



## Lambdoid phages with abundant Chi recombination hotspots reflect diverse viral strategies for recombination-dependent growth

Clarence Zheng, Sherwood R. Casjens, Alan R. Davidson, et al.

*Genome Res.* 2025 35: 1767-1780 originally published online July 10, 2025

Access the most recent version at doi:[10.1101/gr.280248.124](https://doi.org/10.1101/gr.280248.124)

---

**References** This article cites 103 articles, 30 of which can be accessed free at:  
<http://genome.cshlp.org/content/35/8/1767.full.html#ref-list-1>

**Open Access** Freely available online through the *Genome Research* Open Access option.

**Creative Commons License** This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

## Research

# Lambdoid phages with abundant Chi recombination hotspots reflect diverse viral strategies for recombination-dependent growth

Clarence Zheng,<sup>1,4</sup> Sherwood R. Casjens,<sup>2</sup> Alan R. Davidson,<sup>3</sup> Susan K. Amundsen,<sup>1</sup> and Gerald R. Smith<sup>1</sup>

<sup>1</sup>Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA; <sup>2</sup>Department of Pathology, University of Utah, Salt Lake City, Utah 84112, USA; <sup>3</sup>Department of Molecular Genetics and Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1M1, Canada

Many phages encode recombination-mediating enzymes, but characterization of their roles in phage lifecycles is limited, and their impact on phage replication is controversial. To address these issues, we have searched for phages whose growth is impacted by the major recombination-promoting helicase-nuclease of *Escherichia coli*, the RecBCD enzyme. Although no phages inhibited by RecBCD are identified, growth of a newly isolated phage, named LLS, is enhanced by RecBCD. LLS's genome sequence reveals it is related to bacteriophage  $\lambda$  but encodes no recombination-promoting (Rec) proteins or associated RecBCD inhibitor. However, it contains an unexpectedly high number of Chi sites, activators of RecBCD-dependent recombination. Through analysis of 325 genomes of phages related to  $\lambda$  (lambdoid phages), we have found 71 other phage genomes that encode no Rec proteins but mostly possess large numbers of Chi sites. Conversely, phages encoding Rec proteins and a RecBCD inhibitor (collectively a Rec module) mostly lack Chi sites. Lambdoid phages of both diverse enteric bacteria and a pseudomonad have these properties. For this study, we thoroughly analyze the Rec modules of 246 lambdoid phage genomes. These analyses reveal a remarkable heterogeneity of Rec module protein types, both in sequence and in function, and allow us to identify phages that do not contain Rec modules. We conclude that phages lacking their own recombination systems have compensated by becoming enriched in Chi sites, enabling them to use the host's RecBCD to fulfill the requirement for recombination to efficiently replicate. This study highlights the importance of recombination for phage survival and the diversity of strategies to achieve it.

[Supplemental material is available for this article.]

Viruses encode many functions essential for their growth, but host functions are also critical for nearly all viruses. Examples of the latter are functions for energy production (ATP synthesis) and synthesis of precursors for macromolecules (amino acids for proteins and nucleoside triphosphates for DNA and RNA synthesis). In some cases, as studied here, recombination is critical for viral DNA replication, which together produce concatemeric DNA essential for viral packaging. This is the case for viruses as diverse as bacteriophage, such as  $\lambda$ , and human viruses, such as herpes simplex virus (HSV) and pox virus (Weller and Sawitzke 2014; Weerasooriya et al. 2019; Packard and Dembowski 2021; Evans 2022). The recombination functions can be encoded by the virus or be supplied by the host. In some cases, these two sources can be used interchangeably, but in other cases, only the host functions are available because the virus does not encode recombination functions.

An example of a virus requiring host-encoded recombination is *Escherichia coli* temperate phage P1, which requires the host's recombination proteins RecA and RecBCD for growth

(Sternberg and Hoess 1983). These proteins promote recombination of P1's terminal direct repeats of ~5–10 kbp when P1's 94.8 kbp linear DNA is injected into the cell. Terminal-repeat recombination forms circular DNA, which is replicated to form the concatemeric DNA that is essential for packaging and formation of progeny phage. The host RecA promotes exchange of homologous DNA strands to form D-loops and Holliday junctions, which lead to DNA break repair and recombination (Bell and Kowalczykowski 2016). The three-subunit RecBCD enzyme of *E. coli* and other enteric bacteria is a complex helicase-nuclease with multiple enzymatic activities (for review, see Amundsen and Smith 2023). During its rapid unwinding of DNA (up to 1 kb/sec), RecBCD produces single-stranded loops and tails. Upon encountering a properly oriented crossover hotspot instigator (Chi) site 5'-GCTGGTGG-3', RecBCD nicks the strand with that sequence and loads RecA onto the newly generated 3'-end. Polymerization of RecA then forms a single-stranded DNA-protein filament, which can engage homologous duplex DNA to form a D-loop. Replication can be initiated at the invading 3'-end in the D-loop, leading to abundant DNA for formation of new progeny. These essential recombination events are likely stimulated by the 50 Chi sites in P1 DNA, most of which are in

<sup>4</sup>Present address: Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA  
Corresponding author: [gsmith@fredhutch.org](mailto:gsmith@fredhutch.org)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280248.124>. Freely available online through the *Genome Research* Open Access option.

© 2025 Zheng et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

or near its terminal repeats; three P1-related phages have similar properties (Subramaniam and Smith 2022).

Phages related to the well-studied bacteriophage  $\lambda$  (lambdoid phages) also require recombination for their growth (for review, see Smith 1983; for discussion of phage classification nomenclature, see Valencia-Toxqui and Ramsey 2024). The classic bacteriophage  $\lambda$  encodes its own nuclease (Exo) and strand-exchange protein (Beta); the corresponding genes are often designated *red $\alpha$*  (or *exo*) and *red $\beta$*  (or *bet*). Consequently,  $\lambda$  grows on *recA* and *recBCD* mutant hosts.  $\lambda$  also encodes an inhibitor of RecBCD called Gam.  $\lambda$  *red gam* double mutants fail to grow on *recA* mutants because of the lack of both host- and phage-promoted recombination. They do grow on *recBCD* nuclease-negative cells because rolling circle replication, which occurs in the absence RecBCD nuclease activity (with or without recombination), produces concatemeric DNA for packaging.

The RecBCD of *E. coli* and other enteric bacteria is activated by Chi sequences to enhance repair and recombination of broken (linear) DNA (for review, see Smith 2012). *E. coli* strain K-12 (obtained from the NCBI GenBank database [<https://www.ncbi.nlm.nih.gov/genbank/>] under accession number NC\_000913.3) has 1008 Chi sites in its 4.64 Mbp genome, or one per 4.6 kbp on average (Blattner et al. 1997). In contrast, some *E. coli* phages, including the 48.5 kbp bacteriophage  $\lambda$  (GenBank accession number J02459), have no Chi sites (Lam et al. 1974). Consequently, numerous authors (e.g., Biswas et al. 1995; Myers et al. 1995; Anderson and Kowalczykowski 1998; Kobayashi 1998; Jockovich and Myers 2001; Stahl 2005; Dillingham and Kowalczykowski 2008; Handa et al. 2012; Wigley 2013; Levy et al. 2015; Cheng et al. 2020; Millman et al. 2020; Wilkinson et al. 2022) have stated that RecBCD distinguishes *E. coli*'s DNA from other ("foreign") DNA based on Chi and destroys intracellular foreign DNA while saving *E. coli* DNA. For example, the title of one paper (Cheng et al. 2020) states, "... Chi converts RecBCD from phage destruction to DNA repair," but this and other published papers do not specify any phage that is actually destroyed by RecBCD. We recently reported an extensive, but fruitless, search for such phages (Zheng et al. 2024). Instead, we found four independent isolates of a phage whose growth is aided by RecBCD, a behavior opposite to that of the widespread dogma about RecBCD destroying foreign DNA.

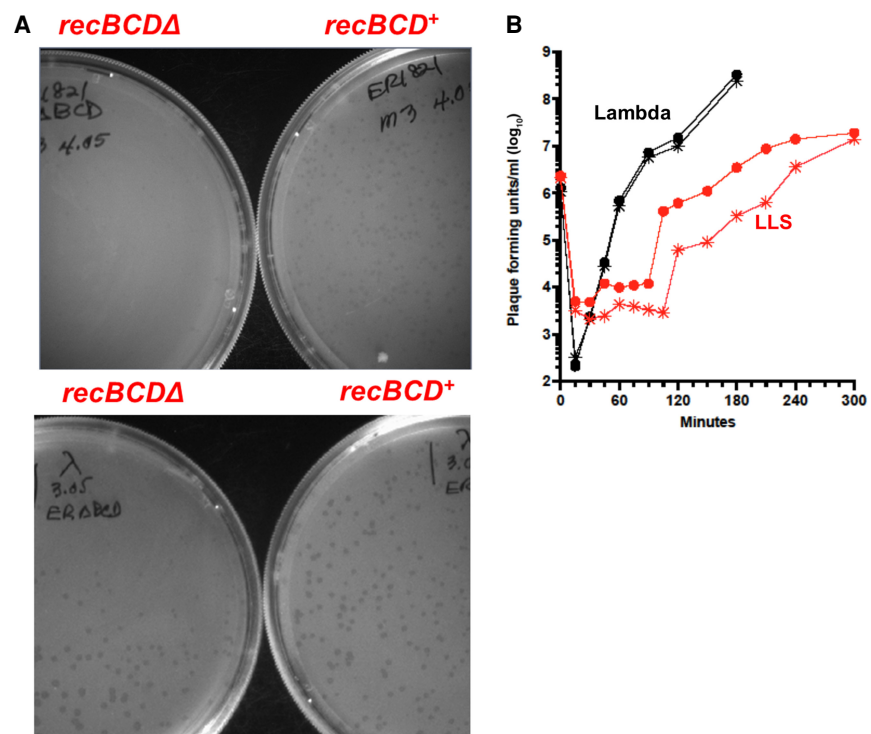
In this report, we describe and analyze the genomes of these four newly isolated lambdoid phages, which carry abundant Chi sites. In parallel with the analysis of these phages, we examined the genome sequences of 313 other lambdoid phages from various *Enterobacteriales* species and of 11 from *Pseudomonas aeruginosa*. To evaluate the interaction of host and phage recombination systems, we grouped these 325 phages based on their recombination

functions and their Chi content. The analysis of these phages and their encoded proteins provides broad perspective and allows discussion of the complex evolution of these phages and their interaction with RecBCD and Chi.

## Results and Discussion

### Four lambdoid phages aided by RecBCD

In our search for phages destroyed by RecBCD in cells, we tested more than 125 sources for phages that could grow in the absence of RecBCD but not in its presence in cells (Zheng et al. 2024). These sources were the feces of warm-blooded animals (mammals and birds), ponds and streams visited by them, and sewage plant effluent. To avoid host restriction of the phages, we plated these samples on *E. coli* strain ER1821 (*recBCD*<sup>+</sup>), which lacks all known restriction enzymes (Jobling et al. 2016), and its isogenic derivative ER1821  $\Delta$ *recBCD*, which lacks RecBCD (Supplemental Fig. S1). Bacterial strains used in this report are described in Supplemental Table S1. Among the more than 80 independent phages tested, we found none that were blocked by the presence of RecBCD enzyme, but we did find four independently isolated phages that made plaques that are more readily visible on *recBCD*<sup>+</sup> than on  $\Delta$ *recBCD*. On both strains, the plaques were small and hazy (Fig. 1A); on some days, they were clearly visible on



**Figure 1.** LLS phage grows less well on the *recBCD* deletion strain than on *recBCD*<sup>+</sup> *E. coli*. (A, top) A suspension of LLS phage containing about 200 phages (pfu) was plated on ER1821  $\Delta$ *recBCD* and ER1821 *recBCD*<sup>+</sup> (wt). After overnight incubation at 37°C, photographs were made. (Bottom) Bacteriophage  $\lambda$  was used similarly as a control, showing that it grows equally well with and without RecBCD, as expected. Phages mouse 3, rabbit 1, and dog 1 behaved like LLS as expected because they have the same DNA sequence. (B) Growth curves show LLS requires RecBCD for optimal growth, but  $\lambda$  does not. ER1821 (*recBCD*<sup>+</sup>; round symbols) and ER1821  $\Delta$ *recBCD* (star symbols) growing in tryptone broth at 37°C were infected with LLS (red symbols) or  $\lambda$  (black symbols) at an moi of about 0.01. Samples were removed at the indicated times into SM with CHCl<sub>3</sub> and titered on ER1821. Similar results were observed in a repeat experiment.

*recBCD*<sup>+</sup> but not visible on  $\Delta$ *recBCD*. These four phages, which behaved similarly, were from a peat bog lake (Larsen Lake, the source of phage Larsen Lake Small [LLS]) and mammalian feces (phages rabbit 1, dog 1, and mouse 3) (Supplemental Fig. S2). The last three phages were isolated about a year after LLS was isolated. We also tested these phages on other isogenic *E. coli recBCD*<sup>+</sup> and *recBCD*-null mutant strains in our laboratory and found similar results.

To quantify RecBCD's aiding the growth of these phages, we examined LLS in more detail with  $\lambda$  as a control. Cells in liquid culture were infected at a multiplicity of infection (moi) of  $\sim 0.01$ . At various times, samples of the cultures were treated with CHCl<sub>3</sub> and titered on ER1821 *recBCD*<sup>+</sup>.  $\lambda$  grew indistinguishably in *recBCD*<sup>+</sup> and  $\Delta$ *recBCD*, but LLS growth was considerably slower in  $\Delta$ *recBCD* relative to *recBCD*<sup>+</sup>, and full phage yield was delayed by  $\sim 1$  h (Fig. 1B). These data are consistent with the plaque size difference mentioned above. LLS also made more readily visible plaques on a *recD* mutant host than on *recB* or *recC* mutant hosts, likely because *recD* mutants are recombination-proficient, although they lack nuclease activity (Amundsen et al. 1986).

We isolated DNA from the above four phages and sequenced their genomes (LLS was sequenced twice). The five sequences, 42,883 bp long, were identical. This sequence (GenBank accession number PQ299149) shows that these phages are related to bacteriophage  $\lambda$  (48.5 kbp) based on their encoded proteins, inferred transcription pattern, and gene expression cascade discussed next (Fig. 2). LLS is thus clearly a lambdoid phage, and its genome has a mosaic relationship to those of other lambdoid phages. Hereafter, we refer to these phages as LLS.

We were initially surprised that the nucleotide sequences of these four phages were identical, because they were isolated at four different sites 3 to 40 km apart (Supplemental Fig. S2), and LLS was isolated about 1 year before the other three. We later noted previous reports of very closely related phages being isolated in even more distant global positions, such as Norway, Greece, and

Chile or such as Puerto Rico and Antarctica, and sometimes across several years (Breitbart et al. 2004; Kalatzis et al. 2017; Bellas et al. 2020). We conjecture that the LLS phage described here has been spread in the Seattle area, perhaps by birds flying among the sites of isolation (Supplemental Fig. S2).

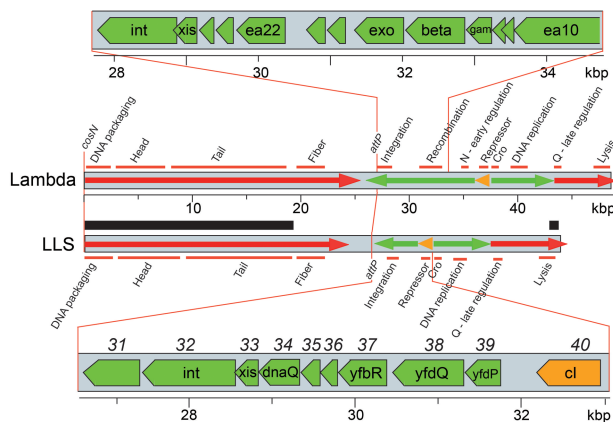
### LLS is a lambdoid phage with novel properties

We analyzed in more detail the relationship of LLS genes to those of other lambdoid phages, proceeding from left to right across the genome (in the standard  $\lambda$  orientation with terminase genes at the left end). The left-most 18.5 kbp and right-most 0.5 kbp of LLS are very similar to lambdoid phage mEp213 (Fig. 2), which was isolated from a human fecal sample in Mexico (Kameyama et al. 1999). This region includes the DNA packaging *cos* site and all head and tail genes except the side tail-fiber genes. Large terminase protein sequences correlate with the type of DNA ends that are generated during packaging (Casjens et al. 2005; Casjens and Gilcrease 2009), and the LLS TerL protein falls with high confidence in the same group as that of bacteriophage  $\lambda$  TerL ( $E=4 \times 10^{-18}$  by BLASTP analysis with 84% query coverage). Thus, like  $\lambda$ , LLS very likely creates cohesive ends with a 5'-overhang. We infer that LLS creates cohesive ends with a 12-nucleotide 5'-overhang (5'-GGGCGGCGCT...-3'), the same as the putative *cos* site of mEp213 (Vahanian et al. 2017; M. Feiss, pers. comm.) but slightly different from the *cos* site of  $\lambda$  (5'-GGGCGGCGACCT...-3', differences underlined) (Sanger et al. 1982). The head proteins of mEp213 have not been studied. They are very different from those of  $\lambda$  but are similar to those of lambdoid phage Gifsy-2, whose pro-head protease and coat protein genes are fused (Effantin et al. 2010). The LLS tail proteins are quite similar to  $\lambda$  tail proteins, but the LLS side-tail-fibers are not closely related to those of any other reported phage. LLS has two adjacent and oppositely oriented putative tail-fiber genes and a DNA invertase gene, suggesting that LLS likely undergoes tail-fiber tip (and thus host receptor) switching similar to that of phages Mu and P1 (Grundy and Howe 1984; Sandmeier et al. 1992).

The integrase (*Int*) gene and *attP* attachment site of LLS were deduced by comparison with known prophages. LLS's *Int* is 99.7% identical to that of a prophage present in the genome of *E. coli* strain FS12139259 (GenBank accession number ABAGKQ01000015); this prophage is integrated in the tmRNA-encoding *ssrA* RNA gene. Thus, LLS contains a 29 bp sequence identical to that of the 3'-end of the *ssrA* gene and should integrate its chromosome into the 3'-end of the *E. coli ssrA* gene without destroying its function, and the putative *attP* site is contained within the LLS bp 26,466–26,494 interval. Curiously, the *E. coli* K-12 mobile element CP4-57 can be integrated in the same host gene and excised by the element's *IntA* protein (Kirby et al. 1994; Trempy et al. 1994; Wang et al. 2009), which is only 17% identical to the LLS *Int*. This appears to be a case in which quite different DNA elements (a phage and a mobile element) have converged to use the same bacterial target.

The lambdoid phages that have been studied experimentally, such as  $\lambda$  and P22, have genes involved in homologous recombination in the middle of the leftward operon expressed early after infection. LLS appears not to have such genes, and this unusual feature is discussed in more detail below.

The closest known relatives of LLS CI repressor and Cro proteins are the cognate proteins encoded by phage mEp460\_4F5 (GenBank accession number LR595868), which are 89% and 78% identical in the two phages, respectively. To our knowledge,



**Figure 2.** Map of the phage LLS genome. The gray rectangles represent phage LLS and bacteriophage  $\lambda$  genomes. Green and red arrows indicate early and late transcription, respectively, and the orange arrows indicate repressor transcripts from the prophages. Gene functions are shown above and below the two maps; gene modules encoding proteins of related function, such as formation of phage heads, are indicated by horizontal orange bars. The thick black bar above the LLS genome marks the region of 98% identity to the phage mEp213 genome (Kameyama et al. 1999). The open reading frames in the early left operon regions are expanded at the top and bottom. Note that  $\lambda$  has a Rec module, with *exo*, *beta*, and *gam*, between the integration genes and the *cl* repressor gene, whereas LLS has genes of unrelated function (see text). LLS genes at the bottom are indicated by names within the green arrows and by numbers above the gray bar.

phage mEp460\_4F5 has not been studied experimentally. Among the phage repressors of known specificity, the LLS repressor is only 47% identical to its closest relative, phage P22 repressor, and this similarity is largely in the C-terminal, non-DNA-binding domain. The operator specificity of LLS repressor remains unknown and is likely novel.

The LLS DNA replication module is unusual among studied lambdoid phages in that it contains only one gene, a homolog of the  $\lambda$  origin-binding O protein. No homolog (or analog) of the  $\lambda$  P helicase loader or other known replication protein is encoded by the LLS replication module. This situation is, however, not unique in the lambdoid phages: 34 members of the 314 *Enterobacteriales* lambdoid phages studied here (see below) encode only a gene O protein homolog with no helicase or helicase loader. None of these 34 phages have been studied experimentally, and replication of this type of lambdoid phage is not yet understood.

No homolog of the  $\lambda$  early antiterminator N protein is found in the LLS genome, so it may have a *put* RNA-mediated early anti-termination mechanism like phage HK022 (Banik-Maiti et al. 1997; Weisberg et al. 1999; Kang et al. 2017). At 56% identical, LLS's late antiterminator Q protein is a "type 5" Q protein according to our earlier analysis of Q proteins (Grose and Casjens 2014), and its closest relative is that of phage  $\lambda$ \_2H10.

The five-gene LLS lysis module (genes 52, 53, 54, 55, and 56) is very similar to that of lambdoid phage 21. It encodes a pinholin that allows release of the signal-anchor-release-type lysozyme through the inner membrane and an antipinholin inhibitor as well as i-spanin and o-spanin proteins that disrupt the outer membrane (see Feiss et al. 2022 and references therein). These analyses establish LLS as a lambdoid phage.

### LLS contains many terminally clustered Chi sites that stimulate recombination

Because LLS grows better on cells with RecBCD than on cells without it (Fig. 1) and because a Chi site mutation enhances growth of  $\lambda$  *red gam* mutants in cells with RecBCD (Lam et al. 1974; Henderson and Weil 1975), we examined the LLS sequence for Chi (5'-GCTGGTGG-3') (Smith et al. 1981). Unexpectedly, we found that LLS has 23 Chi sites, most of which are concentrated near the right end of the genome when aligned with the standard  $\lambda$

map (Fig. 3A). Twenty of them are in the orientation necessary for (1) Chi stimulation of homologous recombination in bacteriophage  $\lambda$  *red gam*  $\chi^+$  mutants and (2) nicking of DNA at Chi by RecBCD enzyme entering the DNA from the right end (as defined in Fig. 2) (Faulds et al. 1979; Taylor et al. 1985). This contrasts sharply with wild-type  $\lambda$ , which has no Chi sites. To our knowledge, the Chi octamer acts only with RecBCD, and the abundance of Chi sites is consistent with LLS requiring RecBCD for optimal growth.

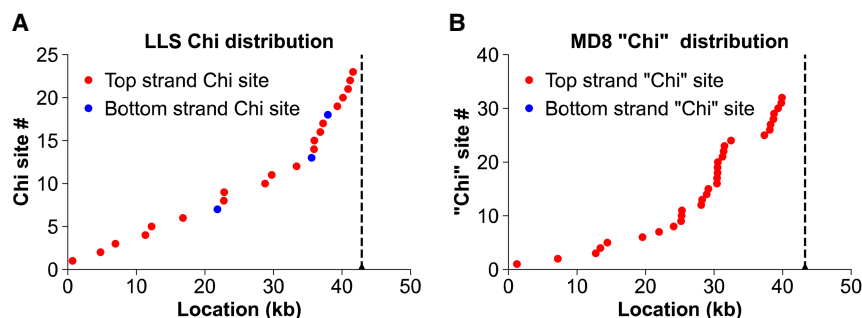
### Lambdoid phage homologous recombination modules

LLS is unusual among lambdoid phages in two seemingly related aspects. It appears not to have genes related to homologous recombination in its early left operon, and it has many Chi sites (Fig. 3A). To explore this finding further and to verify the conclusions, we extended our analysis to homologous recombination genes and Chi site distributions in other lambdoid phages. Each of the early left operons in the well-studied lambdoid phages  $\lambda$ , P22, and HK620 and the defective Rac prophage in *E. coli* K-12 contains a contiguous cluster of homologous recombination genes (Rec modules) (Fig. 2; Table 1). These genes have also been implicated in the switch from circular ( $\theta$  or theta) replication to rolling circle ( $\sigma$  or sigma) replication in bacteriophage  $\lambda$  (Enquist and Skalka 1973; Smith 1983). The genes in the Rec modules of these phages form contiguous clusters with no extraneous genes with unknown or nonrecombination functions. Our examination of this region in a panel of 325 lambdoid phage genomes (for more detail, see Supplemental Material) indicates that genes with some combination of the following four functions are found in the Rec modules of most lambdoid phages: (1) single-strand DNA-binding proteins (SSBs), (2) nucleases, (3) proteins that promote DNA strand-annealing (SAPs), and (4) proteins that inhibit RecBCD helicase-nuclease (anti-RecBCD) (Table 1). For explanations of the Rec protein designation methodology used in this report, see Supplemental Material and Supplemental Table S2. Supplemental Table S3 lists the 325 phages in our lambdoid phage panel, their Rec module type, recombination gene sequence types and protein function types, host species, genome accession number, and information about Chi sites. The proteins that perform each of these functions in the panel phages can be homologs with very different amino acid sequences (different "sequence types") or can be

nonhomologous proteins with similar functions (and thus grouped into different "functional types") (Supplemental Table S4, columns D–G). These in turn allow the definition of module "sequence types" (Supplemental Table S4, columns A,B) and module "functional types" (Supplemental Table S4, column H). The four functional types of recombination proteins are discussed in the following paragraphs.

### Single-strand DNA-binding proteins

The  $\lambda$ , P22, and Rac Rec modules do not encode an SSB, but many other lambdoid phages do (Table 1; Supplemental Tables S3, S4). Although they have not been studied experimentally, Sf6 and ES18 Rec Module proteins have been recognized by sequence homology as robust



**Figure 3.** Phages LLS and MD8 contain many Chi sites, which are concentrated near the right end. Dots are placed at the positions of Chi along the phage genome with equal vertical spacing reflecting their numerical order and showing a high concentration near the right end (vertical dashed line). Red dots are Chi sites in the orientation to activate RecBCD entering the DNA from the right end, as in bacteriophage  $\lambda$  (Kobayashi et al. 1982). (A) *E. coli* phage LLS. Note that the active Chi sites (5'-GCTGGTGG-3' on the top strand) are about seven times more abundant than the inactive ones (20 vs. three). (B) *P. aeruginosa* phage MD8 (Evseev et al. 2021). Note that it has 32 *Pseudomonas* "Chi" sites (5'-GCTGGCGC-3') in the presumably active orientation but none in the opposite orientation. For additional *P. aeruginosa* phages related to MD8, see Supplemental Fig. S5.

**Table 1.** Chi site number and distribution in *Enterobacteriales* lambdoid phages are strongly correlated with absence of anti-RecBCD in diverse Rec modules

Class	Prototype phage <sup>a</sup>	No. of phages	Anti-RecBCD	SSB	Nuclease	SAP	No. of Chi sites (top/bottom strand); range	No. with four or more top-strand Chi sites	No. with right-end clustered Chi sites <sup>b</sup>
A (no Rec module) <sup>c</sup>	Larsen Lake Small (LLS)	68	–	–	–	–	25/0–3/1	62	48
B (no anti-RecBCD)	Gifsy-1	21	–	–	+	+	23/2–0/0	14	4
	Tritos	4	–	–	+	–	21/6–7/2	4	3
	<b>Sum</b>	<b>25</b>						<b>18</b>	<b>7</b>
C (anti-RecBCD)	λ	102	+	–	+	+	15/1–0/0	8	0
	P22	46	+	–	–	+	13/4–0/0	4	0
	HK75	27	+	+	–	+	11/0–0/0	11	2
	ES18	25	+	+	+	+	5/0–0/0	1	0
	Chronis	2	+	+	+	–	0/0	0	0
	<b>Sum</b>	<b>202</b>						<b>24</b>	<b>2</b>
D (other diversity types)	mEp234	7	–	+	–	+	7/0–0/0	2	0
	Pocitis76	6	–	+	+	+	8/0–0/0	1	0
	EhrS_49	3	–	–	–	+	7/0–0/0	1	0
	SEN22	2	–	+	+	–	0/0	0	0
	ZF40	1	+	–	–	–	19/2	1	0
	<b>Sum</b>	<b>19</b>						<b>5</b>	<b>0</b>

<sup>a</sup>The prototype phages of the 13 Rec module functional types illustrated by red brackets in Figure 5 and described in Supplemental Table S4.

<sup>b</sup> $P < 0.1$  for Kolmogorov–Smirnov test for clustering (Massey 1951); phages with fewer than four top-strand Chi sites were not considered to have significant right-end clustering regardless of  $P$  value.

<sup>c</sup>Recombination (Rec) modules are shown in Figure 5 and Supplemental Table S4.

members of the bacterial SSB family (Casjens et al. 2004, 2005), and they have similar AlphaFold 3-predicted structures (Fig. 4A; Abramson et al. 2024). We call this protein sequence type “SSB-1” (Supplemental Table S2). The Rec modules of phage HK97 and phage Chronis encode other very different proteins, gp40 and gp37, whose predicted structures are similar to bacterial SSBs and which we call SSB-1b and SSB-1c, respectively; presumably, these have diverged to the point of having little to no recognizable amino acid sequence similarity to known SSBs (e.g.,  $\leq 15\%$  identical to ES18 SSB-1) (Figs. 4A, 5). The HK620 Rec module encodes protein HhaK (called “SSB-2” here), which binds ssDNA *in vitro* but is not a homolog of bacterial SSB (Fig. 4A; Hutinet et al. 2018).

SSBs are important in many bacterial DNA transactions, and the fact that they are often encoded by genes within lambdoid Rec modules strongly suggests involvement in phage-mediated homologous recombination. *E. coli* SSB is required for bacterial recombination (Glassberg et al. 1979) and recombination in bacteriophage λ *red gam* mutants (Ennis et al. 1987). SSB binds the λ DNA strand-exchange protein Beta (Zakharova et al. 2024) and presumably enhances the activity of Beta and RecA proteins.

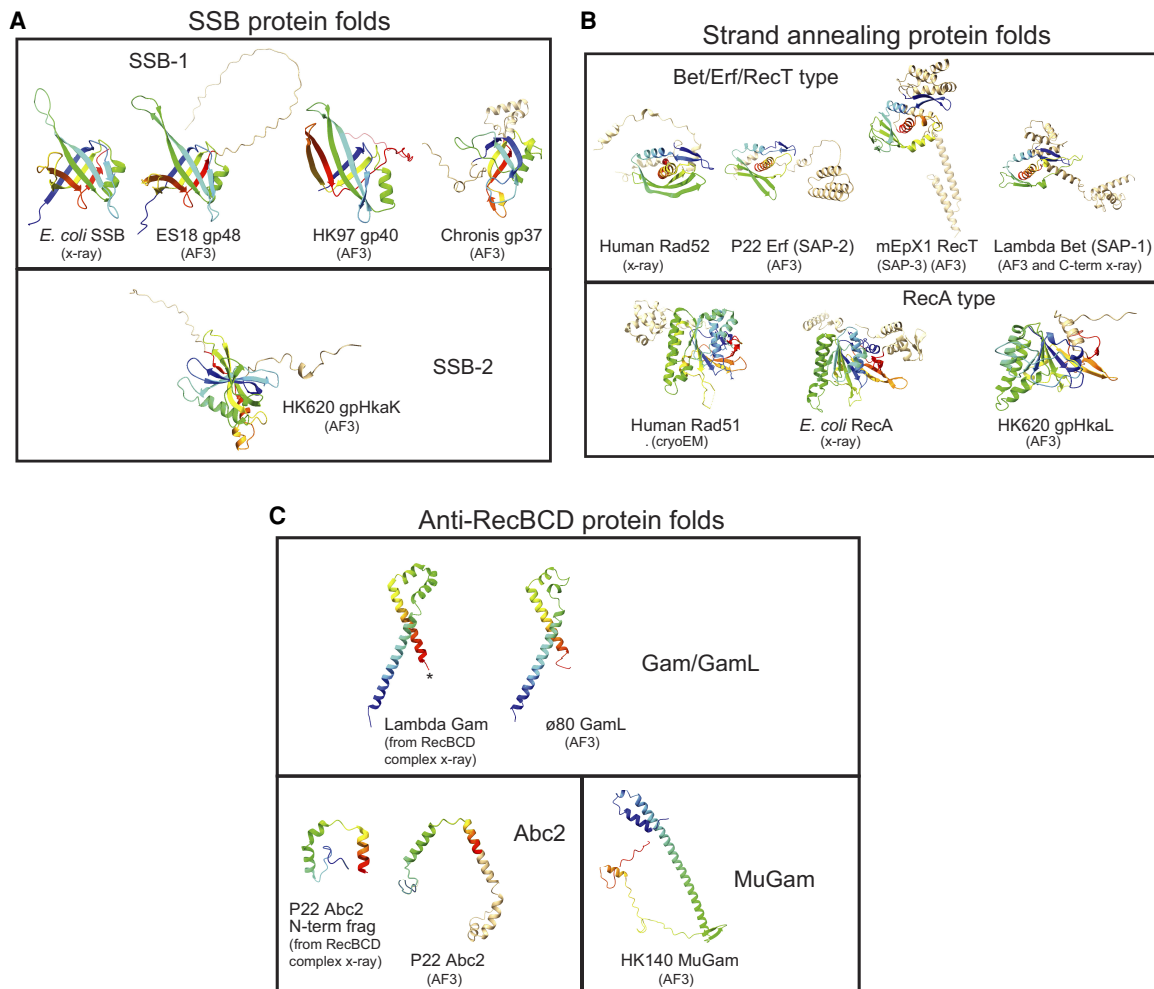
### Nucleases

Bacteriophage λ and prophage Rac Rec modules encode nucleases, but those of P22 and HK620 do not. The Rec modules of the lambdoid panel phages encode four different nuclease types; in addition, two *Pseudomonad* phages encode a nuclease similar to that of RecB, not considered further here. Bacteriophage λ Exo and Rac prophage RecE (also called exonuclease VIII) are 5′-to-3′ exonucleases that promote homologous recombination (Joseph and Kolodner 1983; for review, see Smith 1983). We also discovered

that lambdoid Rec modules can encode DnaQ-like 3′-to-5′ nucleases (Viswanathan and Lovett 1999) that have previously been implicated in DNA polymerase proofreading (Moser et al. 1997) and CRISPR spacer integration (Drabavicius et al. 2018), but, curiously, usually inhibit homologous recombination (Lovett 2011). The fourth nuclease type is composed of HNH-type endonucleases that have been implicated in such varied processes as host defense, intron mobility, DNA packaging and homologous recombination (Xu and Gupta 2013; Kala et al. 2014; Wu et al. 2020). The apparent mobility of some HNH nuclease genes suggests caution in predicting their roles from their gene position. Nonetheless, the tight packing of phage epsilon34 DnaQ and ES18 HNH nuclease genes, for example, with other canonical Rec module genes suggests they are very likely legitimate Rec module members (Fig. 5).

### Strand-annealing proteins

Bacteriophage λ Beta, phage P22 Erf, prophage Rac RecT, and phage HK620 Sak4 (or HkaL) Rec module proteins catalyze strand-annealing and exchange of complementary DNA strands *in vitro* and promote homologous recombination *in vivo* (Botstein and Matz 1970; Lopes et al. 2010; Hutinet et al. 2018; Caldwell and Bell 2019; Brewster and Tolun 2020). The strand-annealing proteins of the lambdoid panel phages can be parsed into eight major sequence types by their strand-annealing domain relationships. Seven of these are structurally related types within the Rad52/Beta/Erf/RecT strand-annealing protein family, called SAP proteins here (Fig. 4B; Supplemental Material; Supplemental Fig. S3C; see also Lopes et al. 2010). We use the uniform nomenclature “SAP-1,” “SAP-2,” etc., for clarity, because the more common names do not indicate their unitary function. The eighth type



**Figure 4.** Polypeptide folds of Rec module proteins. Ribbon diagrams are shown for protein folds for single-strand binding proteins (SSBs; A), strand-annealing proteins (SAPs; B), and anti-RecBCD proteins (C). The diagrams were created with ChimeraX-1.5 (Pettersen et al. 2004) from proteins folded by AlphaFold 3 (AF3) (Abramson et al. 2024) unless otherwise indicated as determined by x-ray diffraction or cryo-electron microscopy. The core folds of proteins are shown in rainbow mode with blue at the N terminus and red at the C terminus. “Extra” polypeptide chain not in the common core fold is shown in tan. PDB accession numbers for experimentally determined structures shown are as follows: *E. coli* SSB, 1SRU and 4M29;  $\lambda$  Gam, 2UJZ and 5MBV; N-terminal fragment of P22 Abc2, 8B1T; C-terminal fragment of  $\lambda$  Beta, 7UJL and 6M9K; human Rad51, 5H1B; human Rad52, 8RIL; and *E. coli* RecA, 7JY9 and 2REB.

belongs to the Rad51/RecA SAP family that has a different polypeptide fold from that of the SAP proteins (Fig. 4B). In addition, three *Pseudomonad* phages encode SAPs not closely related to others studied here.

#### RecBCD inhibitors

The 325 lambdoid phages collectively encode four types of nonhomologous anti-RecBCD proteins; the Gam, GamL, Abc2, and MuGam type proteins inhibit various activities of the host RecBCD helicase-nuclease (Supplemental Tables S3, S4).  $\lambda$  Gam binds RecBCD at its DNA entry site, and MuGam binds DNA ends, so both inhibit all RecBCD activities (d’Adda di Fagagna et al. 2003; Bhattacharyya et al. 2018; Wilkinson et al. 2022). P22 Abc2 binds to the RecC subunit and blocks activation at Chi recombination hotspots (Murphy and Lewis 1993; Wilkinson et al. 2022).  $\lambda$  Gam and ø80 GamL have very different amino acid sequences (i.e., are different sequence types) but are predicted to have similar protein folds (Fig. 4C); GamL blocks RecBCD by an

unknown mechanism but may bind RecBCD like Gam (Rotman et al. 2012).

#### Other possible early left operon DNA transaction proteins

P22 Abc1 and Arf proteins are encoded by genes in its Rec module and aid in, but are not essential for, homologous recombination. Their molecular functions are unknown (Murphy et al. 1987; Poteete et al. 1991). We identified several “RecE-2” proteins that are distantly related to Rac and Gifsy-1 RecE proteins but appear to lack a nuclease domain by AlphaFold 3 polypeptide fold analysis. The roles of such proteins are not known, and these proteins are not considered in detail here.

We considered the possibility of additional unrecognized lambdoid Rec module genes. In addition to the above four proteins, proteins that are very distant relatives of replicon partitioning proteins ParA and ParB are encoded separately in a few lambdoid early left regions, but their very distant relationship to canonical ParA and ParB suggests that their functions may have

changed significantly. ParB-like genes are also sometimes present in the early right operon, something that is not true for the true Rec module genes discussed above. A few lambdoid phages encode RdgC protein homologs in their early left regions. RdgC is a DNA-binding protein that is reported to inhibit RecA-mediated homologous recombination (Briggs et al. 2010); however, its structure is very similar to that of  $\lambda$  RexA protein (Adams et al. 2024), suggesting that it may function in superinfection exclusion and/or prophage induction like  $\lambda$  RexA (Parma et al. 1992; Thomason et al. 2021). Its gene location, when it is present, just left of the *cI* repressor gene in several phages suggests that it may be encoded within the *cI* transcript like RexA. Finally, sporadic HNH nuclease and DnaQ-type nuclease genes are present in some panel phages that do not lie within a recognizable Rec module. When present in the early left operon, the above five DNA-interaction protein types are usually encoded very near *int* at the left end of the early left operon or at the right end near the *cI* gene, rather than in the canonical Rec module location in the middle of the early left operon (Fig. 2). In addition, they are often present in genomes that also have separate canonical Rec modules. Thus, except for the cases in which such a nuclease gene is contiguous with the other Rec module proteins (above), we do not consider these to be Rec module proteins.

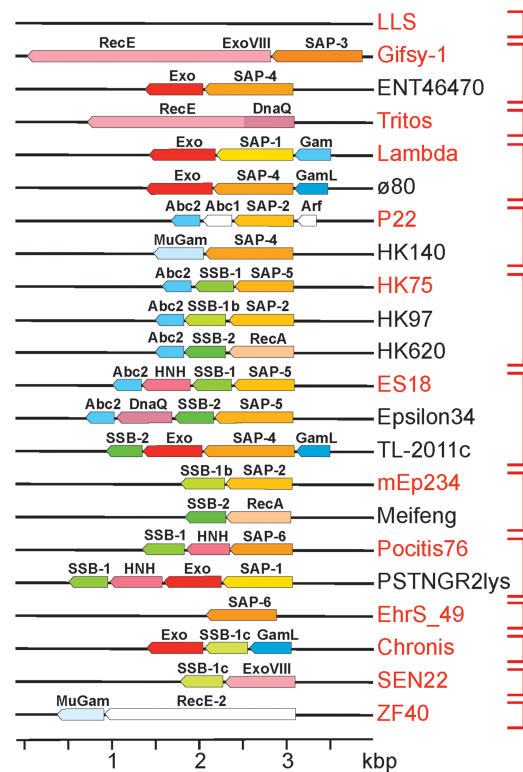
There are very few examples of genes that are neatly placed inside otherwise canonical Rec modules of our lambdoid panel phages whose function cannot be predicted by homology with a protein of known function. However, apparent Rec module protein “SSB-1c” (above) is an example of a protein whose proposed function remains rather speculative. It is neatly inside the Rec module in the phages Chronis and Brookers, for example, and its fold is reminiscent of that of SSB-1 proteins (Fig. 4A). We tentatively include it in the SSB functional group, but additional work will be required to confirm this.

### Rec module diversity

The lambdoid Rec modules are extremely variable, and the phages in our panel carry zero to four putative Rec module genes (Table 1; Supplemental Table S3). The 314 *Enterobacteriales* panel phage Rec modules encode 34 different combinations of protein “sequence types” that can be merged into 13 different module “functional types” (Table 1; Fig. 5; Supplemental Table S4). As might be expected, we find that nearly all putative headful-packaging phages have a Rec module. We have found no other convincing correlations between Rec module gene content and virion morphology, DNA packaging terminase type, or replication protein type, which is consistent with frequent shuffling of Rec and other modules during phage evolution.

The various Rec module protein functional types and sequence types are present in numerous different combinations (Fig. 5; Supplemental Table S4). For example, SAP sequence type 4 can apparently function (1) with no phage-encoded SSB (e.g., in phage HK140) or with SSB-2 (TL-2011c), (2) with no phage-encoded nuclease (HK140) or with Exo ( $\phi$ 80) or RecE (Edno5) type nucleases, and (3) with no anti-RecBCD protein (Edno5) or with GamL ( $\phi$ 80) or MuGam (HK140) (Supplemental Table S4). At least some of this apparent promiscuity is likely owing to shuffling of partner interaction domains, but such an analysis is beyond the scope of this report.

Among the 246 *Enterobacteriales*-infecting panel members with recognizable Rec module genes, only nine lack a SAP gene, whereas 84 lack a nuclease gene, 177 lack an SSB gene, and 43



**Figure 5.** Lambdoid recombination modules. Genes are indicated by pointed boxes drawn approximately to scale as follows: SSB (green), nuclease (red), SAP (orange), and anti-RecBCD (blue). Different shades of the same color indicate different sequence types. Pointed ends on the gene boxes indicate the direction of transcription. Vertical red brackets on the right indicate the 13 module types that encode the same combination of functions regardless of their sequence relationships; the prototype phage for each group (see Supplemental Table S4) is indicated in red text.

lack an anti-RecBCD gene. There are only three phages with more than one gene that encode the same function in the Rec module (GamL and MuGam anti-RecBCD proteins in HK639; SSB-1 [frameshifted pseudogene] and SSB-1c in Brookers; and Exo and HNH nucleases in PSTNGR2lys). It is not possible to be certain of the functionality of the unstudied Rec module genes. In particular, those modules with only one or two genes are somewhat suspect, although they could have evolved to require more host functions than modules with three or four genes. Nonetheless, there is no reason to suspect widespread nonfunctionality, and there appear to be many ways to assemble an apparently functional Rec module.

### LLS and other lambdoid phages with no Rec module

We were initially surprised that 68 (22%) of the genomes in our *Enterobacteriales* lambdoid phage panel have no recognizable Rec module genes (Table 1; Supplemental Tables S3, S4). Unfortunately, none of the phages of this type has, to our knowledge, been studied extensively in the laboratory, so it is not unequivocally known if they truly have no Rec module or if they encode proteins that cannot yet be recognized as having a recombination function; however, no universally present genes of unknown function are present in the early left region of these phages. During preparation of this paper, we noted that Bobay et al. (2013) had analyzed from RefSeq ([ftp://ftp.ncbi.nih.gov/genomes/](http://ftp://ftp.ncbi.nih.gov/genomes/)) the DNA sequences of

237 lambdoid prophages and 38 lambdoid phages of *E. coli* and *Salmonella* for SAPs (homologs of  $\lambda$  Beta, P22 Erf, or HK620 Sak4) and anti-RecBCD proteins (homologs of  $\lambda$  Gam or P22 Abc2); they did not report analysis of Exo or SSB functions. They found that, of these 275 prophages and phages, 134 had no apparent SAP (thus designated “Rec<sup>-</sup>”), and 180 had no apparent anti-RecBCD protein (“Inh<sup>-</sup>”); all 134 Rec<sup>-</sup> were also Inh<sup>-</sup> and may correspond to the “no Rec module” class described here (see below for further discussion).

Phage LLS does not encode any of the known recombination proteins and so belongs to this group, which we designate the “LLS functional type group.” Its early left operon is most closely related to the parallel region of lambdoid phages ST64B (Mmolawa et al. 2003) and  $\phi$ NP (Petty et al. 2011). For example, LLS-encoded proteins gp37, gp38, and gp39 are 76%, 77%, and 90% identical to phage ST64B gp35, gp36, and gp37, respectively. Homologs of these three genes are sometimes called *yfbR*, *yfdQ*, and *yfdP*, respectively; the first is predicted to encode a putative 5'-nucleotide phosphatase, but the latter two have no known function. These gene types are fairly common in the early left operon in lambdoid phages but are not limited to, nor are they universally present in, the phages that lack a canonical Rec module. For example, this cluster is present in *E. coli* lambdoid phage ArgO145 (Krüger et al. 2018), which also carries a complete and separate Rec module with all four functions (Supplemental Table S3). Thus, it seems unlikely that these genes encode currently unrecognizable recombination proteins. We also note that some lambdoid phages carry additional recombination-related proteins in their distal early right operon “*nin*” region; however, LLS has no such gene there.

### The LLS genome has unexpectedly abundant Chi sites

The genomes of the well-studied lambdoid phages  $\lambda$ , 21, 434, Sf6, and P22 contain no Chi recombination hotspot sites, and in our lambdoid phage panel (Supplemental Table S3), the genomes of 172 of 314 *Enterobacteriales* phages have only one or no Chi site. Thus, it was somewhat surprising, as noted above, that the LLS genome contains 23 Chi sites (one site per 1.86 kbp on average). The LLS Chi sites are not randomly distributed: 20 are in the orientation that is active in  $\lambda$  Chi-containing mutants (Faulds et al. 1979; Taylor et al. 1985), and half (10) occur in the rightmost

22% of the genome (as oriented in Fig. 2; Fig. 3A). The LLS Chi density is 2.5 times higher than the 1008 Chi sites in *E. coli* K-12's 4.6 million bp (one per 4.6 kbp). Random nucleotide association of 50% G+C DNA 50 kbp long, approximately as in nearly all phage in our panel (Supplemental Table S3), predicts 0.85 Chi sites per genome. These observations suggest that LLS has actively accumulated Chi sites.

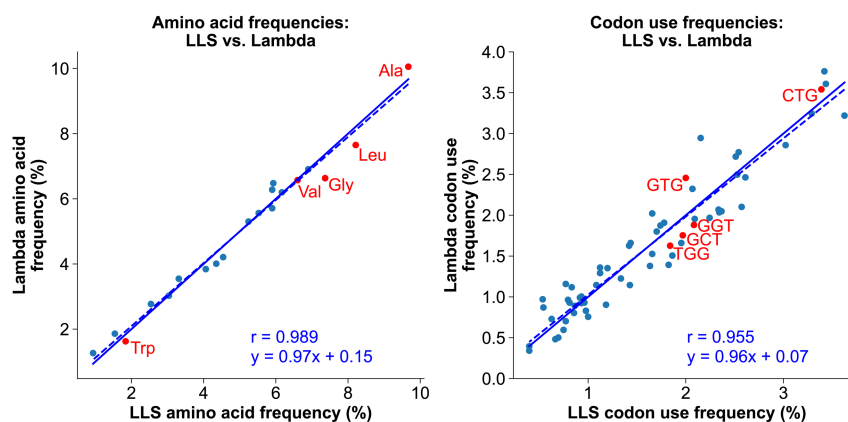
Bobay et al. (2013) noted that some of the 275 lambdoid prophages and phages mentioned above have up to 27 Chi sites, and these phages are generally Rec<sup>-</sup> Inh<sup>-</sup>. The 180 Inh<sup>-</sup> class have on average about 2.5 times more Chi sites than expected from their trinucleotide content, whereas the 95 Inh<sup>+</sup> class, all of which are Rec<sup>+</sup>, have fewer Chi sites than expected. We note, however, that the prophages, which account for 86% of the sequences Bobay et al. analyzed, could include as many as 15 insertion sequences, such as IS3. These sequences likely represent cryptic (inactive) prophages, which would be expected to accumulate Chi just as the rest of the *E. coli* chromosome does. Our analysis below revealed high frequencies of Chi sites in active lambdoid phages infecting a wide variety of *Enterobacteriales* strains and *Pseudomonas*.

We note that, although temperate phage P1 is not a lambdoid phage, it also carries numerous Chi sites (50 in its 94.8 kbp genome, or one per 1.90 kbp) and requires RecBCD for growth (Sternberg and Hoess 1983). The Chi sequence 5' GCTGGTGG 3' is the most frequent octamer in both the LLS and P1 genomes (Supplemental Table S3; Subramaniam and Smith 2022). The next most frequent octamer in each phage (5'-TGCTGGTG-3') occurs 16 times in LLS and 32 times in P1 and shares seven contiguous nucleotides with Chi (underlined if in Chi sequence). In  $\lambda$ , which lacks Chi, the most frequent octamer (5'-GCTGGCTG-3') occurs 18 times and shares only five contiguous nucleotides (underlined) with Chi. The relatively high frequency of these octamers related to Chi could mean that natural mutational processes cause functional Chi sites to be inactivated frequently and inactive sites to mutate to active Chi sites in these phages; namely, Chi sites might be expected to come and go over evolutionary time as modules are shuffled, as discussed further below.

Chi contains the most frequent codon (CTG) for leucine, which is the most or second-most frequent amino acid in proteins of LLS,  $\lambda$ , P1, and their *E. coli* host (Fig. 6; Subramaniam and Smith 2022). Codon usage might thereby account for Chi's high frequency in LLS, but a comparison of  $\lambda$  and LLS argues against this view. The frequencies of amino acids and codons are nearly the same in  $\lambda$  and LLS (Fig. 6), indicating that preferential codon usage does not account for Chi being abundant in LLS but absent from  $\lambda$ . We infer that RecBCD's positive role in the growth of P1 and LLS results in Chi being positively selected to activate RecBCD for recombination.

### LLS represents an important class of lambdoid phages that lack a Rec module

As described above and in Figure 2, analysis of the LLS genome sequence indicated that it is a typical lambdoid phage except that it appears to have no Rec module and has abundant Chi sites. Among the 314 previously sequenced



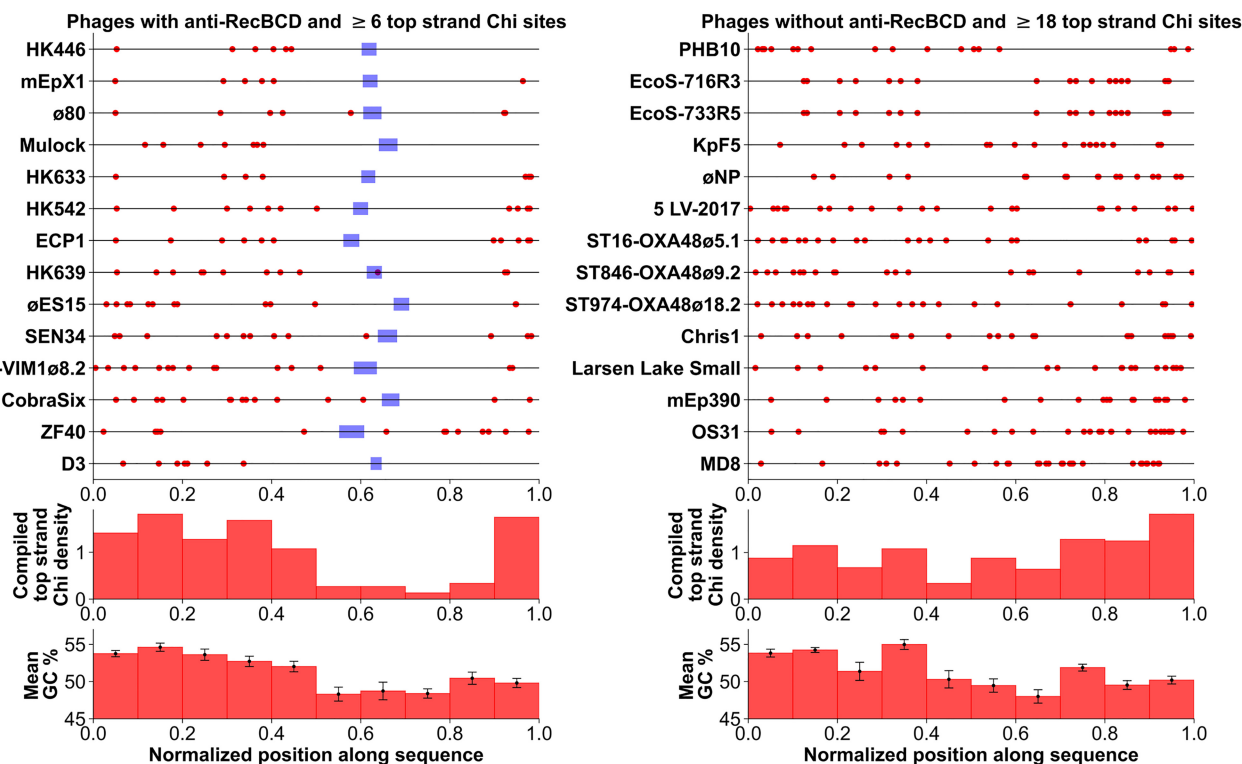
**Figure 6.** Amino acid and codon frequencies are highly similar in bacteriophage  $\lambda$  (with no Chi sites) and phage LLS (with 23 Chi sites). The frequencies of encoded amino acids (left) and codons (right) in  $\lambda$  are plotted versus their frequencies in LLS. Note that the slope of the linear regression line (dotted line) is nearly equal to one, which would indicate equality (solid line). The Pearson regression coefficients ( $r$ ) indicate high correlation. Red dots indicate codons within Chi (right) and their encoded amino acids (left).

*Enterobacteriales* lambdoid phage genomes that we analyzed, 68 have no recognizable Rec module genes (Table 1; Supplemental Table S3). Sixty-two of these phages have eight or more Chi sequences in the proposed active orientation (up to 25 in phage OS31) in their 29.3–58.3 kbp genomes, as well as six or fewer Chi sequences in the inactive orientation. We predict that these phages will be found to have growth requirements similar to those of LLS. These phages infect diverse bacteria across the *Enterobacteriales* order, ranging from *Providencia* to *Hafnia* to *E. coli*, and four others infect a *Pseudomonad* (see below). Thus, this little-recognized class of  $\lambda$ -like phages appears to be widespread in nature.

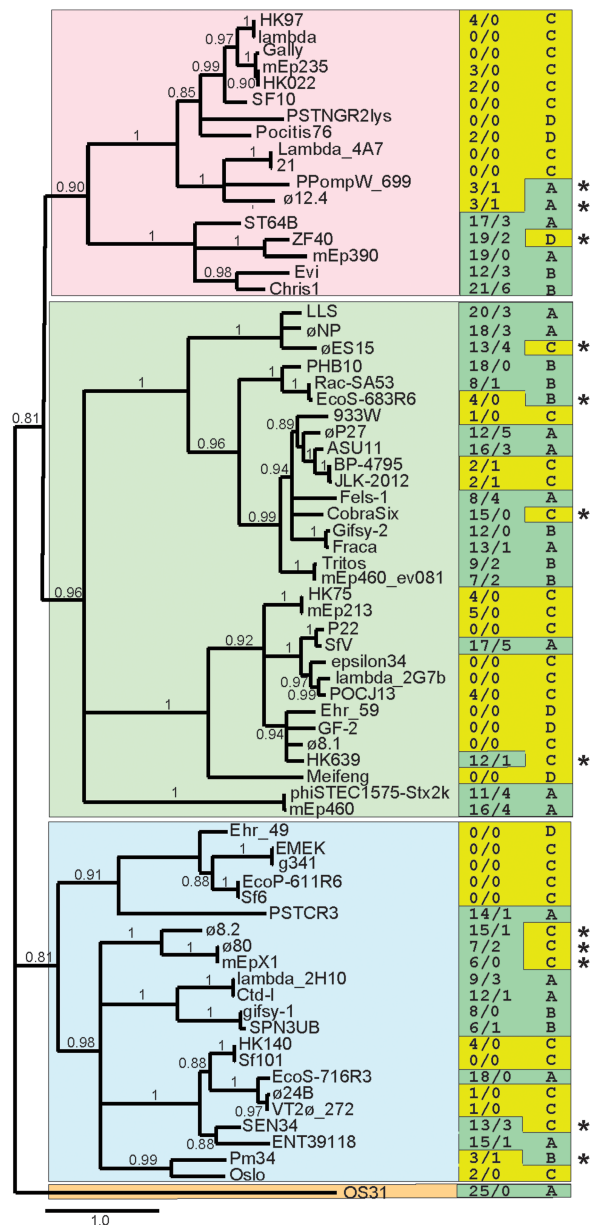
This “no Rec module” class of lambdoid phages typified by LLS likely uses their abundant Chi sites and RecBCD enzyme to their advantage, as is the case for the unrelated phage P1 and its relatives (Subramaniam and Smith 2022). In both cases, we propose that Chi stimulates RecBCD to promote recombination of linear DNA to help form concatemeric DNA essential for packaging and production of complete viable phage virions. This scenario can account for the high frequency and preferential location of the multiple Chi sites toward the phages’ right DNA ends (Fig. 3A; Supplemental Table S3) and the enhancement of LLS, but not  $\lambda$ , growth by RecBCD (Fig. 1).

A few phages among the 314 analyzed above have both a Rec module and abundant Chi sites (Fig. 7) or have no Rec module and relatively few Chi sites (Supplemental Table S3). Such phages could have recently acquired or lost a Rec module region from a phage with a very different Chi content. Such a recombination event could have, for example, placed an intact Rec module and flanking DNA lacking Chi (e.g., from  $\lambda$ ) into a Chi-containing phage without a Rec module (e.g., LLS) and produced a phage with a Rec module but few or no adjacent Chi sites (e.g., phage  $\phi$ ES15). As this phage evolves further, it may eliminate Chi, as  $\lambda$  and its many relatives appear to have done. On the other hand, phages with few or no Chi sites could lose their Rec module, and future evolution might favor accumulation of Chi sites. Such a model predicts that before Chi optimization has occurred, the Chi density in the Rec module region would correlate with the putative parental phage from which it was acquired.

To test this idea, we examined the 14 “unusual” cases of phages that carry a Rec module and have the most abundant Chi sites (six or more on the top, putatively active strand). Among these, most had several kilobase pairs of DNA flanking the Rec module on both sides that lack Chi (Fig. 7). The distribution of Chi in these phages is highly significantly different from that of the 14 phages without a Rec module or lacking an anti-RecBCD function and



**Figure 7.** Rare, exceptional phages with both anti-RecBCD and many Chi sites often lack Chi near the Rec module. The positions of Chi sites (red dots) are placed on the genome sequence of the indicated phage as a fraction of the genome length, to facilitate comparison of genomes with different lengths. (Left) All 13 *Enterobacteriales* phages in the lambdoid panel (Supplemental Table S3) with an encoded anti-RecBCD function and six or more Chi sites are plotted; their Rec module is blue. (Right) All 13 *Enterobacteriales* phages in Supplemental Table S3 without an encoded anti-RecBCD function and 18 or more Chi sites are plotted. *P. aeruginosa* phages MD8 and D3 are included for comparison but are not included in the histograms of G+C content because of their high G+C content reflecting that of their host *P. aeruginosa*. The histogram directly below each Chi position plot shows the density of Chi sites in each 10% bin of the genome. Note that in the left panel there are many fewer Chi sites near the Rec module than far to one side or the other. The distributions in the left and right panels are significantly different ( $P < 0.0001$  by  $\chi^2$  test). This outcome is consistent with this region having recently come from a typical lambdoid phage, such as  $\lambda$  or P22, with a Rec module but without Chi sites (Supplemental Table S3). The histogram below each Chi density histogram shows the mean G+C content in each 10% bin of the genome, with error bars representing standard error. In both the left and right panels, the G+C content of the rightmost three bins is significantly less than that of the leftmost seven bins ( $P < 0.001$  by paired two-sample *t*-test).



**Figure 8.** Correlation of integrase protein sequence with occurrence of Rec modules and Chi sites. BLASTP analysis (Altschul et al. 1997) was used to identify the integrase proteins encoded by the *Enterobacteriales* phages in Supplemental Table S3. A sample of these proteins, chosen before (i.e., independently of) analysis of the phages' Rec modules and Chi sites, that represents the extent of the panel's integrase diversity was used to generate a maximum-likelihood phylogenetic tree at <http://www.phylogeny.fr/> (PhyML program as implemented by Dereeper et al. 2008). Phage names are given at the right of each branch; bootstrap values (0.0 to 1.0) are shown in the tree; branches with bootstrap values less than 0.80 are collapsed; a scale bar of estimated evolutionary distance (changes per site) is shown at the bottom of the figure; and the four major integrase clades are indicated by different background colors. In the right-hand columns, the numbers of Chi sites in the two strands (upper/lower) (Supplemental Table S3) are indicated on the left, and Rec module type as defined in Table 1 is shown on the right. Green and yellow left column background colors highlight high and low numbers of Chi sites, respectively; green and yellow right column backgrounds indicate Table 1 classes whose members usually have high and low numbers of Chi sites, respectively. Asterisks mark 12 phages in which the number of Chi sites is unusual for its Rec module class; eight of these are in Figure 7 (left panel), which suggests they are evolving between Class A or B and Class C.

with the most abundant Chi sites (18 or more on the top, putatively active strand). These observations support the proposed shuffling of Rec modules and Chi sites during evolution to have one or the other but not both or neither. No phages without a Rec module or with only an anti-RecBCD function lack Chi on the top strand (Supplemental Table S3), further pointing to the importance of Chi-stimulated host recombination in the life cycle of phages without their own recombination machinery.

We also tested the possibility that, in the phages without a Rec module, the higher frequency of Chi near the right end of the phages (Fig. 7, right panel) than elsewhere in these phages stems from a higher G+C content on the right (Chi is 75% G+C). We observed the opposite; the G+C content is significantly lower ( $P < 0.001$  by paired two-sample *t*-test) in the three right-most intervals with high Chi density than elsewhere in the genomes. This negative correlation of Chi and G+C content supports our conclusion that Chi is actively selected for in LLS and its relatives without a Rec module.

The proposed shuffling of Rec modules among lambdoid phages is also supported by analysis of a phylogenetic tree of Int proteins (Fig. 8). Int is encoded by a gene close to the Rec module location (Fig. 2) and is sufficiently long (e.g., 391 amino acids in phage LLS) to allow reliable tree building. In multiple cases, a clade of five or more diverse Int types has either a Rec module and few or no Chi sites (e.g., 12 phages with  $\lambda$  subtype of Int) (top portion of branch with pink background in Fig. 8) or no Rec module and many (12 or more) Chi sites (e.g., five phages with ST64B subtype; bottom portion of branch with pink background). Other cases of Int divergence with maintenance of "Rec module-Chi deficiency" indicate that shuffling of Int and Rec has occurred repeatedly, we suppose by recombination between distinct lambdoid phages, such as between  $\lambda$  and LLS as suggested above.

### LLS growth enhancement by RecBCD

LLS grows less well on cells lacking RecB or RecC but grows equally well, or perhaps better, on cells lacking RecD compared with growth on wild type. Mutants that lack *recB* or *recC* or both have the same cellular phenotype. They lack helicase and nuclease activities of RecBCD and are deficient for recombination and DNA break repair (for review, see Amundsen and Smith 2023). On the other hand, *recD* mutants have a markedly different phenotype; they retain helicase and RecA-loading activities and are recombination-proficient even though they lack nuclease and Chi activation. These phenotypes could account for LLS growth differentials.

We suppose the requirement for RecB and RecC, but not RecD, for optimal growth of LLS reflects RecBCD enzyme being needed to promote Chi-stimulated recombination. This proposal is consistent with the high density of Chi sites near the right end of these DNAs (Fig. 3A). This is the end of  $\lambda$  DNA that is available to RecBCD after DNA packaging is initiated by the terminase complex (Kobayashi et al. 1982). As noted above, we infer that LLS packages DNA as  $\lambda$  does, with its left end occluded by terminase but with the right end open to RecBCD. RecBCD entering the right DNA end encounters Chi from its 3' side when Chi is on the "upper" DNA strand, that with the 3'-end on the right. Such an encounter in  $\lambda$  activates RecBCD, but a Chi site in the opposite orientation in  $\lambda$  is inactive, because RecBCD encounters it from the opposite direction (Faulds et al. 1979; Taylor et al. 1985). Twenty of the 23 Chi sites in the LLS genome are in the active orientation if RecBCD enters from the right and proceeds leftward (Fig. 2). This high nonrandomness strongly suggests that Chi sites

are important for LLS propagation in wild-type *E. coli*. We suppose that in *recD* mutants, in which LLS grows well, recombination is still required and proceeds in a Chi-independent manner; alternatively, rolling-circle replication may occur in *recD* mutants, which lack RecBCD nuclease activity (Enquist and Skalka 1973; Amundsen et al. 1986).

### “No Rec module–high Chi density” feature is not limited to *Enterobacteriales* phages

During these sequence analyses, we noted that phage MD8 (GenBank accession number KX198612) of *P. aeruginosa* (Evseev et al. 2021), a member of the Pseudomonadales bacterial order, has a lambdoid gene arrangement but lacks a Rec module and has 32 “Chi” sites (5′-GCTGGCGC-3′), all on the “top,” putatively active strand. This DNA sequence, identical to *E. coli* Chi at six of the eight positions (underlined), is site-specifically cut by the purified *Pseudomonas syringae* RecBCD enzyme and stimulates homologous recombination in *E. coli* expressing *P. syringae* RecBCD (Pavankumar et al. 2010, 2018). Phage MD8, originally isolated from Lake Baikal in eastern Russia (Evseev et al. 2021), strictly requires RecA for growth but not RecB (Supplemental Fig. S4); RecC and RecD mutants have not been tested. We propose that MD8, like phage P1, strictly requires host-promoted recombination for growth and that this recombination is stimulated by *P. aeruginosa* RecBCD acting at the “Chi” sites, which are concentrated near the right end of the MD8 genome as for LLS and many other lambdoid phages lacking a Rec module (Figs. 3, 7). MD8 lacks all potential parts of the early left operon Rec module, but it does have a  $\lambda$  NinG homolog encoded in its early right operon, which in the case of  $\lambda$  strongly aids RecBCD-promoted recombination between the phage and a plasmid (Hollifield et al. 1987).

Three other *P. aeruginosa* phages (JBD68,  $\Phi$ 2, and F10) related to MD8 (Evseev et al. 2021) also lack a Rec module and have abundant “Chi” sites all on the top strand, as in MD8 (Supplemental Fig. S5; Supplemental Table S3). In the absence of RecBCD nuclease activity (i.e., in the *recB*-null mutant tested), MD8 and perhaps these other similar phages may readily switch from theta to rolling-circle replication and thereby produce the concatemeric DNA needed for packaging, as in bacteriophage  $\lambda$  (for review, see Smith 1983). Although these points remain to be tested, they are consistent with the idea that MD8 and its relatives, like LLS and other phages noted above, are not destroyed by RecBCD and have evolved abundant “Chi” sites to use RecBCD for their benefit. Thus, as previously noted, the Chi-RecBCD paradigm appears to extend beyond the enteric bacteria, of which all tested species use Chi as a recombination hotspot (Schultz and Smith 1986; Smith et al. 1986; McKittrick and Smith 1989; Subramaniam and Smith 2022).

In phage MD8, “Chi” is the most frequent octamer among its 37,849 unique octamers. In the genome of its host, *P. aeruginosa* PAO1 (GenBank accession number AE004091.2), “Chi” is only the 43rd most abundant octamer among its 65,499 unique octamers. In MD8, “Chi” occurs 32 times, all in the presumed active orientation and concentrated near the right end (Fig. 3B); “Chi” is also significantly concentrated near the right end of  $\Phi$ 2 (Supplemental Fig. S5). The fourth most abundant octamer in MD8 (5′-GCCGGCGC 3′) occurs 26 times and shares 7 nucleotides with *Pseudomonas* “Chi,” suggesting that MD8 may evolve further to increase its “Chi” frequency and enhance its recombination by RecBCD. In contrast, *P. aeruginosa* phage D3 (GenBank accession number AF165214.2) (Kropinski 2000) is a lambdoid with a Rec module and has only seven top-strand (but no bottom strand)

“Chi” sites, all in the left one-third of the genome (Fig. 7, left). Thus, D3 has the features of several *Enterobacteriales* lambdoids noted earlier, such as HK639 (Fig. 7, left), with both a Rec module and distant Chi sites. D3 may have also recently arisen from recombination between phages with and without a Rec module but differing in “Chi” content, as proposed above for HK639 and its relatives (Fig. 7, left). Six other *P. aeruginosa* phages (Supplemental Table S3) have a variety of Rec modules and “Chi” site abundances similar to those of the *Enterobacteriales* phages discussed above, indicating that the patterns noted for the *Enterobacteriales* phages extend to Pseudomonad phages.

The findings reported here are consistent with the idea that phages LLS and MD8, like phage P1, enrich Chi in their genomes to promote recombination mediated by RecBCD.  $\lambda$  encodes its own recombination machinery, which acts independently of Chi, and has thus not enriched its genome for Chi. The absence of a Rec module and the abundance of Chi therefore distinguish this LLS class from the other well-studied lambdoid classes of phages, whose Rec module genes we have extensively analyzed and categorized. The diversity of lambdoid phages likely reflects frequent gene exchanges among them by homologous recombination, which is strongly influenced by the Rec modules themselves.

## Methods

### Bacterial strains

Strains, derivatives of *E. coli* K-12 or *P. aeruginosa* PAO1, and their genotypes are listed in Supplemental Table S1.

### Growth media and phage suspension medium

Cells were grown in tryptone broth or on tryptone agar plates with 1.0% agar in the bottom layer and 0.75% agar in the top layer. MgCl<sub>2</sub> or MgSO<sub>4</sub> (10 mM) was sometimes added but made little difference in phage or bacterial behavior. Incubation was at 37°C. Phage were suspended and diluted in suspension medium (20 mM Tri-HCl at pH 7.4, 1 mM MgSO<sub>4</sub>, 0.5% NaCl, 0.01% gelatin), designated SM.

### Phage isolation and analysis

Fecal samples were mixed with about 3 volumes of water, agitated to make a uniform slurry, and centrifuged at 21 kg for 1 min. The supernate was passed through a 0.22 micron filter (Millipore); liquid samples, such as pond water or sewage, were similarly filtered. A small volume, typically 0.2–0.5 mL, was mixed with 0.1 mL of freshly grown host cells and incubated for 10 min at 37°C. Top agar (2.5 or 3 mL) was added, and the mixture poured onto a bottom agar plate (~35 mL). In some cases, the samples were diluted before plating to result in about 200 plaques per plate. In other cases, ~10  $\mu$ L samples were spotted onto lawns of host cells. Plates were incubated overnight at 37°C.

Phage from individual plaques of each type (clear or turbid, large or small, etc.) from a given source were picked with a Pasteur pipette into 1 mL of SM with ~30  $\mu$ L of chloroform. After about 5 min with occasional gentle shaking, a sample of the aqueous supernate was spotted onto lawns of bacteria and streaked with a fine needle. Phages showing differential growth on *recBCD*<sup>+</sup> and  $\Delta$ *recBCD* cells were purified by picking an isolated plaque to SM or by streaking with a fine platinum–iridium needle directly from a plaque onto a fresh lawn of host cells. After purification, a confluent spot of phage was scraped from the plate with a sterile spatula, suspended in SM with chloroform, and stored at 4°C.

## DNA preparation, sequencing, and analysis

Phage were grown in 25 mL cultures of strain ER1821 (initial OD<sub>650</sub> about 0.5 and moi about 0.1) and harvested after 3–4 h of vigorous shaking (or overnight) at 37°C. In some cases, bacteria were concentrated 10-fold by centrifugation before infection and subsequently diluted for growth. CHCl<sub>3</sub> (~1% vol/vol) was added and, after ~10 min, the culture was centrifuged at 10 kg for 5 min. The supernate was titered, and DNA was extracted from ~2 × 10<sup>10</sup> plaque-forming units with a modification of the protocol in the New England Biolabs Monarch genomic DNA purification kit. Briefly, phage lysate was treated with RNase and DNase I and incubated with 10% PEG-8000 and 1 M NaCl overnight at 4°C, followed by centrifugation and resuspension of the pellet. The resuspension was treated with Proteinase K and Monarch tissue lysis buffer. DNA was precipitated using sodium acetate, ethanol, and Novagen's pellet paint coprecipitant.

DNA (~1 µg) was sequenced by Plasmidsaurus using Nanopore sequencing methods. Sequences were compiled from about 25–100 overlapping reads ~1 to ~40 kbp long. LLS was sequenced twice, and the other three isolates (rabbit 1, dog 1, and mouse 3) once. Sequencing of PCR amplicons was used to resolve any ambiguities. These sequences did not include the ends of the DNAs surrounding the *cos* site, so DNA was PCR-amplified using primers with 3'-ends at nucleotides 93 ("leftward" directed) and 42,707 ("rightward" directed), which produced a fragment that crossed the *cos* site (see text) and allowed completion of the sequences.

Sequences were analyzed and compared using SnapGene (<https://www.snapgene.com>), PHASTER (Arndt et al. 2016), PHASTE (Wishart et al. 2023), DNA Strider (Douglas 1994), and NCBI-BLAST (Altschul et al. 1997).

## Analysis of Rec module proteins and Chi sites

Our panel of 325 lambdoid phage genomes and the details of their analysis are described in the [Supplemental Material](#). Recombination proteins encoded by these genomes were identified by systematic psiBLAST (at the University of Utah) (Altschul et al. 1997) and hidden Markov models (HMMs; at University of Toronto, Fred Hutchinson Cancer Center, and MIT) (Finn et al. 2015). In addition, structures predicted by AlphaFold 3 (Abramson et al. 2024) and Clustal X neighbor-joining trees (Larkin et al. 2007) were used as sensitive tests of structural similarity. For descriptions of the results of these analyses, see [Supplemental Tables S2–S4](#).

Statistics regarding the abundance and distribution of Chi sites were computed using Python 3.12 (<http://www.python.org/>) with Numpy 2.0.0 (Harris et al. 2020). Briefly, the rank of Chi out of all octamers was calculated based on each octamer's number of occurrences. The number of Chi sites on each strand was counted, and indices for all Chi sites were extracted. Subsequently, Kolmogorov–Smirnov testing (Massey 1951) was used to classify the top-strand Chi sites as left-end clustered, right-end clustered, or not significantly clustered to either end.

## Data access

All sequencing data generated in this study have been submitted to the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession number PQ299149.1. Phage LLS is available at the Félix d'Hérelle Reference Center for Bacterial Viruses at the University of Laval, Quebec, Canada (<https://www.phage.ulaval.ca>) under phage HER number 843.

## Competing interest statement

The authors declare no competing interests.

## Acknowledgments

We thank Curtis Furukawa for isolating and initially characterizing with C.Z. phage LLS; Jerry Liu for rabbit 1 and dog 1 phages, and Han Lin for mouse 3 phage; Lise Raleigh (New England Biolabs) for *E. coli* strain ER1821; Konstantin Miroshnikov (Russian Academy of Sciences) and Konstantin Severinov (Rutgers University) for phage MD8; Colin Manoil (University of Washington) for *P. aeruginosa* PAO1 and its *recA* mutant; Alex Hong and Joe Bondy-Denomy (University of California at San Francisco) for *P. aeruginosa* PAO1 *recB* mutant; Graham Hatfull (University of Pittsburgh) for phage hunting advice; Mike Feiss (University of Iowa) for analysis of the mEp213 *cos* site; Tom Milac and Randy Hyppa for helpful comments on the manuscript; Sandra Weller, David Evans, Harmit Malik, and Pravurtha Raman for helpful discussions of human viruses and phylogenetic trees; and the Interlake High School internship program for C.Z. This research was supported by U.S. National Institute of General Medical Sciences grant R35 GM118120 to G.R.S. and operating grant FDN-15427 from the Canadian Institutes of Health Research (CIHR) to A.R.D., who is also supported by a Tier 1 Canada Research Chair (950-232058).

**Author contributions:** G.R.S. conceived and supervised the study. C.Z., S.K.A., and G.R.S. designed the experiments. C.Z., S.R.C., and S.K.A. performed all the experiments and visualization. C.Z., S.R.C., and A.R.D. performed the data analysis. G.R.S. and S.R.C. wrote the original draft. All authors revised the manuscript and read and approved the final manuscript.

## References

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**: 493–500. doi:10.1038/s41586-024-07487-w
- Adams MC, Schiltz CJ, Sun J, Hosford CJ, Johnson VM, Pan H, Borbat PP, Freed JH, Thomason LC, Court C, et al. 2024. The crystal structure of bacteriophage lambda RexA provides novel insights into the DNA binding properties of Rex-like phage exclusion proteins. *Nucleic Acids Res* **52**: 4659–4675. doi:10.1093/nar/gkac212
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402. doi:10.1093/nar/25.17.3389
- Amundsen SK, Smith GR. 2023. RecBCD enzyme: mechanistic insights from mutants of a complex helicase-nuclease. *Microbiol Mol Biol Rev* **87**: e0004123. doi:10.1128/mmb.00041-23
- Amundsen SK, Taylor AF, Chaudhury AM, Smith GR. 1986. *recD*: the gene for an essential third subunit of exonuclease V. *Proc Natl Acad Sci* **83**: 5558–5562. doi:10.1073/pnas.83.15.5558
- Anderson DG, Kowalczykowski SC. 1998. Reconstitution of an SOS response pathway: derepression of transcription in response to DNA breaks. *Cell* **95**: 975–979. doi:10.1016/S0092-8674(00)81721-3
- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**: W16–W21. doi:10.1093/nar/gkw387
- Banik-Maiti S, King RA, Weisberg RA. 1997. The antiterminator RNA of phage HK022. *J Mol Biol* **272**: 677–687. doi:10.1006/jmbi.1997.1251
- Bell JC, Kowalczykowski SC. 2016. RecA: regulation and mechanism of a molecular search engine. *Trends Biochem Sci* **41**: 491–507. doi:10.1016/j.tibs.2016.04.002
- Bellas CM, Schroeder DC, Edwards A, Barker G, Anesio AM. 2020. Flexible genes establish widespread bacteriophage pan-genomes in cryoconite hole ecosystems. *Nat Commun* **11**: 4403. doi:10.1038/s41467-020-18236-8
- Bhattacharyya S, Soniat MM, Walker D, Jang S, Finkelstein IJ, Harshey RM. 2018. Phage Mu Gam protein promotes NHEJ in concert with *Escherichia coli* ligase. *Proc Natl Acad Sci* **115**: E11614–E11622. doi:10.1073/pnas.1816606115
- Biswas I, Maguin E, Ehrlich SD, Gruss A. 1995. A 7-base-pair sequence protects DNA from exonucleolytic degradation in *Lactococcus lactis*. *Proc Natl Acad Sci* **92**: 2244–2248. doi:10.1073/pnas.92.6.2244

- Blattner FR, Plunkett III G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al. 1997. The complete genome sequence of *Escherichia coli* K12. *Science* **277**: 1453–1462. doi:10.1126/science.277.5331.1453
- Bobay LM, Touchon M, Rocha EP. 2013. Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. *PLoS Genet* **9**: e1003825. doi:10.1371/journal.pgen.1003825
- Botstein D, Matz MJ. 1970. A recombination function essential to the growth of bacteriophage P22. *J Mol Biol* **54**: 417–440. doi:10.1016/0022-2836(70)90119-1
- Breitbart M, Miyake JH, Rohwer F. 2004. Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett* **236**: 249–256. doi:10.1111/j.1574-6968.2004.tb09654.x
- Brewster JL, Tolun G. 2020. Half a century of bacteriophage lambda recombinase: in vitro studies of lambda exonuclease and Red-beta annealase. *IUBMB Life* **72**: 1622–1633. doi:10.1002/iub.2343
- Briggs GS, Yu J, Mahdi AA, Lloyd RG. 2010. The RdgC protein employs a novel mechanism involving a finger domain to bind to circular DNA. *Nucleic Acids Res* **38**: 6433–6446. doi:10.1093/nar/gkq509
- Caldwell BJ, Bell CE. 2019. Structure and mechanism of the Red recombination system of bacteriophage  $\lambda$ . *Prog Biophys Mol Biol* **147**: 33–46. doi:10.1016/j.pbiomolbio.2019.03.005
- Casjens SR, Gilcrease EB. 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol Biol* **502**: 91–111. doi:10.1007/978-1-60327-565-1\_7
- Casjens S, Winn-Stapley DA, Gilcrease EB, Morona R, Kühlewein C, Chua JE, Manning PA, Inwood W, Clark AJ. 2004. The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J Mol Biol* **339**: 379–394. doi:10.1016/j.jmb.2004.03.068
- Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW. 2005. The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol* **187**: 1091–1104. doi:10.1128/JB.187.3.1091-1104.2005
- Cheng K, Wilkinson M, Chaban Y, Wigley DB. 2020. A conformational switch in response to Chi converts RecBCD from phage destruction to DNA repair. *Nat Struct Mol Biol* **27**: 71–77. doi:10.1038/s41594-019-0355-2
- d'Adda di Fagnana F, Weller GR, Doherty AJ, Jackson SP. 2003. The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. *EMBO Rep* **4**: 47–52. doi:10.1038/sj.embor.embor709
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465–W469. doi:10.1093/nar/gkn180
- Dillingham MS, Kowalczykowski SC. 2008. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* **72**: 642–671. doi:10.1128/MMBR.00020-08
- Douglas SE. 1994. DNA strider. A macintosh program for handling protein and nucleic acid sequences. *Methods Mol Biol* **25**: 181–194. doi:10.1385/0-89603-276-0:181
- Drabavicius G, Sinkunas T, Silanskas A, Gasiumas G, Venclovas C, Siksnys V. 2018. DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. *EMBO Rep* **19**: e45543. doi:10.15252/embr.201745543
- Effantin G, Figueroa-Bossi N, Schoehn G, Bossi L, Conway JF. 2010. The tripartite capsid gene of *Salmonella* phage Gifsy-2 yields a capsid assembly pathway engaging features from HK97 and lambda. *Virology* **402**: 355–365. doi:10.1016/j.virol.2010.03.041
- Ennis DG, Amundsen SK, Smith GR. 1987. Genetic functions promoting homologous recombination in *Escherichia coli*: A study of inversions in phage  $\lambda$ . *Genetics* **115**: 11–24. doi:10.1093/genetics/115.1.11
- Enquist LW, Skalka A. 1973. Replication of bacteriophage lambda DNA dependent on the function of host and viral genes. I. interaction of *red*, *gam*, and *rec*. *J Mol Biol* **75**: 185–212. doi:10.1016/0022-2836(73)90016-8
- Evans DH. 2022. Poxvirus recombination. *Pathogens* **11**: 896. doi:10.3390/pathogens11080896
- Evseev P, Lukianova A, Sykilinda N, Gorshkova A, Bondar A, Shneider M, Kabilov M, Drucker V, Miroshnikov K. 2021. *Pseudomonas* phage MD8: genetic mosaicism and challenges of taxonomic classification of lambdaoid bacteriophages. *Int J Mol Sci* **22**: 10350. doi:10.3390/ijms221910350
- Faulds D, Dower N, Stahl MM, Stahl FW. 1979. Orientation-dependent recombination hotspot activity in bacteriophage  $\lambda$ . *J Mol Biol* **131**: 681–695. doi:10.1016/0022-2836(79)90197-9
- Feiss M, Young R, Ramsey J, Adhya S, Georgopoulos C, Hendrix RW, Hatfull GF, Gilcrease EB, Casjens SR. 2022. Hybrid vigor: importance of hybrid lambda phages in early insights in molecular biology. *Microbiol Mol Biol Rev* **86**: e0012421. doi:10.1128/mmb.00124-21
- Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. *Nucleic Acids Res* **43**: W30–W38. doi:10.1093/nar/gkv397
- Glassberg J, Meyer RR, Kornberg A. 1979. Mutant single strand binding protein of *Escherichia coli*: genetic and physiological characterization. *J Bacteriol* **140**: 14–19. doi:10.1128/jb.140.1.14-19.1979
- Grose JH, Casjens SR. 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family *Enterobacteriaceae*. *Virology* **468-470**: 421–443. doi:10.1016/j.virol.2014.08.024
- Grundy FJ, Howe MM. 1984. Involvement of the invertible G segment in bacteriophage mu tail fiber biosynthesis. *Virology* **134**: 296–317. doi:10.1016/0042-6822(84)90299-X
- Handa N, Yang L, Dillingham MS, Kobayashi I, Wigley DB, Kowalczykowski SC. 2012. Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence,  $\chi$ , by RecBCD enzyme. *Proc Natl Acad Sci* **109**: 8901–8906. doi:10.1073/pnas.1206076109
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, et al. 2020. Array programming with NumPy. *Nature* **585**: 357–362. doi:10.1038/s41586-020-2649-2
- Henderson D, Weil J. 1975. Recombination-deficient deletions in bacteriophage lambda and their interaction with *Chi* mutations. *Genetics* **79**: 143–174. doi:10.1093/genetics/79.2.143
- Hollifield WC, Kaplan EN, Huang HV. 1987. Efficient RecABC-dependent, homologous recombination between coliphage lambda and plasmids requires a phage *ninR* region gene. *Mol Gen Genet* **210**: 248–255. doi:10.1007/BF00325690
- Hutinet G, Besle A, Son O, McGovern S, Guerois R, Petit MA, Ochsenbein F, Lecoate F. 2018. Sak4 of phage HK620 is a RecA remote homolog with single-strand annealing activity stimulated by its cognate SSB protein. *Front Microbiol* **9**: 743. doi:10.3389/fmicb.2018.00743
- Jobling MG, Raleigh EA, Frank DN. 2016. Complete genome sequence of *Escherichia coli* ER1821R, a laboratory K-12 derivative engineered to be deficient in all methylcytosine and methyladenine restriction systems. *Genome Announc* **4**: e00763-16. doi:10.1128/genomeA.00763-16
- Jockovich ME, Myers RS. 2001. Nuclease activity is essential for RecBCD recombination in *Escherichia coli*. *Mol Microbiol* **41**: 949–962. doi:10.1046/j.1365-2958.2001.02573.x
- Joseph JW, Kolodner R. 1983. Exonuclease VIII of *Escherichia coli*. II. mechanism of action. *J Biol Chem* **258**: 10418–10424. doi:10.1016/S0021-9258(17)44473-5
- Kala S, Cumby N, Sadowski PD, Hyder BZ, Kanelis V, Davidson AR, Maxwell KL. 2014. HNH proteins are a widespread component of phage DNA packaging machines. *Proc Natl Acad Sci* **111**: 6022–6027. doi:10.1073/pnas.1320952111
- Kalatzis PG, Rorbo NI, Castillo D, Mauritzen JJ, Jørgensen J, Kokkari C, Zhang F, Katharios P, Middelboe M. 2017. Stumbling across the same phage: comparative genomics of widespread temperate phages infecting the fish pathogen *Vibrio anguillarum*. *Viruses* **9**: 122. doi:10.3390/v9050122
- Kameyama L, Fernandez L, Calderon J, Ortiz-Rojas A, Patterson TA. 1999. Characterization of wild lambdaoid bacteriophages: detection of a wide distribution of phage immunity groups and identification of a nus-dependent, nonlambdaoid phage group. *Virology* **263**: 100–111. doi:10.1006/viro.1999.9888
- Kang JY, Olinares PD, Chen J, Campbell EA, Mustaev A, Chait BT, Gottesman ME, Darst SA. 2017. Structural basis of transcription arrest by coliphage HK022 Nun in an *Escherichia coli* RNA polymerase elongation complex. *eLife* **6**: e25478. doi:10.7554/eLife.25478
- Kirby JE, Trempey JE, Gottesman S. 1994. Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli*. *J Bacteriol* **176**: 2068–2081. doi:10.1128/jb.176.7.2068-2081.1994
- Kobayashi I. 1998. Selfishness and death: raison d'être of restriction, recombination and mitochondria. *Trends Genet* **14**: 368–374. doi:10.1016/S0168-9525(98)01532-7
- Kobayashi I, Murialdo H, Crasemann JM, Stahl MM, Stahl FW. 1982. Orientation of cohesive end site *co* determines the active orientation of  $\chi$  sequence in stimulating *recA*. *recBC*-mediated recombination in phage  $\lambda$  lytic infections. *Proc Natl Acad Sci* **79**: 5981–5985. doi:10.1073/pnas.79.19.5981
- Kropinski AM. 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J Bacteriol* **182**: 6066–6074. doi:10.1128/JB.182.21.6066-6074.2000
- Krüger A, Burgan J, Friedrich AW, Rossen JWA, Lucchesi PMA. 2018. Argo145, a Stx2a prophage of a bovine O145:H- STEC strain, is closely related to phages of virulent human strains. *Infect Genet Evol* **60**: 126–132. doi:10.1016/j.meegid.2018.02.024
- Lam ST, Stahl MM, McMilin KD, Stahl FW. 1974. Rec-mediated recombination hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics* **77**: 425–433. doi:10.1093/genetics/77.3.425

- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948. doi:10.1093/bioinformatics/btm404
- Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, Edgar R, Qimron U, Sorek R. 2015. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**: 505–510. doi:10.1038/nature14302
- Lopes A, Amarir-Bouhram J, Faure G, Petit MA, Guerois R. 2010. Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res* **38**: 3952–3962. doi:10.1093/nar/gkq096
- Lovett ST. 2011. The DNA exonucleases of *Escherichia coli*. *EcoSal Plus* **4**: 10.1128/ecosalplus.4.4.7. doi:10.1128/ecosalplus.4.4.7
- Massey FJ Jr. 1951. The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* **46**: 68–78. doi:10.1080/01621459.1951.10500769
- McKittrick NH, Smith GR. 1989. Activation of Chi recombinational hotspots by RecBCD-like enzymes from enteric bacteria. *J Mol Biol* **210**: 485–495. doi:10.1016/0022-2836(89)90125-3
- Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voicheck M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R. 2020. Bacterial retrons function in anti-phage defense. *Cell* **183**: 1551–1561.e12. doi:10.1016/j.cell.2020.09.065
- Mmolawa PT, Schmieger H, Heuzenroeder MW. 2003. Bacteriophage ST64B, a genetic mosaic of genes from diverse sources isolated from *Salmonella enterica* serovar typhimurium DT 64. *J Bacteriol* **185**: 6481–6485. doi:10.1128/JB.185.21.6481-6485.2003
- Moser MJ, Holley WR, Chatterjee A, Mian IS. 1997. The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic Acids Res* **25**: 5110–5118. doi:10.1093/nar/25.24.5110
- Murphy KC, Lewis LJ. 1993. Properties of *Escherichia coli* expressing bacteriophage P22 *abc* (anti-RecBCD) proteins, including inhibition of Chi activity. *J Bacteriol* **175**: 1756–1766. doi:10.1128/jb.175.6.1756-1766.1993
- Murphy KC, Fenton AC, Poteete AR. 1987. Sequence of the bacteriophage P22 anti-RecBCD (*abc*) genes and properties of P22 *abc* region deletion mutants. *Virology* **160**: 456–464. doi:10.1016/0042-6822(87)90017-1
- Myers RS, Kuzminov A, Stahl FW. 1995. The recombination hot spot  $\chi$  activates RecBCD recombination by converting *Escherichia coli* to a *recD* mutant phenocopy. *Proc Natl Acad Sci* **92**: 6244–6248. doi:10.1073/pnas.92.14.6244
- Packard JE, Dembowski JA. 2021. HSV-1 DNA replication—coordinated regulation by viral and cellular factors. *Viruses* **13**: 2015. doi:10.3390/v13102015
- Parma DH, Snyder M, Sobolevski S, Nawroz M, Brody E, Gold L. 1992. The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes Dev* **6**: 497–510. doi:10.1101/gad.6.3.497
- Pavankumar TL, Sinha AK, Ray MK. 2010. All three subunits of RecBCD enzyme are essential for DNA repair and low-temperature growth in the Antarctic *Pseudomonas syringae* Lz4W. *PLoS One* **5**: e9412. doi:10.1371/journal.pone.0009412
- Pavankumar TL, Sinha AK, Ray MK. 2018. Biochemical characterization of RecBCD enzyme from an Antarctic *Pseudomonas* species and identification of its cognate Chi ( $\chi$ ) sequence. *PLoS One* **13**: e0197476. doi:10.1371/journal.pone.0197476
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612. doi:10.1002/jcc.20084
- Petty NK, Feltwell T, Pickard D, Clare S, Toribio AL, Fookes M, Roberts K, Monson R, Nair S, Kingsley RA, et al. 2011. *Citrobacter rodentium* is an unstable pathogen showing evidence of significant genomic flux. *PLoS Pathog* **7**: e1002018. doi:10.1371/journal.ppat.1002018
- Poteete AR, Fenton AC, Semerjian AV. 1991. Bacteriophage P22 accessory recombination function. *Virology* **182**: 316–323. doi:10.1016/0042-6822(91)90675-2
- Rotman E, Kouzminova E, Plunkett G III, Kuzminov A. 2012. Genome of enterobacteriophage Lulu/phi80 and insights into its ability to spread in the laboratory environment. *J Bacteriol* **194**: 6802–6817. doi:10.1128/JB.01353-12
- Sandmeier H, Iida S, Arber W. 1992. DNA inversion regions Min of plasmid p15B and Cin of bacteriophage P1: evolution of bacteriophage tail fiber genes. *J Bacteriol* **174**: 3936–3944. doi:10.1128/jb.174.12.3936-3944.1992
- Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB. 1982. Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol* **162**: 729–773. doi:10.1016/0022-2836(82)90546-0
- Schultz DW, Smith GR. 1986. Conservation of Chi cutting activity in terrestrial and marine enteric bacteria. *J Mol Biol* **189**: 585–595. doi:10.1016/0022-2836(86)90489-4
- Smith GR. 1983. General recombination. In *Lambda II* (ed. Hendrix RW, et al.), pp. 175–209. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Smith GR. 2012. How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view. *Microbiol Mol Biol Rev* **76**: 217–228. doi:10.1128/MMBR.05026-11
- Smith GR, Kunes SM, Schultz DW, Taylor A, Triman KL. 1981. Structure of Chi hotspots of generalized recombination. *Cell* **24**: 429–436. doi:10.1016/0092-8674(81)90333-0
- Smith GR, Roberts CM, Schultz DW. 1986. Activity of Chi recombinational hotspots in *Salmonella typhimurium*. *Genetics* **112**: 429–439. doi:10.1093/genetics/112.3.429
- Stahl FW. 2005. Chi: a little sequence controls a big enzyme. *Genetics* **170**: 487–493. doi:10.1093/genetics/170.2.487
- Stemberg N, Hoess R. 1983. The molecular genetics of bacteriophage P1. *Annu Rev Genet* **17**: 123–154. doi:10.1146/annurev.ge.17.120183.001011
- Subramaniam S, Smith GR. 2022. RecBCD enzyme and Chi recombination hotspots as determinants of self vs. non-self: myths and mechanisms. *Adv Genet* **109**: 1–37. doi:10.1016/bs.adgen.2022.06.001
- Taylor AF, Schultz DW, Ponticelli AS, Smith GR. 1985. RecBC enzyme nicking at Chi sites during DNA unwinding: location and orientation dependence of the cutting. *Cell* **41**: 153–163. doi:10.1016/0092-8674(85)90070-4
- Thomason LC, Schiltz CJ, Court C, Hosford CJ, Adams MC, Chappie JS, Court DL. 2021. Bacteriophage lambda RexA and RexB functions assist the transition from lysogeny to lytic growth. *Mol Microbiol* **116**: 1044–1063. doi:10.1111/mmi.14792
- Trempey JE, Kirby JE, Gottesman S. 1994. Alp suppression of Lon: dependence on the *slpA* gene. *J Bacteriol* **176**: 2061–2067. doi:10.1128/jb.176.7.2061-2067.1994
- Vahanian N, Oh CS, Sippy J, Feiss M. 2017. Natural history of a viral cohesive end site: *cosN* of the  $\lambda$ -like phages. *Virology* **509**: 140–145. doi:10.1016/j.virol.2017.06.016
- Valencia-Toxqui G, Ramsey J. 2024. How to introduce a new bacteriophage on the block: a short guide to phage classification. *J Virol* **98**: e0182123. doi:10.1128/jvi.01821-23
- Viswanathan M, Lovett ST. 1999. Exonuclease X of *Escherichia coli*. A novel 3'-5' DNase and DnaQ superfamily member involved in DNA repair. *J Biol Chem* **274**: 30094–30100. doi:10.1074/jbc.274.42.30094
- Wang X, Kim Y, Wood TK. 2009. Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *ISME J* **3**: 1164–1179. doi:10.1038/ismej.2009.59
- Weerasooriya S, DiScipio KA, Darwish AS, Bai P, Weller SK. 2019. Herpes simplex virus 1 ICP8 mutant lacking annealing activity is deficient for viral DNA replication. *Proc Natl Acad Sci* **116**: 1033–1042. doi:10.1073/pnas.1817642116
- Weisberg RA, Gottesmann ME, Hendrix RW, Little JW. 1999. Family values in the age of genomics: comparative analyses of temperate bacteriophage HK022. *Annu Rev Genet* **33**: 565–602. doi:10.1146/annurev.genet.33.1.565
- Weller SK, Sawitzke JA. 2014. Recombination promoted by DNA viruses: phage lambda to herpes simplex virus. *Annu Rev Microbiol* **68**: 237–258. doi:10.1146/annurev-micro-091313-103424
- Wigley DB. 2013. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat Rev Microbiol* **11**: 9–13. doi:10.1038/nrmicro2917
- Wilkinson M, Wilkinson OJ, Feyerherm C, Fletcher EE, Wigley DB, Dillingham MS. 2022. Structures of RecBCD in complex with phage-encoded inhibitor proteins reveal distinctive strategies for evasion of a bacterial immunity hub. *eLife* **11**: e83409. doi:10.7554/eLife.83409
- Wishart DS, Han S, Saha S, Oler E, Peters H, Grant JR, Stothard P, Gautam V. 2023. PHATEST: faster than PHASTER, better than PHAST. *Nucleic Acids Res* **51**: W443–W450. doi:10.1093/nar/gkad382
- Wu CC, Lin JIJ, Yuan HS. 2020. Structures, mechanisms, and functions of His-Me finger nucleases. *Trends Biochem Sci* **45**: 935–946. doi:10.1016/j.tibs.2020.07.002
- Xu SY, Gupta YK. 2013. Natural zinc ribbon HNH endonucleases and engineered zinc finger nicking endonuclease. *Nucleic Acids Res* **41**: 378–390. doi:10.1093/nar/gks1043
- Zakharova K, Liu M, Greenwald JR, Caldwell BC, Qi Z, Wyszocki VH, Bell CE. 2024. Structural basis for the interaction of Red $\beta$  single-strand annealing protein with *Escherichia coli* single-stranded DNA-binding protein. *J Mol Biol* **436**: 168590. doi:10.1016/j.jmb.2024.168590
- Zheng C, Furukawa C, Liu J, Sankaran S, Lin H, Munugeti N, Wang M, Smith GR. 2024. Debunking the dogma that RecBCD nuclease destroys phage. *Genetics* **230**: iyae199. doi:10.1093/genetics/iyae199

Received November 25, 2024; accepted in revised form May 21, 2025.