Serine recombinases are conserved genetic markers of antiphage defense systems

Shelby E. Andersen[1], Joshua M. Kirsch[1], Navtej Singh[1], Stephen R. Garret[3,4], John C. Whitney[3,4], Jay R. Hesselberth[2], Breck A. Duerkop[1,#]

[1]Department of Immunology and Microbiology, University of Colorado School – Anschutz Medical Campus, School of Medicine, Aurora, CO, USA, 80045

[2]Department of Biochemistry and Molecular Genetics, University of Colorado School – Anschutz Medical Campus, School of Medicine, Aurora, CO, USA, 80045

[3]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada, L8S 4L8

[4]Michael DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada L8S 4L8

[#]Correspondence: Breck A. Duerkop, breck.duerkop@cuanschutz.edu

Abstract

Antiphage defense systems confer bacteriophage (phage) resistance in bacteria. Renewed interest in phage therapy indicates a need to understand the breadth and molecular mechanisms of antiphage defenses. Traditionally, strategies to identify antiphage defenses lack throughput or are biased toward model bacteria. Herein, we developed a bioinformatic pipeline that uses a serine recombinase to identify known and unknown antiphage defense systems. Using this approach to query reference genomes and metagenomes, we show that serine recombinase genes are genetically linked to antiphage defense systems and serve as bait for finding these systems across diverse bacterial phyla. Using co-transcription predictions and statistical analysis of protein domain abundances, we experimentally validated our informatic approach by discovering that KAP P-loop NTPases are fused to putative antiphage effector domains and prokaryotic Schlafen proteins support phage defense. Our work shows that serine recombinases are a reliable genetic marker for the discovery of antiphage defenses across diverse bacterial phyla.

54 **Introduction**

55 Multi-drug resistant (MDR) bacterial infections are a serious public health threat. With limited
56 treatment options available for these infections, bacteriophage (phage) therapy has re-emerged as a
57 potential treatment option[1-3]. Antiphage defense systems adapt to phages to overcome predation which
58 threatens the efficacy of phage therapies[4]. Therefore, understanding the breadth, activity, and evolution
59 of antiphage defenses in bacteria will inform the development of successful phage therapies.

60 Antiphage defense systems are routinely discovered by the tendency of these systems to localize
61 in so-called 'defense islands', wherein uncharacterized phage defense systems are identified by their
62 proximity to other known defense systems, or by generating genomic libraires and screening these in a
63 heterologous bacterial host for phage defense[5-9]. To date, a marker gene that is broadly associated with
64 various antiphage defenses across diverse bacteria has not been identified. Additionally, our knowledge
65 of antiphage defense systems stems mostly from model bacteria such as *Escherichia coli* and *Bacillus*
66 *subtilis*[5,8,10-13]. Well-studied antiphage defenses, such as abortive infection, restriction modification, and
67 CRISPR/Cas, are widespread throughout bacterial phyla; however, the majority of the more recently
68 discovered antiphage defense systems are poorly represented among diverse phyla and are biased to
69 the Proteobacteria[14]. This indicates that utilizing defense islands or random cloning as the primary means
70 of discovering novel systems likely excludes much of the diversity in antiphage defense biology and
71 necessitates deeper studies across the bacterial domain of life.

72 Mobile genetic elements (MGEs), including plasmids, temperate phages, and transposons are
73 significant drivers of genetic diversity in bacteria. MGEs are capable of spreading genetic traits that
74 support bacterial fitness, such as antibiotic resistance genes and bacterial immunity genes to toxins[15,16].
75 Recent studies show that antiphage defense systems localize within the boundaries of MGEs[17-20]. For
76 example, temperate phages and other phage-like elements are "hotspots" for antiphage defenses, and
77 mobilizable antibiotic resistance cassettes can harbor and disseminate antiphage defenses[10,21]. Finally,
78 genes that support DNA mobilization including recombinases, integrases, and transposon-encoding
79 insertion sequence elements are often embedded within or near antiphage defense islands[19,22]. This

80 indicates that broadly conserved MGE-specific genes could be leveraged as bait to identify antiphage
81 defenses beyond those already discovered.

82 We recently characterized a Type IV restriction enzyme (TIV-RE, EF_B0059) from *Enterococcus*
83 *faecalis* that is encoded on a bicistronic mRNA with a serine recombinase (EF_B0058)[23]. The TIV-RE
84 restricts phage replication in *E. faecalis*, while the co-transcribed serine recombinase is dispensable for
85 phage resistance, suggesting that instead, the serine recombinase has evolved to perform an unknown
86 function associated with the genetic maintenance of this TIV-RE defense system. Thus, we hypothesized
87 that other antiphage defense systems could be found in the genomic regions surrounding homologs of
88 these serine recombinases and may also be co-transcribed. To explore this idea, we developed a
89 bioinformatic pipeline termed Recombinase Associated Defense Search (RADS) that searches for serine
90 recombinase homologs and captures their surrounding genomic context. Using the RADS pipeline to
91 query bacterial genomes and metagenomic contigs, we identified a plethora of antiphage defense
92 systems. RADS identifies antiphage defenses spanning diverse bacterial phyla. Using binomial statistical
93 testing coupled with co-transcription prediction to prioritize genes to further interrogate, we discover
94 previously uncharacterized antiphage defense systems. We show that KAP NTPase domains regulate
95 the function of linked effector modules that mediate antiphage defense. Because it relies on a conserved
96 serine recombinase rather than known antiphage defense genes, RADS is a uniquely unbiased method
97 for locating antiphage defense systems in diverse bacteria.
98

99 **Materials & Methods**

100 **Plasmids**

101 Plasmid constructs were generated by Genscript through synthesis of the nucleotide sequence
102 of interest and cloning each gene into the plasmid pLZ12A that harbors a constitutive promoter[24]. Site
103 directed mutagenesis of antiphage defense genes were generated where mutations of interest were
104 designed into the primers used for amplification. These primers were used for whole plasmid PCR
105 amplification using Q5 High-Fidelity 2X DNA polymerase Master Mix (New England Biolabs) to create a

106 linear DNA product. PCR products were purified using a QIAquick PCR Purification Kit (Qiagen) and
107 linearized products were ligated using T4 DNA ligase (New England Biolabs). Ligated plasmids were
108 transformed into electrocompetent *E. coli* TG1 DUOs (Lucigen). Plasmids were purified from *E. coli* using
109 a Plasmid Mini Prep Kit (Qiagen) and transformed into electrocompetent *E. faecalis* using the lysozyme
110 treatment method[24]. Plasmids and primers used can be found in Supplemental Table 1.

111

112 **Bacteria and bacteriophages**

113 A list of bacterial and bacteriophage strains used in this study can be found in Supplemental Table
114 1. *E. faecalis* strains were grown with aeration in Todd Hewitt broth (THB, Becton Dickinson) or on THB
115 agar at 37°C. *Escherichia coli* was grown in Lennox lysogeny broth (LB, Fisher) with aeration or on LB
116 agar at 37°C. When necessary, for the selection of *E. coli* or *E. faecalis* carrying a plasmid of interest,
117 15 µg/mL chloramphenicol (Research Products International) was added to the media. Phage plaque
118 assays were performed using THB base and top agar supplemented with 10 mM $MgSO_4$ as described
119 elsewhere[25].

120

121 **Bacterial growth curves**

122 Overnight cultures of *E. faecalis* were grown in THB broth with 15µg/mL chloramphenicol.
123 Overnight cultures were diluted to an OD_{600} of 0.3 in THB broth with 15µg/mL chloramphenicol and 10mM
124 $MgSO_4$. 200µL of the diluted culture was added to the appropriate wells of a 96-well plate. The appropriate
125 volume of phage diluted in SM-plus buffer (100 mM NaCl, 50 mM Tris-HCl, 8 mM $MgSO_4$, 5 mM $CaCl_2$
126 [pH 7.4]) was added to the appropriate samples. Samples were set up in technical triplicate. The 96-well
127 plate was then moved to 37°C with orbital shaking. Bacterial growth was measured by optical density
128 (OD_{600}) every 5 minutes using a Biotek Synergy H1 Microplate reader.

129

130 **Bioinformatics**

131 ***Development of the Recombinase Associated Defense Search (RADS) pipeline.*** RADS first predicts
132 ORFs in genomes or metagenomic contigs and translates them to amino acid sequences using
133 prodigal[26]. Prodigal is used for de novo generation of ORFs within genomes and contigs rather than using
134 the protein files from NCBI because the nascent ORF headers have the data in an easily queried format
135 for downstream extraction. These amino acid sequences are then used to generate a BLASTp database
136 using Diamond[27]. The Diamond BLASTp database is searched for homologs of the EF_B0058 serine
137 recombinase with high (≥65%), or moderate (≥30%) amino acid identity. Once the genomic coordinates
138 of serine recombinase homologs are identified, Seqkit is used to extract the 5kB flanking DNA sequence
139 up- and downstream of the serine recombinase gene (designated as a RADS contig)[28]. RADS can be
140 tuned to extract a region of a preferred size, but defaults to 5 kb on each side of the serine recombinase
141 gene for a total of 10kb. These contigs are then run through InterProScan, returning a table of putative
142 domains and predicted structural similarities for each ORF across the 10kb region queried[29]. RADS will
143 also generate a list of predicted co-transcribed ORFs in which the downstream ORF must be on the same
144 DNA strand and within 100 base pairs of the serine recombinase gene (**Fig. S2A**). RADS has been
145 incorporated with functions so that each step of the pipeline may be independently called as long as the
146 dependent files have been generated.

147

148 ***Analysis of RADS output data.*** Binomial statistical analysis for enrichment of domains in RADS contigs
149 (**Fig. 1B, S2B**) is run in RStudio. The binomial analysis compares the frequency of a domain being found
150 in a RADS contig to its frequency in the entirety of the genome to reveal domains that are truly enriched
151 in RADS contigs (**Fig. S2B**). To run the binomial analysis, InterProScan must be run on all sequences
152 where serine recombinase homologs were found. Known antiphage systems are confirmed using
153 DefenseFinder[14]. Recombinase discovery rate was calculated by the number of recombinase homologs
154 discovered divided by the number of input sequences. Antiphage discovery rate was calculated by the
155 number of antiphage defense systems identified by DefenseFinder HMMer divided by the number of
156 contigs discovered by RADS. All protein modeling was performed with Alphafold3[30].

157

158 ***Pipeline availability.*** RADS is available for download at https://github.com/Seandersen/RADS.

159

160 ***Computational recommendations.*** As with all bioinformatic approaches, RADS exhibits computational
161 limits. While RADS was successfully run on 190,280 genomes from the Bacillota phylum for this study,
162 the pipeline took considerable computation time and resources. This is due in part to translation of each
163 coding sequence in a genome or contig, as well as the time it takes for InterProScan to run analyses on
164 each ORF for domain prediction. While possible to run RADS on an entire phylum of bacterial nucleotide
165 sequences from NCBI, the pipeline is more manageable with a smaller number of input files. This study
166 was completed on a supercomputing system with 80CPU and allocating up to 62GB of RAM. The
167 maximum cores allotted for this pipeline was 40CPU for the BLASTp step. The pipeline has been
168 incorporated with a threads option where the number of CPU to be used can be specified.

169

170 ***Metagenomics analyses.*** Assembled metagenomes were downloaded from their respective data
171 sources (Human Microbiome Project or Ma *et al*. 2023)[31,32]. Taxonomy classification was performed using
172 Kraken2[33]. The human fecal microbiome metagenomes were taxonomically classified with the Phanta
173 database, while the soil microbiome metagenome was taxonomically classified with the with the
174 Plus_PF_16 database.

175

176 **Results**

177 **Development of an informatic pipeline that identifies antiphage defense systems**

178 The observation that a TIV-RE defense system (EF_B0059) in *Enterococcus faecalis* is encoded
179 on a shared mRNA transcript with a predicted serine recombinase (EF_B0058, **Fig. 1A**)[23] led us to
180 hypothesize that serine recombinases may co-occur in genomic regions with antiphage defense genes.
181 To test this, we devised a bioinformatic approach that identifies serine recombinases homologous to the
182 *E. faecalis* serine recombinase EF_B0058, extracts the surrounding DNA sequence, and predicts the

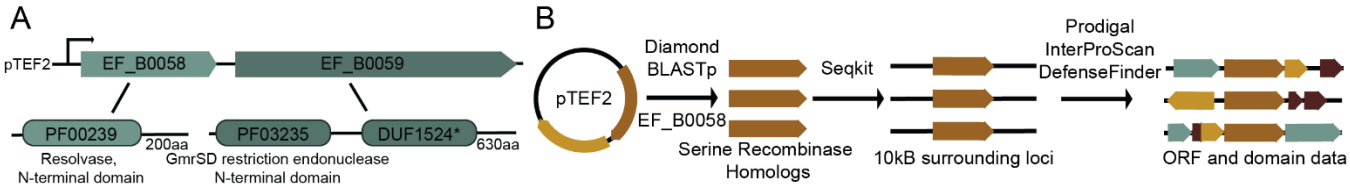

Figure 1. RADS leverages a serine recombinase associated with antiphage defense. (A) Graphic depicting the locus that inspired RADS. EF_B0058 is a serine recombinase that is co-transcribed with EF_B0059, a TIV restriction enzyme on the large conjugative plasmid pTEF2. (B) Graphic depicting the RADS bioinformatic pipeline.

183 surrounding open reading frames (ORFs) that are then assessed for antiphage defense genes (**Fig. 1B**).
184 We call this bioinformatic strategy Recombinase Associated Defense Search (RADS).

185 *E. faecalis* belongs to the Bacillota phylum. Therefore, to test whether RADS could identify
186 previously discovered antiphage defense genes, we ran the pipeline on all genomes belonging to the
187 Bacillota from NCBI (accessed August 2022). RADS identified 904 high homology (≥65% amino acid
188 identity) EF_B0058-like serine recombinases in 95,140 genomes, resulting in a discovery rate of 0.95%.
189 We next sought to provide evidence that the homologs of EF_B0058 identified by RADS were likely to
190 be proteins with demonstrated serine recombinase activity. Therefore, we took the nine lowest homology
191 matches to EF_B0058 and ran Alphafold3 modelling on these homologs and EF_B0058[30]. These
192 predicted structures were then aligned to the predicted structure of the *Escherichia* phage Mu serine
193 recombinase – a previously characterized serine recombinase (**Fig. 2A**)[34]. The predicted structures of
194 these proteins are nearly identical with an average RMSD value of 1.4Å between them, suggesting that
195 these are true serine recombinases, despite their diverse sequences.

196 To assess whether the contigs identified by RADS (hereafter referred to as RADS contigs) had
197 known antiphage defense systems we used DefenseFinder to query these DNA sequences[14]. Previously
198 characterized antiphage defense systems were grouped into one of eight categories: abortive infection,
199 CBASS, nucleic acid restriction, CRISPR-Cas, toxin/antitoxin, retron, RNA degradation, and unknown
200 mechanism. Known antiphage defense systems were found in high abundance within these contigs at a
201 discovery rate of 36.5% (**Fig.2B**). The most abundant antiphage defense category identified among the
202 Bacillota were restriction enzymes (**Fig. 2B)**, which were found to be distributed across many nucleic
203 acid restriction categories (**Fig. 2C**). EF_B0058 was originally discovered upstream of a TIV-RE system,
204 indicating these recombinases are often associated with nucleic acid restriction systems. The second

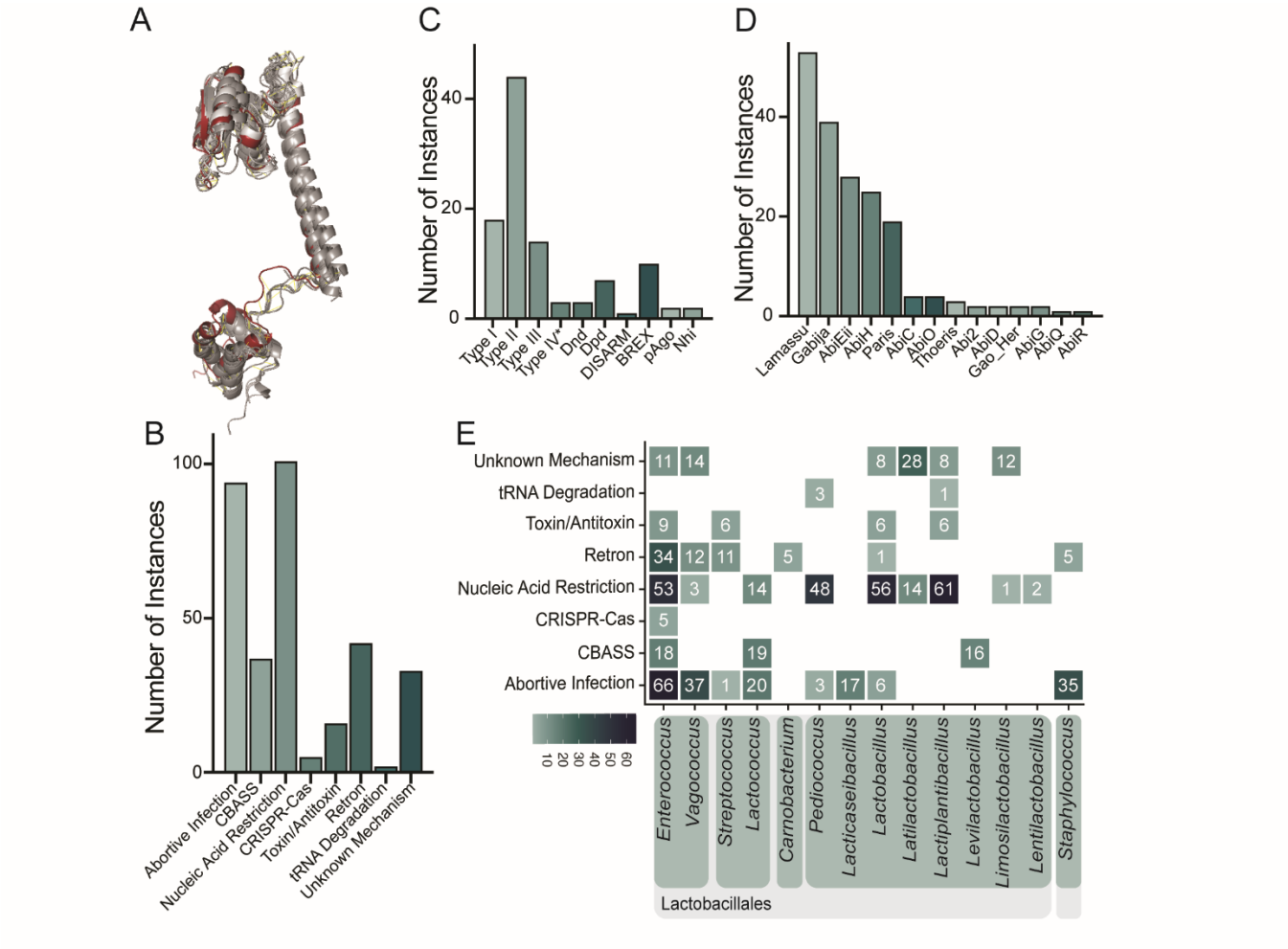

Figure 2. RADS identifies known antiphage defenses across genera in the Bacillota phylum. (A) Alphafold3 modeling of EF_B0058 and its 9 lowest-homology homologs from the Bacillota dataset (gray) aligned to the *Escherichia* phage Mu serine recombinase (red). RMSD=1.4. (B) Quantification of antiphage system categories, (C) nucleic acid restriction types, or (D) abortive infection types identified by DefenseFinder in Bacillota RADS contigs, respectively. (E) Heatmap of antiphage system categories identified across genera in the Bacillota phylum. Green boxes around genus names indicate taxonomic families, while gray boxes indicate orders.

most common category identified was abortive infection systems (**Fig. 2B**), with those identified representing a wide array of systems and mechanisms (**Fig. 2D**). These data show that characterized antiphage defense systems can be identified in RADS contigs.

Traditional antiphage defense system search methods has historically limited the diversity of bacteria interrogated. We sought to demonstrate that RADS could identify known antiphage defense systems in diverse bacteria. To first illustrate this, we assessed the distribution of known antiphage defense systems across genera in the Bacillota phylum. Antiphage defense systems were observed in contigs from fourteen bacterial genera, spanning five families and two orders (**Fig. 2E**). These data show

213 that RADS can identify antiphage systems in bacteria that have been previously neglected for antiphage
214 defense discovery.
215
216 **RADS identifies an abundance of antiphage defense systems from microbiomes**
217 To test the utility of RADS for finding antiphage defense systems in diverse phyla of bacteria from
218 polymicrobial communities, we ran the pipeline on a large publicly available human fecal microbiome
219 dataset[32]. We acquired 711 metagenome assemblies from pooled and un-pooled samples
220 (encompassing a total of 2,441 samples) from the Human Microbiome Project Data Analysis and
221 Coordination Center (DACC)[32] (accessed July 2024). Using these data, we asked whether serine
222 recombinases were distributed evenly across phyla. We first determined the bacterial taxonomy using
223 kraken2 with the phanta database[33], revealing Bacillota (67.14%), Bacteroidota (24.23%),
224 Pseudomonadota (3.21%), Uroviricota (2.26%), and Actinomycetota (1.75%) to be the most abundant
225 phyla (**Fig. 3A**). Contigs containing moderate homology to the EF_B0058 serine recombinase (≥30%
226 amino acid identity) were identified using RADS with a discovery rate of 0.016%. Serine recombinase-
227 containing contigs shared a similar taxonomic distribution compared to the input dataset, revealing
228 contigs from Bacillota (69.41%), Bacteroidota (21.34%), Pseudomonadota (3.26%), Uroviricota (3.03%),
229 and Actinomycetota (2.27%) to be most abundant (**Fig. 3A**). This shows that these serine recombinases
230 are well distributed across microbial phyla and that RADS can identify them in a manner that is not biased
231 toward a single phylum.
232 We next sought to determine if RADS could identify known antiphage defense systems in the
233 diverse phyla from these metagenomes. Abortive infection, CBASS, CRISPR-Cas, nucleic acid
234 restriction, retron, toxin/antitoxin, and systems of unknown mechanism were identified across nine phyla
235 at high abundance with an antiphage discovery rate of 30.2% in RADS contigs (**Fig. 3B**). Notably, the
236 distribution of phyla in which antiphage systems were identified is similar to the distribution of phyla in
237 the input and output datasets.
238 To demonstrate that RADS is broadly applicable to bacteria from different niches, we ran the
239 pipeline on a previously published soil microbiome dataset encompassing 3,304 metagenomes[31]. RADS

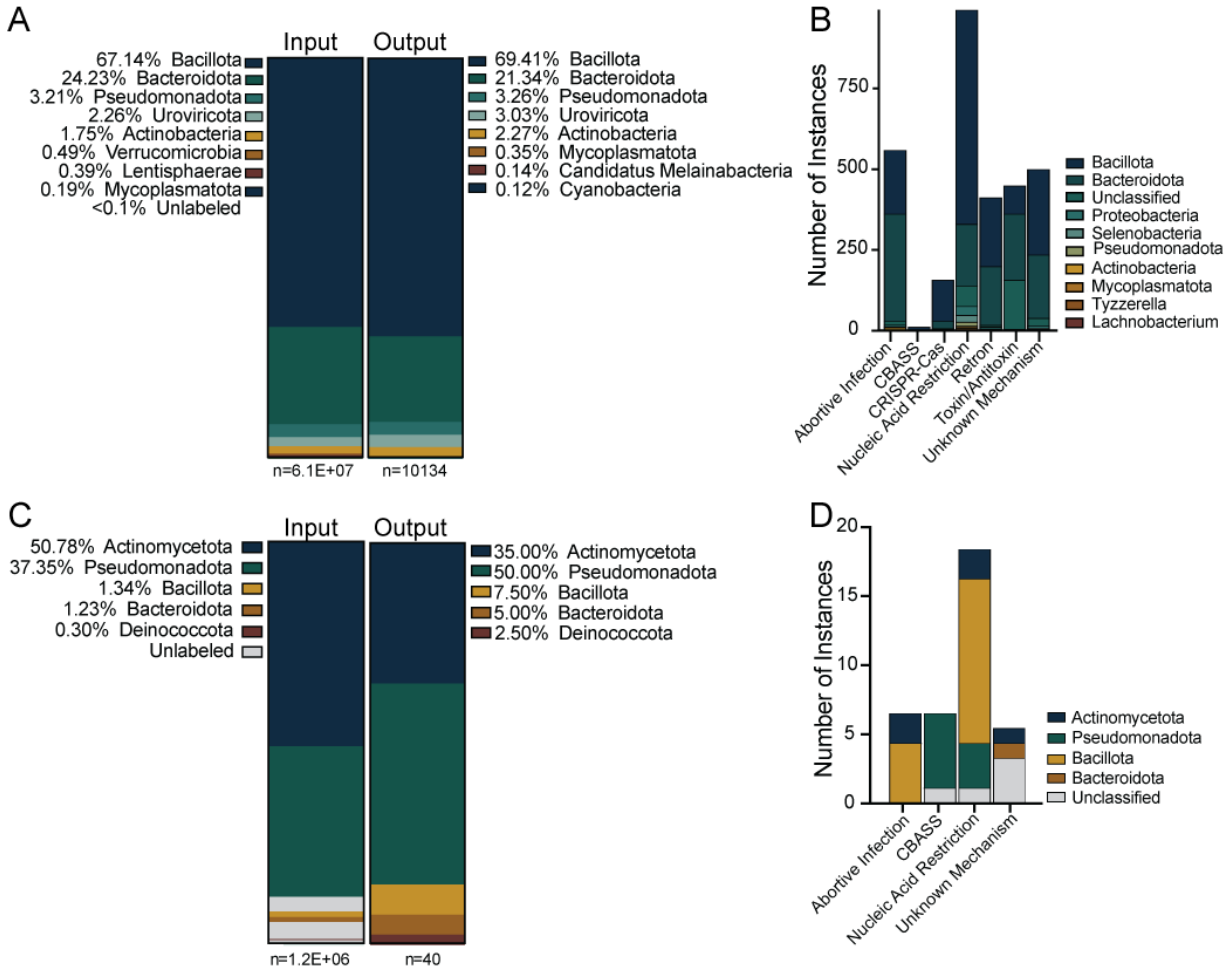

Figure 3. RADS identifies known antiphage defenses across diverse phyla. (A,B) Human fecal microbiome metagenomes or (C,D) soil microbiome metagenomes were run through the RADS pipeline, respectively. (A,C) Distribution of phyla identified by Kraken2 in the input dataset vs the contigs that came out of the RADS pipeline (output). (B,D) Quantification of antiphage system categories identified by DefenseFinder, where fill color represents the distribution of that category across different phyla.

contigs were identified at a discovery rate of 0.003% using moderate homology. While this discovery rate is approximately five-fold lower than observed for the gut microbiome dataset, this may be explained by contig quality of the input data. The average contig length for the soil metagenomes was 249.6bp, with a maximum length of 7,103bp. Meanwhile, the mean average and mean maximum length of the gut microbiome dataset was 1796bp and 186,996bp, respectively. Importantly, input and output phyletic distributions were similar in the soil metagenomes with Bacillota being a minority in the output dataset, further supporting RADS as an unbiased approach (**Fig. 3C**). We identified 34 antiphage defense systems in these RADS contigs, with a discovery rate of 85% (**Fig. 3D**).

To further substantiate that RADS can identify antiphage defenses in phyla divergent from the Bacillota, we ran RADS on genomes from the Bacteroidota. We pulled all Bacteroidota entries from NCBI (accessed July 2024), then randomly selected 33,520 entries, which encompassed metagenomic contigs and assembled genomes from a variety of genera within the phylum. RADS contigs were identified using moderate homology at a discovery rate of 9.8%. Not only were an overwhelming abundance of known antiphage defense systems discovered using RADS on the Bacteroidota (77.4% discovery rate) (**Fig. 4A**), but these systems were distributed across 13 families encompassing 4 orders (**Fig. 4B**). The most abundant defense system category of the Bacteroidota were toxin/antitoxin systems followed closely by nucleic acid defense systems similar to that observed for the Bacillota (**Fig. 4C, 4D**).

Altogether, these data validate RADS as a robust method for identifying putative antiphage defenses across highly diverse bacterial genomes in an unbiased manner.

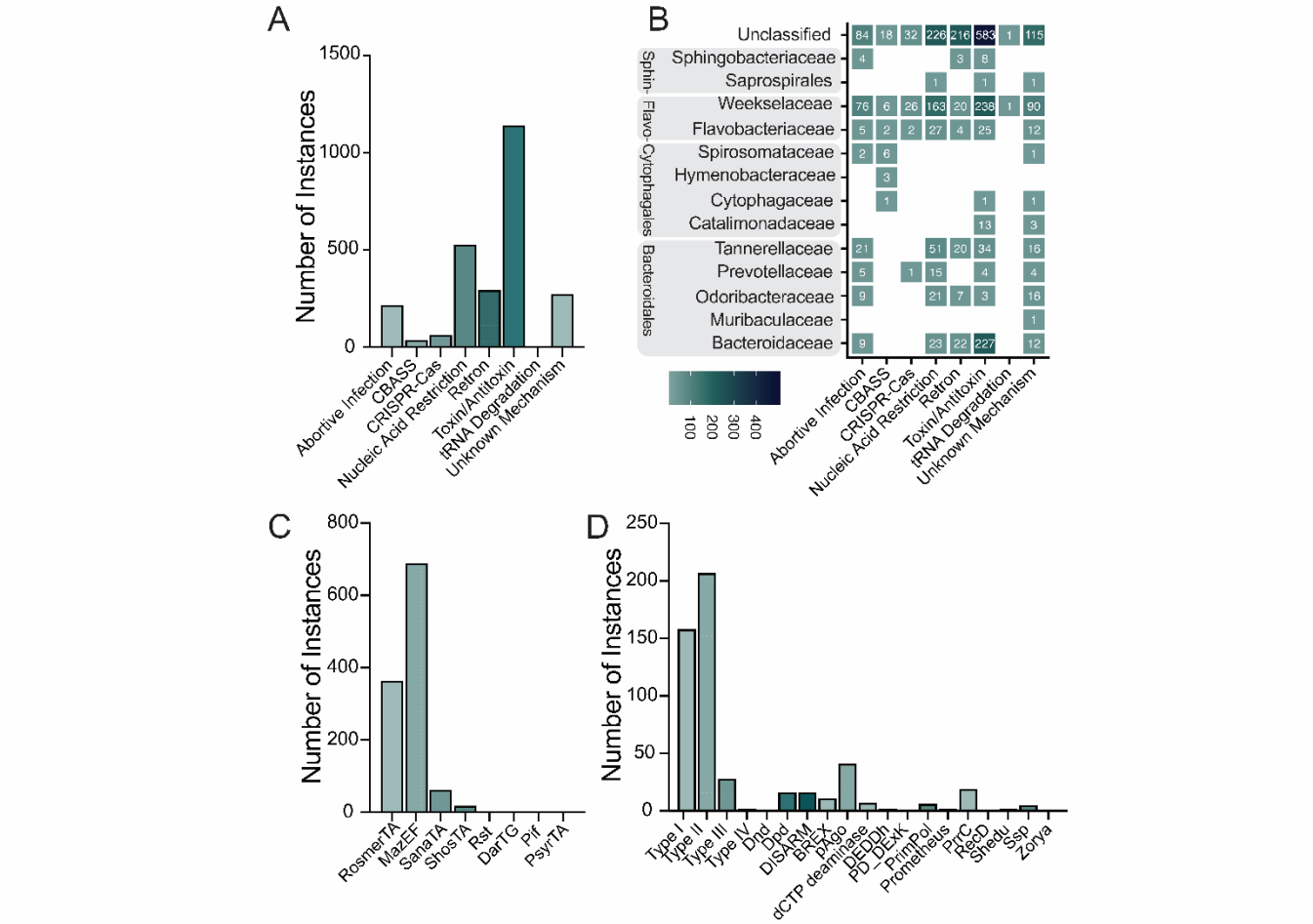

Figure 4. RADS identifies known antiphage defenses in the Bacteroidota. (A) Quantification of antiphage defense categories identified in the Bacteroidota. (B) Distribution of antiphge defense categories across families in the Bacteroidota. Gray boxes around family names indicate taxonomic orders. (C,D) Quantification of systems identified by Defense Finder that are in the (C) toxin/Antitoxin or (D) nucleic Acid Restriction categories.

With the success of using the EF_B0058 serine recombinase as bait for identifying antiphage defenses, we questioned whether this may be a general feature of recombinases. To test this, we utilized a tyrosine recombinase as bait. Similar to the serine recombinases, tyrosine recombinases are widely distributed in prokaryotic genomes and are involved in transposon mobility and prophage integration[35]. We ran RADS with either the *Bacteroides fragilis* tyrosine recombinase *tsr25*[36] or a *Clostridia* tyrosine-type recombinase/integrase (NCBI accession MFQ9799422.1). This recombinase was selected from *Clostridia* because it was the most closely related tyrosine recombinase to *tsr25* in the Bacillota phylum. Using these baits, RADS was then run on either the Bacillota phylum or the 33,520 randomly selected Bacteroidota genomes (**Fig. S1)**.

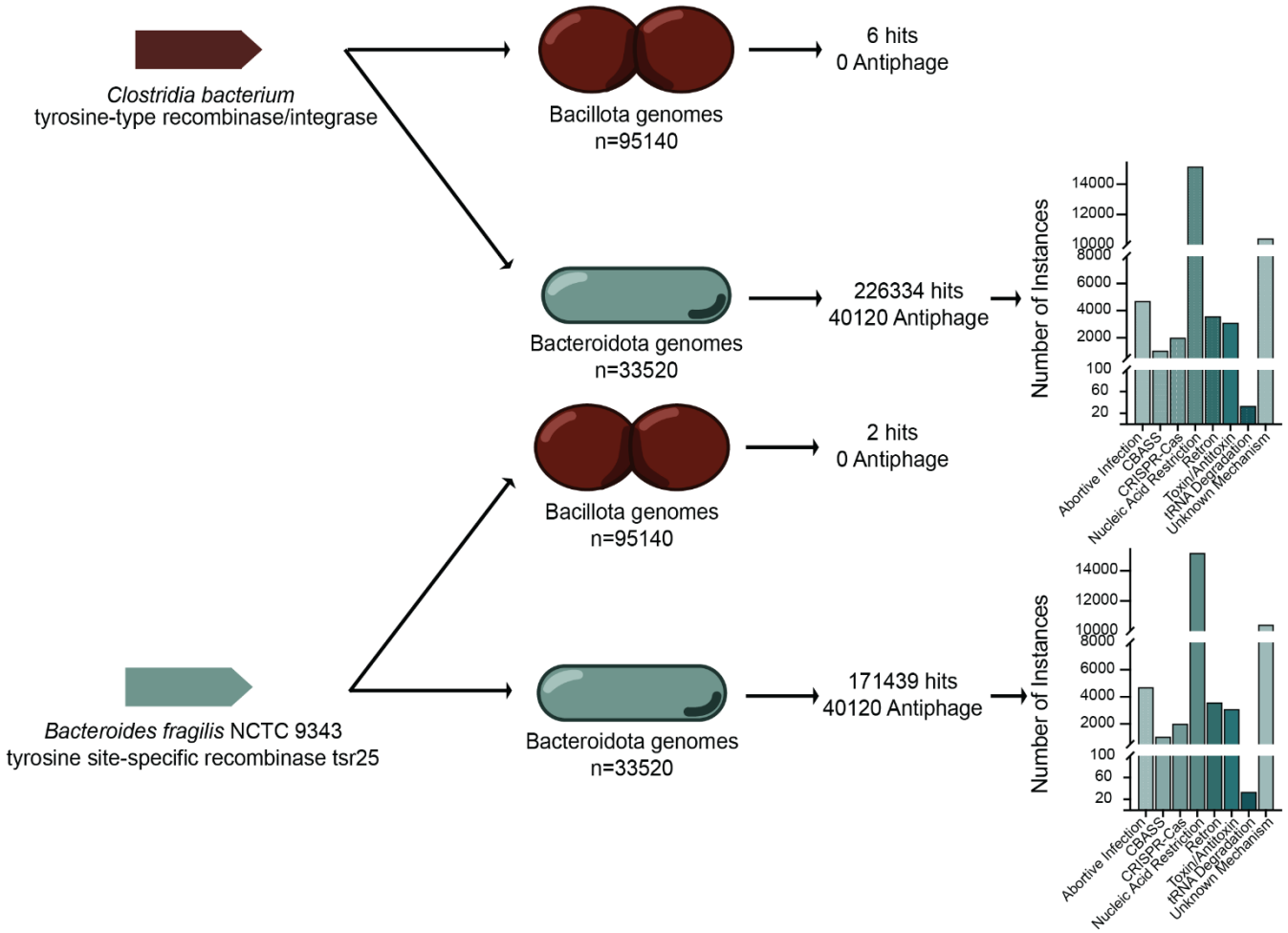

Figure S1. Tyrosine recombinases do not reliably identify antiphage defenses across phyla. Tyrosine recombinases were used as queries in the RADS pipeline on the Bacillota and Bacteroidota phyla. Graphic depicts searches performed with number of hits (recombinase homologs) and antiphage systems discovered noted. Graphs represent the antiphage defense system categories identified in tyrosine recombinase contigs.

268 Notably, these tyrosine recombinases exhibited phylum bias with very few homologs of both baits
269 found in the Bacillota and a multitude of homologs of both baits found in the Bacteroidota dataset. When
270 homologs were identified, an overwhelming abundance of antiphage defenses were noted in their
271 surrounding 10kb contig, particularly restriction enzymes (**Fig. S1**). Notably, the same antiphage
272 defenses were identified by both baits in the Bacteroidota phylum. This data highlights that while
273 recombinases more generally may be found surrounding antiphage defense systems, serine
274 recombinases exhibit unique phylum-level impartiality.
275
276 **Binomial and co-transcriptional analyses reveal novel antiphage defense genes using RADS**
277 Knowing that we could bioinformatically identify previously characterized antiphage defense systems
278 using RADS, we next sought to determine if RADS could be used to identify uncharacterized antiphage
279 defense systems. We focused our efforts on the dataset generated from the Bacillota phylum (**Fig. 2**). To
280 prioritize potential antiphage defense systems for further investigation, we employed a binomial statistical
281 analysis paired with co-transcriptional prediction between the serine recombinase gene and genes
282 containing one or more assigned pfam domain (identified by InterProScan, **Table S2**). The co-
283 transcriptional analysis determines if an ORF following a serine recombinase gene is on the same DNA

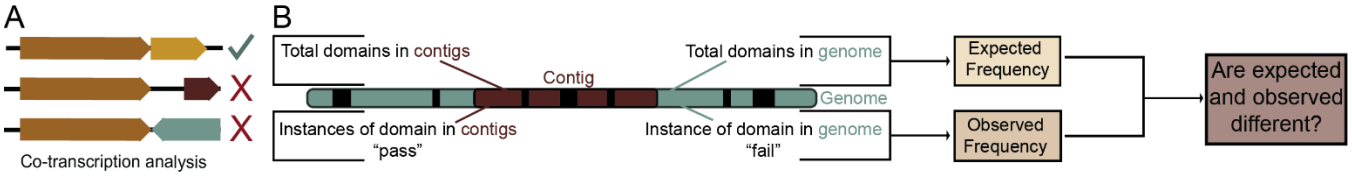

Figure S2. Additional analyses utilized to select ORFs to test for antiphage activity. (A) Graphic depicting the co-transcriptional analysis. ORFs were required to be on the same strand and within 100bp to determine they are likely to be co-transcribed. (B) Graphic depicting the binomial analysis. The frequency of finding a given domain in RADS contigs versus the genome was compared to the overall distribution of domains in contigs compared to domains in the genome. Comparing these frequencies allowed us to assign a statistical measure to how enriched domains were in contigs.

284 strand and if its predicted start site is within the broad distance of 100bp of the end of the serine
285 recombinase[37,38] **(Fig. S2A)**. If both conditions are met, the ORF is flagged as likely to be co-transcribed.
286 The binomial statistical analysis reports the enrichment of a given domain within a RADS contig compared
287 to the entire pangenome of the input dataset **(Fig. S2B)**.
288

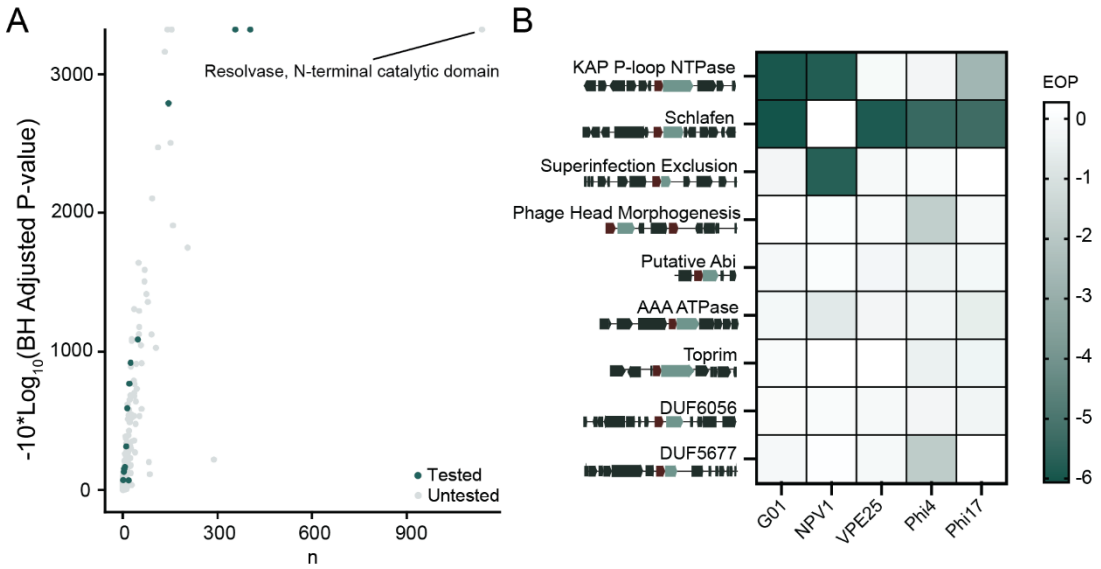

Figure 5. Discovery of previously uncharacterized antiphage defenses in RADS contigs. (A) Scatter plot of domains found in RADS contigs where n represents the number of instances in RADS contigs. The higher the value on the y axis, the more statistically enriched the domain. Green points represent domains chosen to test. (B) Heatmap of mean Log10(fold-change efficiency of plaquing (EOP)) of phages (X-axis) on *E. faecalis* OG1RF strains carrying ORF of interest (Y-axis) compared to strains carrying an empty vector. The more negative the EOP value, the better the antiphage defense. ORFs from each contig selected are depicted under the name of each system. Light green arrows indicates the ORF tested. Red arrows indicate recombinases.

289 Nine ORFs predicted to be co-transcribed with significantly enriched domains by binomial
290 statistical testing using Benjamini-Hochberg-corrected P-values (**Fig. 5A, Table S3**) were selected for
291 testing for antiphage defense activity (**Fig. 5B**). Tested ORFs include genes encoding domains of
292 unknown function (DUF) - DUF5677 and DUF6056, and genes predicted to encode putative Abi, AAA
293 ATPase, phage head morphogenesis, superinfection exclusion, Schlafen, and KAP (Kidins220/ARMS
294 and PifA) P-Loop NTPase domains (**Table S1, S3**). We prioritized the selection of ORFs from
295 *Enterococcus* or other lactic acid bacteria to maximize the likelihood of successfully expressing these
296 proteins in our model bacterium *E. faecalis* OG1RF, a strain that lacks prophages and other MGEs[39-41].
297 ORFs were selected from four *Enterococcus* species; *faecalis, faecium, hirae,* and *avium,* as well as
298 *Lactobacillus helveticus* and *Vagococcus fluvialis* (**Table S1**).

299 We cloned these candidate antiphage defense genes and constitutively expressed them in *E.*
300 *faecalis* OG1RF. We tested a panel of six genomically diverse *E. faecalis* phages for their ability to infect
301 these strains[24,25,42,43]. Plaque forming units (PFU) per mL were measured and compared between *E.*
302 *faecalis* cells expressing a potential antiphage defense system and cells carrying the empty vector to

303 generate a fold change Efficiency of Plaquing (EOP) (**Fig. 5B, S3A**). This experiment revealed that three
304 of the nine ORFs tested were potent antiphage defenses with EOP being decreased by up to six log-fold
305 change. These included the KAP P-Loop NTPase, Schlafen, and superinfection exclusion domain

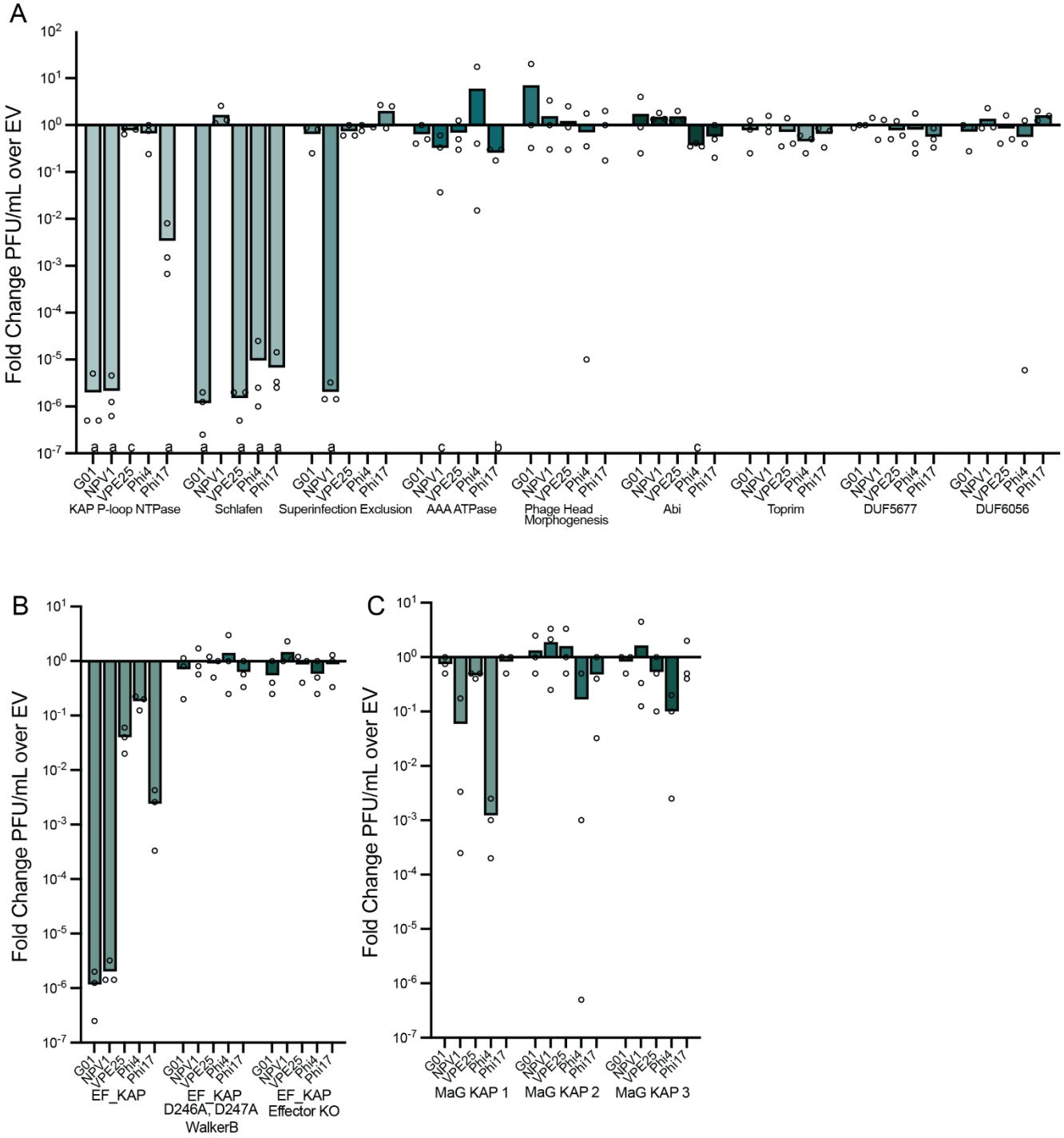

Figure S3. Raw values of EOP experiments. (A,B,C) Raw triplicate values (points) and means (bars) of EOP experiments in Figure 5B (A), Figure 6B (B), and Figure 6D (C). P-values determined by one-tailed one-sample t test compared against theoretical value of 1 (n=3, a < 0.0001, b < 0.01, c < 0.05).

306 containing proteins (**Fig. 5B, S3A).** An additional two showed weak to moderate antiphage defense; a
307 AAA ATPase and an undescribed Abi system domain containing protein (**Fig. 5B, S3A**).
308
309 **KAP P-loop NTPase domains are fused to antiphage effector domains**
310 As KAP P-loop NTPAases have not been robustly explored in association with antiphage defense, we
311 focused our efforts on this system. To further probe the function of the *E. faecalis* KAP NTPase ORF
312 (EF_KAP), we tested this system against various multiplicities of infection (MOIs) of phage. We first
313 determined that EF_KAP does not function through abortive infection as it is equally protective at low (1)
314 and high (10) MOIs **(Fig. S4).** AlphaFold3 modeling of EF_KAP suggests it consists of two N-terminal

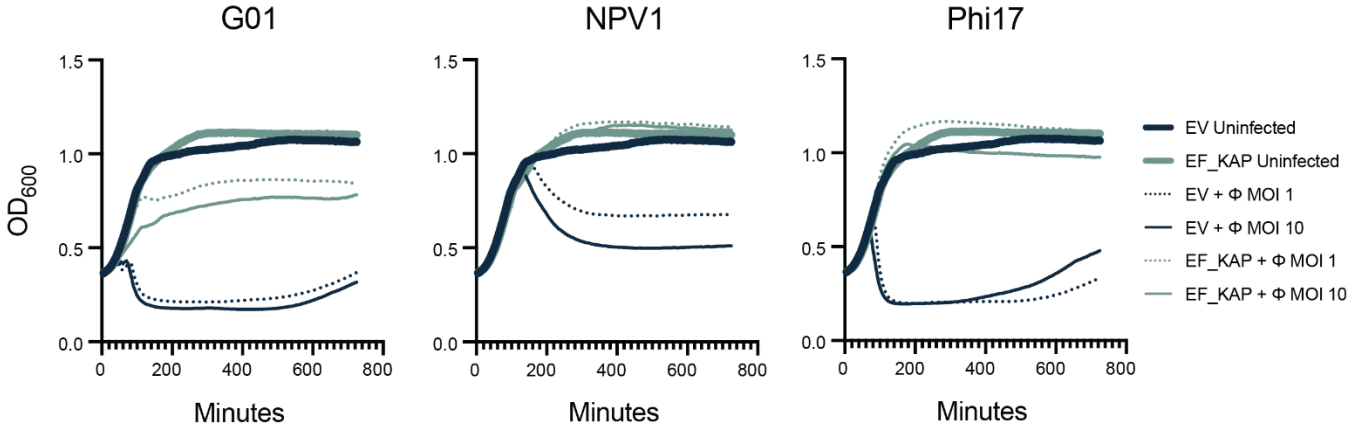

Figure S4. EF_KAP does not protect from phage via abortive infection. OD_{600} of cultures was measured over time in the presence or absence of phage infection (phages G01, NPV1, and Phi17) at MOI of 1 or 10. EV indicates host strain (*E. faecalis* OG1RF) carries empty vector, while EF_KAP indicates host strain carries plasmid expressing EF_KAP. Lines represent mean of two biological replicates of technical triplicates.

315 transmembrane helices, followed by the KAP NTPase domain **(Fig. 6A)**. KAP P-loop NTPases have been
316 previously reported to regulate diverse proteins across the tree of life[44]. Notably, in prokaryotes these
317 enzymes are hypothesized to regulate membrane-associated signaling complexes involved in restriction
318 of foreign DNA[44]. A KAP P-loop NTPase domain is found in PifA, a protein encoded by the enterobacterial
319 F plasmid that restricts phage T7, however it has been noted that phage restriction is likely not the primary
320 function of this protein nor is the mechanism of action known[45]. In the case of EF_KAP, the
321 transmembrane and KAP NTPase regions are followed by a 322 amino acid region with no known
322 domains or predicted similarity to proteins of known function. This suggests that this KAP-NTPase is
323 fused to an uncharacterized antiphage effector domain that may be activated by KAP-NTPase activity.
324

325 To begin to unravel how EF_KAP mediates protection against phage, we first generated two
326 amino acid substitutions (D246A, D247A) in the Walker B motif of the KAP NTPase domain, which has
327 been previously shown to be necessary for catalytic function by nucleotide hydrolysis[44]. Mutation of the
328 Walker B motif ablated antiphage defense (**Fig. 6B, S3B**). These single amino acid mutations of the
329 Walker B region suggest that this protein is not preventing phage adsorption, but rather its enzymatic
330 activity prevents phage replication.

331 We next sought to test whether KAP NTPase domains may regulate other fused putative effector
332 domains that have antiphage function and are divergent from EF_KAP. To explore this, we selected three
333 KAP P-Loop NTPase domain-containing ORFs from the human fecal microbiome metagenomic RADS
334 contigs with diverse putative effector regions (**Fig. 6C**). These ORFs were selected from diverse host

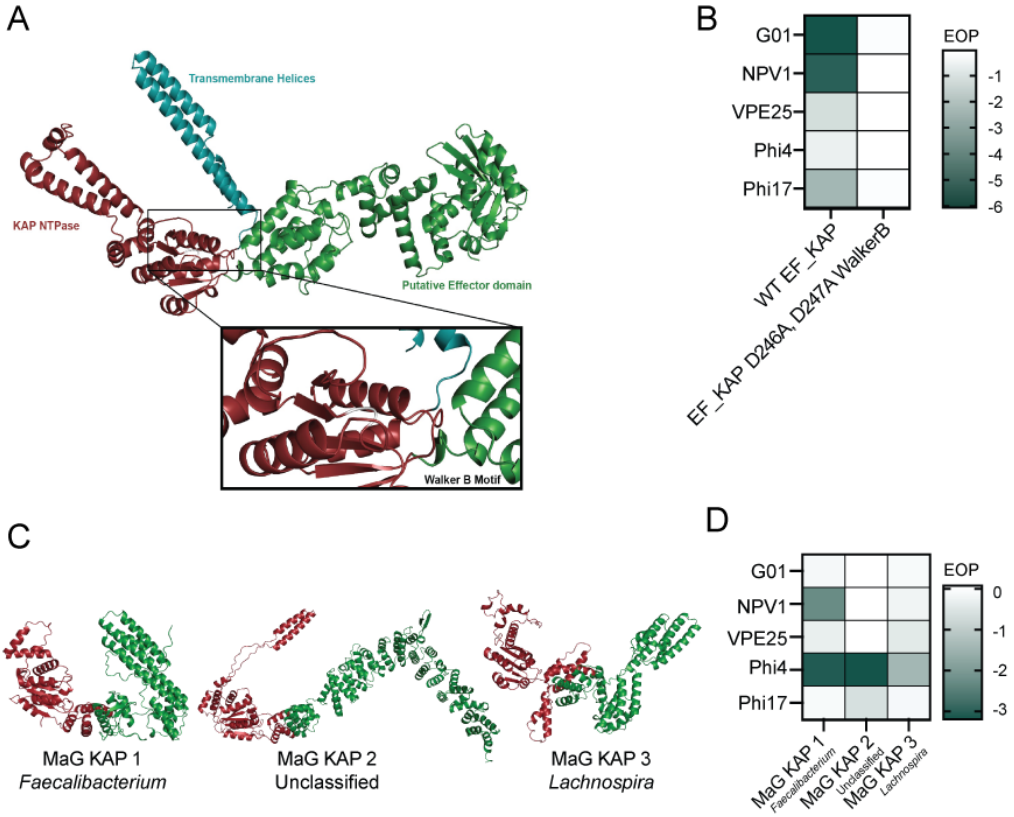

Figure 6. KAP P-Loop NTPases are associated with antiphage effector regions. (A) Alphafold3 modeling of EF_KAP where teal indicates the transmembrane helices, red indicates the KAP NTPase domain, white indicates the Walker B motif within the KAP NTPase region, and green indicates the effector region. (B) Heatmap of mean Log10(fold-change efficiency of plaquing (EOP)) of phages (Y) on E. faecalis OG1RF strains carrying EF_KAP or EF_KAP with WalkerB Mutation (X). The more negative the EOP value, the better the antiphage defense. (C) Alphafold3 modeling of KAP P-Loop NTPase domain (red) containing proteins with diverse effector regions (green). (D) Heatmap of mean Log10(fold-change efficiency of plaquing (EOP)) of phages (Y) on E. faecalis OG1RF strains carrying each diverse KAP P-Loop NTPase domain protein compared to strain carrying empty vector. The more negative the EOP value, the better the antiphage defense.

335 bacteria (*Faecalibacterium, Lachnospira*, and an unclassified contig) and predicted structural similarity to
336 characterized proteins. Importantly, we found that these three ORFs conferred protection from phage
337 infection in *E. faecalis* (**Fig. 6D, S3C**).

338 Together, these data indicate that KAP P-Loop NTPases that are fused to putative effector
339 domains belong to an undescribed class of antiphage defense proteins.

340

341 **Discussion**

342 In this study we developed a bioinformatic pipeline for identifying antiphage defense systems, which
343 we have named Recombinase Associated Defense Search (RADS). This pipeline leverages a serine
344 recombinase as genomic 'bait' to identify adjacently encoded putative antiphage defense systems. RADS
345 identifies antiphage defense systems in a manner unbiased by bacterial taxonomy. Additionally, we
346 leveraged the predictive power of RADS to identify previously uncharacterized antiphage defense
347 systems.

348 Prior work has utilized bioinformatic strategies to identify antiphage defense systems[5-7]. However,
349 these studies use known antiphage defense systems, prophages, or DNA genomic libraries to identify
350 genomic regions that may assemble into defense islands[5-9,46]. Although an effective strategy, this
351 approach is limited in the context of the genomic regions queried in a reference genome and is skewed
352 toward specific bacterial species. RADS represents a major step forward in the identification of antiphage
353 defenses as it begins to more evenly assess these systems across diverse bacterial phyla. RADS
354 includes a ranking scheme for ORFs likely to be involved in antiphage defense through their statistical
355 enrichment in contigs and their likelihood of being co-transcribed with a serine recombinase gene based
356 on genomic proximity. RADS streamlines the identification of candidate antiphage defense systems in a
357 high throughput manner.

358 Herein, we test nine ORFs identified using RADS for their ability to protect against phage infection,
359 discovering three highly potent and two moderately potent antiphage defense systems. However, this
360 only begins to scratch the surface of ORFs present in these RADS identified contigs. There are 425
361 domains that are enriched in Bacillota contigs that have potential to be antiphage defense systems,

362 regardless of co-transcription status. Multiple ORFs with diverse sequences can contain identical
363 domains, thus studying how a given domain functions across diverse ORFs could provide new insights
364 into domain function. Finally, ORFs identified with minimal antiphage defense activity may simply not be
365 active against the phages tested, or *E. faecalis* OG1RF is an unsuitable heterologous host.

366 RADS offers a path toward the continued discovery of new biology related to antiphage defense
367 systems. For instance, KAP P-loop NTPases were hypothesized to be antiphage defense systems by
368 Aravind *et al.* over 20 years ago[44]. More recently the protein Tmn, a YobI-like P-loop NTPase (as
369 determined by InterProScan), with transmembrane helices following the Walker A motif, synergizes with
370 the Gabija defense system[6,44,47]. Here we show proteins harboring the KAP P-loop NTPase domain are
371 potent antiphage defenses (**Fig. 5, 6**). We discovered a variety of KAP P-loop NTPases that offer cross-
372 genus protection against phages (**Fig. 6**). These KAP P-loop NTPases are associated with effector
373 regions that lack known domains or any structural similarity to characterized proteins. We also show that
374 the KAP P-Loop NTPases described in our study likely constitute a new class of proteins where the KAP
375 P-Loop NTPase domain is fused to an antiphage defense effector region.

376 Expanding on the discovery of previously unrecognized antiphage defense systems, we show
377 that a Schlafen domain-containing protein from *E. faecium* is a broad and potent antiphage defense
378 system (**Fig. 5**). Schlafens are mammalian antiviral proteins with extensive activity against eukaryotic
379 viruses[48] (**Fig. S5A**). However, a recent study also identified Schlafen proteins in prokaryotes that support
380 antiphage defense, independently corroborating our findings[49]. One such system mediates abortive
381 infection via host bacterial tRNA cleavage upon recognizing a phage tail assembly protein[49]. This
382 mechanism is likely conserved with eukaryotic Schlafen protein function where Schlafen domains are
383 fused to sensor domains that trigger Schlafen antiviral activity[49] (**Fig. S5A**). To date, characterized
384 prokaryotic Schlafen proteins are fused to sensor domains that impart antiphage function (**Fig. S5A**).
385 Intriguingly, the Schlafen protein identified in our study exhibits broad antiphage activity via abortive
386 infection despite its C-terminus lacking any known domains **(Fig. S5A,B).** This suggests that a variety of
387 yet to be described phage sensing mechanisms are mediated by Schlafen proteins, reminiscent of what
388 has been observed in eukaryotic organisms[39,48].

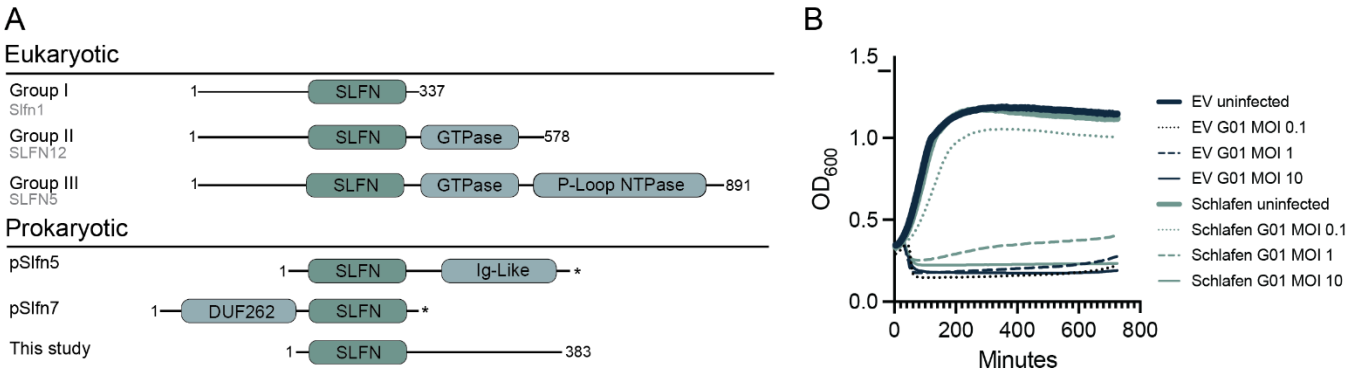

Figure S5. Schlafen proteins are conserved antivirals. (A) Comparison between prokaryotic and eukaryotic Schlafen proteins of different groups. Asterisk indicates size of protein is not currently reported. (B) OD_{600} of E. faecalis OG1RF cultures were measured over time in the presence or absence of phage G01 infection at MOI of 0.1, 1, or 10. EV indicates host strain carries empty vector, while Schlafen indicates host strain carries plasmid expressing Schlafen (This study). Lines represent mean of three biological replicates of technical triplicates

389

390 Other antiphage defense systems identified in this study had phage-specific and/or moderate to
391 weak effects on phage infection. This includes a predicted superinfection exclusion protein, a AAA
392 ATPase, and a putative Abi system of unknown function. Superinfection exclusion proteins support inter-
393 phage defense, often encoded in prophages, that restrict invading phages[50]. Interestingly, the
394 superinfection exclusion protein identified in our study is not found within a prophage as determined by
395 PHASTER[51]. This may represent the hijacking of a superinfection exclusion mechanism without
396 maintaining the prophage or potential recombination between a phage carrying a superinfection exclusion
397 mechanism and the host bacterial chromosome. Further study of this superinfection exclusion
398 mechanism may reveal unrecognized adaptive relationships between infecting phages and their host
399 bacteria that promotes antiphage defense.

400 Due to the breadth of antiphage defense genes found associated with serine recombinase genes,
401 it is likely that this co-occurrence represents more than just evolutionary synteny. Rather, we propose
402 that this link between serine recombinase genes and antiphage defense systems represents a broader
403 evolutionary adaptation. One potential hypothesis is that antiphage defense systems are mobilized by
404 horizontal gene transfer via prophages or plasmids[17-20]. However, maintenance of such genetic elements
405 can impose fitness constraints, such as the observation that plasmid maintenance can decrease bacterial
406 growth rate[52]. We postulate that these serine recombinases may help mobilize antiphage defenses to the
407 chromosome, thereby minimizing the need for carriage of costly mobile genetic elements and reducing

408 fitness costs. Additionally, serine recombinases similar to the recombinase used in RADS, function as
409 invertases that support DNA inversion events[53]. DNA inversion may be a means to regulate the
410 expression of antiphage defense genes during phage infection. Notably, a recent study showed that the
411 serine recombinase PinQ of the cryptic prophage Qin inverts DNA to form a protein coding sequence that
412 supports antiphage defense by interfering with phage adsorption[54]. Finally, these serine recombinases
413 may serve to diversify the host genome upon phage predation while working in tandem with the antiphage
414 defense system to block phage infection. Diverse populations are more resilient to infection[55], and this
415 could be a mechanism whereby genetic flexibility of phage defense systems through rapid recombination
416 and horizontal gene transfer supports population resilience.
417

418 **Acknowledgements**
419 This work was supported by NIH grants R01AI1414791 and R01AI171046 (BAD) and National Science
420 Foundation Graduate Research Fellowship AWD-230884 (SEA).
421

422 **References**

423 1 Kortright, K. E., Chan, B. K., Koff, J. L. & Turner, P. E. Phage therapy: A renewed approach to
424 combat antibiotic-resistant bacteria. *Cell Host Microbe* **25**, 219-232 (2019).
425 doi:10.1016/j.chom.2019.01.014

426 2 Strathdee, S. A., Hatfull, G. F., Mutalik, V. K. & Schooley, R. T. Phage therapy: From biological
427 mechanisms to future directions. *Cell* **186**, 17-31 (2023). doi:10.1016/j.cell.2022.11.017

428 3 Suh, G. A., Lodise, T. P., Tamma, P. D., Knisely, J. M., Alexander, J., Aslam, S., Barton, K. D.,
429 Bizzell, E., Totten, K. M. C., Campbell, J. L., Chan, B. K., Cunningham, S. A., Goodman, K. E.,
430 Greenwood-Quaintance, K. E., Harris, A. D., Hesse, S., Maresso, A., Nussenblatt, V., Pride, D.,
431 Rybak, M. J., Sund, Z., van Duin, D., Van Tyne, D. & Patel, R. Considerations for the use of
432 phage therapy in clinical practice. *Antimicrob Agents Chemother* **66**, e0207121 (2022).
433 doi:10.1128/aac.02071-21

434 4 Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community
435 resource. *Nat Rev Microbiol* **18**, 113-119 (2020). doi:10.1038/s41579-019-0278-2

436 5 Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G. & Sorek, R.
437 Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**
438 (2018). doi:10.1126/science.aar4120

439 6 Gao, L., Altae-Tran, H., Böhning, F., Makarova, K. S., Segel, M., Schmid-Burgk, J. L., Koob, J.,
440 Wolf, Y. I., Koonin, E. V. & Zhang, F. Diverse enzymatic activities mediate antiviral immunity in
441 prokaryotes. *Science* **369**, 1077-1084 (2020). doi:10.1126/science.aba0372

442 7 Millman, A., Melamed, S., Leavitt, A., Doron, S., Bernheim, A., Hör, J., Garb, J., Bechon, N.,
443 Brandis, A., Lopatina, A., Ofir, G., Hochhauser, D., Stokar-Avihail, A., Tal, N., Sharir, S., Voichek,
444 M., Erez, Z., Ferrer, J. L. M., Dar, D., Kacen, A., Amitai, G. & Sorek, R. An expanded arsenal of
445 immune systems that protect bacteria from phages. *Cell Host Microbe* **30**, 1556-1569.e1555
446 (2022). doi:10.1016/j.chom.2022.09.017

447 8 Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I. C. & Laub, M. T. A functional
448 selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome.
449 *Nat Microbiol* **7**, 1568-1579 (2022). doi:10.1038/s41564-022-01219-4

450 9 Rodriguez-Rodriguez, L., Pfister, J., Schuck, L., Martin, A. E., Mercado-Santiago, L. M.,
451 Tagliabracci, V. S. & Forsberg, K. J. Metagenomic selections reveal diverse antiphage defenses
452 in human and environmental microbiomes. *Cell Host Microbe* **33**, 1381-1395 e1387 (2025).
453 doi:10.1016/j.chom.2025.07.005

454 10 Rousset, F., Depardieu, F., Miele, S., Dowding, J., Laval, A.-L., Lieberman, E., Garry, D., Rocha,
455 E. P. C., Bernheim, A. & Bikard, D. Phages and their satellites encode hotspots of antiviral
456 systems. *Cell Host & Microbe* **30**, 740-753.e745 (2022). doi:10.1016/j.chom.2022.02.018

457 11 Ofir, G., Herbst, E., Baroz, M., Cohen, D., Millman, A., Doron, S., Tal, N., Malheiro, D. B. A.,
458 Malitsky, S., Amitai, G. & Sorek, R. Antiviral activity of bacterial TIR domains via immune
459 signalling molecules. *Nature* **600**, 116-120 (2021). doi:10.1038/s41586-021-04098-7

460 12 Tal, N., Millman, A., Stokar-Avihail, A., Fedorenko, T., Leavitt, A., Melamed, S., Yirmiya, E.,
461 Avraham, C., Brandis, A., Mehlman, T., Amitai, G. & Sorek, R. Bacteria deplete
462 deoxynucleotides to defend against bacteriophage infection. *Nat Microbiol* **7**, 1200-1209 (2022).
463 doi:10.1038/s41564-022-01158-0

464 13 Whiteley, A. T., Eaglesham, J. B., de Oliveira Mann, C. C., Morehouse, B. R., Lowey, B.,
465 Nieminen, E. A., Danilchanka, O., King, D. S., Lee, A. S. Y., Mekalanos, J. J. & Kranzusch, P. J.
466 Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194-199
467 (2019). doi:10.1038/s41586-019-0953-5

468 14 Tesson, F., Hervé, A., Mordret, E., Touchon, M., d'Humières, C., Cury, J. & Bernheim, A.
469 Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun* **13**, 2561
470 (2022). doi:10.1038/s41467-022-30269-9

471 15 Yamaguchi, T., Ono, D. & Sato, A. Staphylococcal cassette chromosome *mec* (SCC*mec*)
472 analysis of MRSA. *Methods Mol Biol* **2069**, 59-78 (2020). doi:10.1007/978-1-4939-9849-4_4

473 16 Ross, B. D., Verster, A. J., Radey, M. C., Schmidtke, D. T., Pope, C. E., Hoffman, L. R., Hajjar,
474 A. M., Peterson, S. B., Borenstein, E. & Mougous, J. D. Human gut bacteria contain acquired
475 interbacterial defence systems. *Nature* **575**, 224-228 (2019). doi:10.1038/s41586-019-1708-z

476 17 Beamud, B., Benz, F. & Bikard, D. Going viral: The role of mobile genetic elements in bacterial
477 immunity. *Cell Host & Microbe* **32**, 804-819 (2024). doi:10.1016/j.chom.2024.05.017

478 18 Beavogui, A., Lacroix, A., Wiart, N., Poulain, J., Delmont, T. O., Paoli, L., Wincker, P. & Oliveira,
479 P. H. The defensome of complex bacterial communities. *Nat Commun* **15**, 2146 (2024).
480 doi:10.1038/s41467-024-46489-0

481 19 Hochhauser, D., Millman, A. & Sorek, R. The defense island repertoire of the *Escherichia coli*
482 pan-genome. *PLoS Genet* **19**, e1010694 (2023). doi:10.1371/journal.pgen.1010694

483 20 Fillol-Salom, A., Rostøl, J. T., Ojiogu, A. D., Chen, J., Douce, G., Humphrey, S. & Penadés, J. R.
484 Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against
485 competitors. *Cell* **185**, 3248-3262.e3220 (2022). doi:10.1016/j.cell.2022.07.014

486 21 Hossain, M., Aslan, B. & Hatoum-Aslan, A. Tandem mobilization of anti-phage defenses
487 alongside SCC*mec* elements in staphylococci. *Nat Commun* **15**, 8820 (2024).
488 doi:10.1038/s41467-024-53146-z

489 22 Botelho, J. Defense systems are pervasive across chromosomally integrated mobile genetic
490 elements and are inversely correlated to virulence and antimicrobial resistance. *Nucleic Acids*
491 *Res* **51**, 4385-4397 (2023). doi:10.1093/nar/gkad282

492 23 Bullen, N. P., Johnson, C. N., Andersen, S. E., Arya, G., Marotta, S. R., Lee, Y. J., Weigele, P.
493 R., Whitney, J. C. & Duerkop, B. A. An enterococcal phage protein inhibits type IV restriction
494 enzymes involved in antiphage defense. *Nat Commun* **15**, 6955 (2024). doi:10.1038/s41467-
495 024-51346-1

496 24 Chatterjee, A., Johnson, C. N., Luong, P., Hullahalli, K., McBride, S. W., Schubert, A. M.,
497 Palmer, K. L., Carlson, P. E., Jr. & Duerkop, B. A. Bacteriophage resistance alters antibiotic-
498 mediated intestinal expansion of enterococci. *Infect Immun* **87** (2019). doi:10.1128/iai.00085-19

499 25 Duerkop, B. A., Huo, W., Bhardwaj, P., Palmer, K. L. & Hooper, L. V. Molecular basis for lytic
500 bacteriophage resistance in enterococci. *mBio* **7** (2016). doi:10.1128/mBio.01304-16

501 26 Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W. & Hauser, L. J. Prodigal:
502 prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**,
503 119 (2010). doi:10.1186/1471-2105-11-119

504 27 Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using
505 DIAMOND. *Nat Methods* **18**, 366-368 (2021). doi:10.1038/s41592-021-01101-x

506 28 Shen, W., Sipos, B. & Zhao, L. SeqKit2: A Swiss army knife for sequence and alignment
507 processing. *Imeta* **3**, e191 (2024). doi:10.1002/imt2.191

508 29 Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J.,
509 Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong,
510 S.-Y., Lopez, R. & Hunter, S. InterProScan 5: genome-scale protein function classification.
511 *Bioinformatics* **30**, 1236-1240 (2014). doi:10.1093/bioinformatics/btu031

512 30 Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore,
513 L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M.,
514 Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O.,
515 Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M.,
516 Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy,
517 P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M.,
518 Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D. & Jumper, J. M. Accurate structure
519 prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500 (2024).
520 doi:10.1038/s41586-024-07487-w

521 31 Ma, B., Lu, C., Wang, Y., Yu, J., Zhao, K., Xue, R., Ren, H., Lv, X., Pan, R., Zhang, J., Zhu, Y. &
522 Xu, J. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic
523 resources. *Nat Commun* **14**, 7318 (2023). doi:10.1038/s41467-023-43000-z

524 32 Gevers, D., Knight, R., Petrosino, J. F., Huang, K., McGuire, A. L., Birren, B. W., Nelson, K. E.,
525 White, O., Methé, B. A. & Huttenhower, C. The Human Microbiome Project: a community
526 resource for the healthy human microbiome. *PLoS Biol* **10**, e1001377 (2012).
527 doi:10.1371/journal.pbio.1001377

528 33 Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome*
529 *Biology* **20**, 257 (2019). doi:10.1186/s13059-019-1891-0

530 34 Plasterk, R. H., Brinkman, A. & van de Putte, P. DNA inversions in the chromosome of
531 *Escherichia coli* and in bacteriophage Mu: relationship to other site-specific recombination
532 systems. *Proc Natl Acad Sci U S A* **80**, 5355-5358 (1983). doi:10.1073/pnas.80.17.5355

533 35 Jayaram, M., Ma, C. H., Kachroo, A. H., Rowley, P. A., Guga, P., Fan, H. F. & Voziyanov, Y. An
534 overview of tyrosine site-specific recombination: From an Flp perspective. *Microbiol Spectr* **3**
535 (2015). doi:10.1128/microbiolspec.MDNA3-0021-2014

536 36 Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N.,
537 Poxton, I., Duerden, B., Harris, B., Quail, M. A., Barron, A., Clark, L., Corton, C., Doggett, J.,
538 Holden, M. T., Larke, N., Line, A., Lord, A., Norbertczak, H., Ormond, D., Price, C.,
539 Rabbinowitsch, E., Woodward, J., Barrell, B. & Parkhill, J. Extensive DNA inversions in the *B.*
540 *fragilis* genome control variable gene expression. *Science* **307**, 1463-1465 (2005).
541 doi:10.1126/science.1107008

542 37 Wang, L., Trawick, J. D., Yamamoto, R. & Zamudio, C. Genome-wide operon prediction in
543 *Staphylococcus aureus*. *Nucleic Acids Res* **32**, 3689-3702 (2004). doi:10.1093/nar/gkh694

544 38 Charaniya, S., Mehra, S., Lian, W., Jayapal, K. P., Karypis, G. & Hu, W. S. Transcriptome
545 dynamics-based operon prediction and verification in *Streptomyces coelicolor*. *Nucleic Acids*
546 *Res* **35**, 7222-7236 (2007). doi:10.1093/nar/gkm501

547 39 Chen, J. & Kuhn, L. A. Deciphering the three-domain architecture in schlafens and the
548 structures and roles of human *schlafen12* and *serpinB12* in transcriptional regulation. *J Mol*
549 *Graph Model* **90**, 59-76 (2019). doi:10.1016/j.jmgm.2019.04.003

550 40 Palmer, K. L., Godfrey, P., Griggs, A., Kos, V. N., Zucker, J., Desjardins, C., Cerqueira, G.,
551 Gevers, D., Walker, S., Wortman, J., Feldgarden, M., Haas, B., Birren, B. & Gilmore, M. S.

552 Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E.*
553 *faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. *mBio* **3**, e00318-
554 00311 (2012). doi:10.1128/mBio.00318-11

555 41 Bourgogne, A., Garsin, D. A., Qin, X., Singh, K. V., Sillanpaa, J., Yerrapragada, S., Ding, Y.,
556 Dugan-Rocha, S., Buhay, C., Shen, H., Chen, G., Williams, G., Muzny, D., Maadani, A., Fox, K.
557 A., Gioia, J., Chen, L., Shang, Y., Arias, C. A., Nallapareddy, S. R., Zhao, M., Prakash, V. P.,
558 Chowdhury, S., Jiang, H., Gibbs, R. A., Murray, B. E., Highlander, S. K. & Weinstock, G. M.
559 Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain
560 OG1RF. *Genome Biol* **9**, R110 (2008). doi:10.1186/gb-2008-9-7-r110

561 42 Sheriff, E. K., Andersen, S. E., Chatterjee, A. & Duerkop, B. A. Complete genome sequence of
562 enterococcal phage G01. *Microbiol Resour Announc* **13**, e0121723 (2024).
563 doi:10.1128/mra.01217-23

564 43 Trotter, K. M. & Dunny, G. M. Mutants of *Enterococcus faecalis* deficient as recipients in mating
565 with donors carrying pheromone-inducible plasmids. *Plasmid* **24**, 57-67 (1990).
566 doi:10.1016/0147-619X(90)90025-8

567 44 Aravind, L., Iyer, L. M., Leipe, D. D. & Koonin, E. V. A novel family of P-loop NTPases with an
568 unusual phyletic distribution and transmembrane segments inserted within the NTPase domain.
569 *Genome Biol* **5**, R30 (2004). doi:10.1186/gb-2004-5-5-r30

570 45 Schmitt, C. K., Kemp, P. & Molineux, I. J. Genes 1.2 and 10 of bacteriophages T3 and T7
571 determine the permeability lesions observed in infected cells of *Escherichia coli* expressing the
572 F plasmid gene *pifA*. *J Bacteriol* **173**, 6507-6514 (1991). doi:10.1128/jb.173.20.6507-6514.1991

573 46 Egorov, A. A., Hauryliuk, V. & Atkinson, G. C. Systematic annotation of hyper-variability hotspots
574 in phage genomes and plasmids. *bioRxiv*, 2024.2010.2015.618418 (2024).
575 doi:10.1101/2024.10.15.618418

576 47 Wu, Y., Garushyants, S. K., van den Hurk, A., Aparicio-Maldonado, C., Kushwaha, S. K., King,
577 C. M., Ou, Y., Todeschini, T. C., Clokie, M. R. J., Millard, A. D., Gencay, Y. E., Koonin, E. V. &
578 Nobrega, F. L. Bacterial defense systems exhibit synergistic anti-phage activity. *Cell Host*
579 *Microbe* **32**, 557-572 e556 (2024). doi:10.1016/j.chom.2024.01.015

580 48 Jo, U. & Pommier, Y. Structural, molecular, and functional insights into Schlafen proteins. *Exp*
581 *Mol Med* **54**, 730-738 (2022). doi:10.1038/s12276-022-00794-0

582 49 Taboada, V. P., Wu, Y., Cassidy, R., Medvedev, K. E., Loeff, L., Nemudraia, A. & Nemudryi, A.
583 Bacterial Schlafens mediate anti-phage defense. *bioRxiv*, 2025.2007.2024.666596 (2025).
584 doi:10.1101/2025.07.24.666596

585 50 Bucher, M. J. & Czyż, D. M. Phage against the machine: The SIE-ence of superinfection
586 exclusion. *Viruses* **16** (2024). doi:10.3390/v16091348

587 51 Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. & Wishart, D. S. PHASTER: a
588 better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**, W16-21 (2016).
589 doi:10.1093/nar/gkw387

590 52 Baltrus, D. A. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol* **28**, 489-495
591 (2013). doi:10.1016/j.tree.2013.04.002

592 53 Stark, W. M. The serine recombinases. *Microbiol Spectr* **2** (2014).
593 doi:10.1128/microbiolspec.MDNA3-0046-2014

594 54 Kirigo, J., Huelgas-Mendez, D., Tomas, M., Benedik, M. J., Garcia-Contreras, R. & Wood, T. K.
595 Adsorption of phage T2 is inhibited due to inversion of cryptic prophage DNA by the serine
596 recombinase PinQ. *Nucleic Acids Res* **53** (2025). doi:10.1093/nar/gkaf1041

597 55 Gibson, A. K. Genetic diversity and disease: The past, present, and future of an old idea.
598 *Evolution* **76**, 20-36 (2022). doi:10.1111/evo.14395

599