences



**FIG 6** Host-phage population dynamics in re-inoculation experiments. Fluctuation in the number of host (colony forming units per ml) and phage (plaque forming units per ml) in the serial transfer experiment, over time (days). The “asterisked” series represent re-inoculation of the samples at defined time points, after 3 and 21 transfers in the MOI-2A series (3\* and 21\*, respectively—top panel); after 210 and 216 transfers in the MOI-2B series (210\* and 216\*, respectively—centered panel); and after 172 and 180 transfers in the MOI-10 series (172\* and 180\*, respectively—bottom panel). Data shows that a tipping point has been reached for phage loss over time at the latest re-inoculated time point in each series (shown in red in each panel).

the accumulation of point mutations and selects for advantageous homologous recombination events between genetically related phages. The selective advantage of reduced susceptibility to CRISPR silencing could explain why recombination is a common evolutionary phenomenon in natural phage populations.

## MATERIALS AND METHODS

**Model system: overview of the experimental design, sampling scheme, and sequence analysis pipeline.** *Streptococcus thermophilus* strain DGCC7710 and co-culture of DGCC7710 with phage 2972 were grown, incubated, transferred, and stored as described in reference 13. We used the procedures from this study for the determination of CFU and PFU and detection of phage 2766. The *Streptococcus thermophilus* host was cultured over time (MOI-0 series) or challenged with lytic phage 2972 at two different multiplicities of infection (host:phage ratios of 1:2 for the MOI-2 series [MOI-2A and MOI-2B] and 1:10 for the MOI-10 series) (Fig. 1). Then, samples were transferred daily as a 1% (vol/vol) inoculum. Host counts (in CFU per milliliter) and phage counts (in PFU per milliliter) were monitored over time, and samples obtained at various time points were subjected to DNA extraction (50-ng minimum) and subsequently used as the templates for library preparation and deep sequencing using 100-bp paired-ends sequencing with the Illumina system. (Selection of the time points was based on dynamically interesting growth curve fluctua-

tuations of culture/co-culture experiments distributed across time and covering a wide range of host:virus ratios.)

**Metagenomic sequencing.** DNA samples were subsequently used as the templates for library preparation and deep sequencing by using 100-bp paired-ends sequencing with the Illumina system. A total of 165 Gb of sequence data were generated and subjected to bioinformatic analyses, including host CRISPR locus spacer detection, host and phage genome assembly, and comparison with wild-type sequences for the host chromosome, phage 2972, and phage 2766 sequences. Finally, postassembly as well as comparative analyses were performed to identify SNPs, indels, and recombination events.

**Computational analyses.** Total genomic DNA (bacterium and phage) extracted at various time points from all experimental series was sequenced using Illumina high-throughput technology at the WM Keck Center for Comparative and Functional Genomics (University of Illinois at Urbana---Champaign) (for details, see Table S1 in the supplemental material). Reads were quality filtered by trimming both ends, using sickle (a program available at the GitHub, a Web-based Git repository hosting service) and deposited the sequence information with NCBI (see “Nucleotide sequence accession numbers,” below). Only paired reads were used in the assemblies. Assemblies were evaluated using *idba\_ud* (27) and default parameters.

Phage genomes were reconstructed from each sample separately as follows. First, all reads that belonged to the genome of *S. thermophilus*

DGCCC7710 were removed based on alignment to the reference genome using Bowtie (28). Remaining reads were assembled using Velvet (29) with parameters were adjusted to the expected coverage, considering the number of reads and the length of phage 2972. Only a subset of the reads was used when the expected coverage exceeded  $\sim 200\times$ . For phage genome analysis, we used miniassembly procedures described previously (30).

SNPs were identified by using a program that takes read mapping information as input and computes base calling for every position on the target genome. C++ source code for the program is available at <https://github.com/CK7/SNPs>.

Spacer sequences were extracted using a custom Ruby script (included in SOM) that searched for each of the exact repeat sequences from each of the CRISPR loci, as well as their reverse complement sequences in the full read set. We grouped the spacers that shared 85% in length and 85% sequence identity to avoid overrepresentation of spacer types due to sequencing errors.

**Nucleotide sequence accession numbers.** The sequences identified in this work have been deposited with NCBI under Bioproject number PRJNA275232 and SRA main accession number SRP055779.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00262-15/-DCSupplemental>.

- Figure S1, TIF file, 1.5 MB.
- Table S1, PDF file, 0.1 MB.
- Table S2, PDF file, 0.1 MB.
- Table S3, PDF file, 0.05 MB.
- Table S4, PDF file, 0.2 MB.
- Table S5, PDF file, 0.05 MB.

## ACKNOWLEDGMENTS

This work was supported by grant no. W911NF-10-0046 from the Army Research Office and DuPont Nutrition and Health.

We thank Alvaro Hernandez (Keck Center at the University of Illinois at Urbana-Champaign) for assistance with sequencing.

D.P.-E. conducted experiments and performed data analysis. W.M. and B.S. conducted experiments and collected samples. I.S. and B.C.T. provided analytical methods and performed data analysis. D.P.-E., R.B., and J.F.B. designed the study, interpreted the data, and wrote the manuscript.

## REFERENCES

1. Pal C, Maciá MD, Oliver A, Schachar I, Buckling A. 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450:1079–1081. <http://dx.doi.org/10.1038/nature06350>.
2. Levin BR, Bull JJ. 1994. Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol* 2:76–81. [http://dx.doi.org/10.1016/0966-842X\(94\)90538-X](http://dx.doi.org/10.1016/0966-842X(94)90538-X).
3. Gómez P, Buckling A. 2011. Bacteria-phage antagonistic coevolution in soil. *Science* 332:106–109. <http://dx.doi.org/10.1126/science.1198767>.
4. Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <http://dx.doi.org/10.1038/nrmicro1750>.
5. Jansen R, Embden JD, Gastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43:1565–1575. <http://dx.doi.org/10.1046/j.1365-2958.2002.02839.x>.
6. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712. <http://dx.doi.org/10.1126/science.1138140>.
7. Garneau JE, Dupuis MÉ, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67–71. <http://dx.doi.org/10.1038/nature09523>.
8. Horvath P, Romero DA, Coûté-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R. 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 190:1401–1412. <http://dx.doi.org/10.1128/JB.01415-07>.
9. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190:1390–1400. <http://dx.doi.org/10.1128/JB.01412-07>.
10. Paez-Espino D, Morovic W, Sun CL, Thomas BC, Ueda K, Stahl B, Barrangou R, Banfield JF. 2013. Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* 4:1430. <http://dx.doi.org/10.1038/ncomms2440>.
11. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 30:1335–1342. <http://dx.doi.org/10.1038/emboj.2011.41>.
12. Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, Siksnys V. 2013. In vitro reconstitution of cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J* 32:385–394. <http://dx.doi.org/10.1038/emboj.2012.352>.
13. Young JC, Dill BD, Pan C, Hettich RL, Banfield JF, Shah M, Fremaux C, Horvath P, Barrangou R, Verberkmoes NC. 2012. Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS One* 7:e38077. <http://dx.doi.org/10.1371/journal.pone.0038077>.
14. Levin BR, Moineau S, Bushman M, Barrangou R. 2013. The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLoS Genet* 9:e1003312. <http://dx.doi.org/10.1371/journal.pgen.1003312>.
15. Barrangou R, Coûté-Monvoisin AC, Stahl B, Chavichvily I, Damange F, Romero DA, Boyaval P, Fremaux C, Horvath P. 2013. Genomic impact of CRISPR immunization against bacteriophages. *Biochem Soc Trans* 41:1383–1391. <http://dx.doi.org/10.1042/BST20130160>.
16. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* 108:10098–10103. <http://dx.doi.org/10.1073/pnas.1104144108>.
17. Sapranaukas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 39:9275–9282. <http://dx.doi.org/10.1093/nar/gkr606>.
18. Mojica FJ, Díez-Villaseñor C, García-Martínez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155:733–740. <http://dx.doi.org/10.1099/mic.0.023960-0>.
19. Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF. 2013. Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* 15:463–470. <http://dx.doi.org/10.1111/j.1462-2920.2012.02879.x>.
20. Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck A, Boekema EJ, Dickman MJ, Doudna JA, Boekema EJ, Boekema E, Dickman MJ, Dickman M, Doudna JA. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* 108:10092–10097. <http://dx.doi.org/10.1073/pnas.1102716108>.
21. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821. <http://dx.doi.org/10.1126/science.1225829>.
22. Edgar R, Qimron U. 2010. The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J Bacteriol* 192:6291–6294. <http://dx.doi.org/10.1128/JB.00644-10>.
23. Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, Clulow JS, Richter C, Przybyski R, Pitman AR, Fineran PC. 2013. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLOS Genet* 9:e1003454. <http://dx.doi.org/10.1371/journal.pgen.1003454>.
24. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507:62–67. <http://dx.doi.org/10.1038/nature13011>.
25. Lucchini S, Desiere F, Brüßow H. 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J Virol* 73:8647–8656.



26. Andersson AF, Banfield JF. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320: 1047–1050. <http://dx.doi.org/10.1126/science.1157358>.
27. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <http://dx.doi.org/10.1093/bioinformatics/bts174>.
28. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. <http://dx.doi.org/10.1186/gb-2009-10-3-r25>.
29. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
30. Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. 2013. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res* 23:111–120. <http://dx.doi.org/10.1101/gr.142315.112>.
31. Eppley JM, Tyson GW, Getz WM, Banfield JF. 2007. Strainer: software for analysis of population variation in community genomic datasets. *BMC Bioinformatics* 8:398.