





Genome archaeology of two laboratory *Salmonella enterica enterica* sv Typhimurium

Julie Zaworski ,¹ Oyut Dagva ,¹ Anthony W. Kingston,¹ Alexey Fomenkov ,¹ Richard D. Morgan,¹ Lionello Bossi,² and Elisabeth A. Raleigh ^{1,*}

¹Research Department, New England Biolabs, Ipswich, MA 01938-2723, USA

²CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), University Paris-Saclay, Gif-sur-Yvette 91198, France

*Corresponding author: Research Department, New England Biolabs, 240 County Road, Ipswich, MA 01935-2723, USA. Email: raleigh@neb.com

Abstract

The *Salmonella* research community has used strains and bacteriophages over decades, exchanging useful new isolates among laboratories for the study of cell surface antigens, metabolic pathways and restriction-modification (RM) studies. Here we present the sequences of two laboratory *Salmonella* strains (STK005, an isolate of LB5000; and its descendant ER3625). In the ancestry of LB5000, segments of ~15 and ~42 kb were introduced from *Salmonella enterica* sv Abony 803 into *S. enterica* sv Typhimurium LT2, forming strain SD14; this strain is thus a hybrid of *S. enterica* isolates. Strains in the SD14 lineage were used to define flagellar antigens from the 1950s to the 1970s, and to define three RM systems from the 1960s to the 1980s. LB5000 was also used as a host in phage typing systems used by epidemiologists. In the age of cheaper and easier sequencing, this resource will provide access to the sequence that underlies the extensive literature.

Keywords: *Salmonella* LB5000; ER3625; prophages; hybrid genome; transduction scars; H antigen; IS200 mobility; Gifsy prophages; 6-methyladenine; SenLT2I (StyLT); SenLT2II (StySB); SenLT2III (StySA)

Introduction

Bacteria, phages, and conjugal plasmids are in a constant arms race to develop respectively defense against DNA invasion and methods of evading defenses. Bacteria express variable surface features to thwart entry (e.g., flagellae, outer membrane proteins, and O-antigen sugar chains) and internal limitation mechanisms to abort establishment (e.g., restriction systems and abortive-infection suicide factors) (Doron et al. 2018; Azam and Tanji 2019). Phages and plasmids may defeat host restriction-modification (RM) systems with anti-restriction proteins, which frequently block type I DNA RM enzymes (e.g., Zavilgelsky and Rastorguev 2009; Serfotis-Mitsa et al. 2010; Balabanov et al. 2012; Roberts et al. 2012; Piya et al. 2017; González-Montes et al. 2020). Low-specificity DNA modification enzymes expressed only for the entering molecule are also found in prophages (Murray et al. 2018) and plasmids (Fomenkov et al. 2020).

Classical studies of horizontal exchange processes within and between genera frequently employed strains in the ancestry of LB5000, particularly for study of cell surface antigens and restriction systems. Early hybrid serovar ancestors were constructed by transduction from *S. enterica* sv Abony into *S. enterica* sv Typhimurium. These studies defined flagellar antigen types, the phenomenon of flagellar phase variation (Joys and Stocker 1965; Joys and Stocker 1969), and the use of O-antigen as phage receptor (Ornellas and Stocker 1974; Stocker et al. 1980). Three

restriction systems were identified using this hybrid lineage (Bullas and Colson 1975). In LB5000 itself, the three restriction systems of strain LT2 were inactivated (Bullas and Ryu 1983), allowing use for epidemiologic purposes as a host for phages in the Anderson phage typing system (Schmieger 1999). Intergeneric conjugal transfer of chromosomal markers was studied using *Escherichia coli* K-12 derivatives and *Salmonella enterica* sv Typhimurium lysotypes LT2 and LT7, using F' plasmids and integrated F factors (Sanderson et al. 1987). LB5000 and relatives are particularly useful as an intermediate host in molecular genetic applications, moving constructions between strains and genera (Sanderson and Roth 1988; Segall and Roth 1989; Chaudhuri et al. 2009; Lerche et al. 2017; Yoon et al. 2018; Griewisch et al. 2020; Jones-Carson et al. 2020).

Genome sequences are available for LT2 (McClelland et al. 2001) and LT7 (Zaworski et al. 2021), but none within this hybrid lineage.

Here, we present annotated sequences of two *S. enterica* sv Typhimurium variant genomes: STK005 (our isolate of LB5000, from Anca Segall's G251), and ER3625, a derivative of STK005 built in our lab for studies of HGT events. The LB5000 genome carries scars from genetic steps in its descent from LT2, including the replacement of two flagellum-encoding segments of the LT2 genome by segments of *S. enterica* sv Abony, with concomitant loss of prophage Fels2. Unusually, new insertions of the indolent mobile element IS200 were found. Recombination events

Received: April 06, 2021. Accepted: June 29, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

involving two prophages document intralinear allele shuffling. Mutations linked to the amino acid requirement of the strain were identified as well.

Materials and methods

Salmonella growth

For each strain, a single colony was grown in RB (10 g soy peptone, 5 g yeast extract, and 5 g NaCl per liter) with the appropriate antibiotics (kanamycin 40 µg/ml, streptomycin 100 µg/ml) overnight at 37°C with 250 rpm agitation.

DNA preparation, sequencing, and *de novo* assembly

DNA extraction

STK005 and ER3625 were prepared for PacBio RSII or Sequel sequencing. gDNA was extracted with the Monarch Genomic DNA purification kit (#T3010 New England Biolabs; Ipswich, MA, USA). For Oxford Nanopore sequencing of STK005, gDNA was extracted from 1 ml of overnight culture with the kit Monarch HMW DNA Extraction kit (T#3060 New England Biolabs; Ipswich, MA, USA) following the manufacturer instructions for High Molecular Weight DNA extraction from gram-negative bacteria. After the extraction, the HMW DNA was homogenized following by heat and pipetting before QC. A clean peak >60 kb was found on analysis with the Genomic DNA ScreenTape® (Agilent Technology, Santa Clara, CA).

Library preparation

For the PacBio RSII sequencing platform, STK005 SMRTbell libraries were constructed from genomic DNA samples sheared to between 10 and 20 kb using the G-tubes protocol (Covaris; Woburn, MA, USA), end repaired, and ligated to PacBio hairpin adapters. Incompletely formed SMRTbell templates and linear DNAs were digested with a combination of Exonuclease III and Exonuclease VII (New England Biolabs; Ipswich, MA, USA). DNA qualification and quantification were performed using the Qubit fluorimeter (Invitrogen, Eugene, OR) and 2100 Bioanalyzer (Agilent Technology, Santa Clara, CA, USA). The SMRTbell library was prepared according to PacBio sample preparation protocol sequenced with C4-P6 chemistry with a 300 min collection time. A total of 751,460 polymerase reads with N50 length of 17,626 bp was obtained (approximately 360X coverage) before filtering.

ER3625 libraries were sequenced using the PacBio Sequel I platform. Briefly, SMRTbell libraries were constructed from genomic DNA samples following the PacBio protocol for Sequel using the kit 100-938-900. DNA qualification and quantification were performed using the Qubit fluorimeter (Invitrogen, Eugene, OR) and 2100 Bioanalyzer (Agilent Technology, Santa Clara, CA, USA). The resulting library, of average size 11,640 bp, was prepared for binding following the PacBio guidelines generated by SMRT Link and run on a Sequel I machine.

The STK005 Nanopore library was prepared with 1.7 µg of HMW DNA following the Genomic DNA by ligation (SQK-LSK109 protocol) from Oxford Nanopore. The adapter ligation was performed for 1 hour instead of 10 minutes. The flow cell was primed and loaded with ~100 ng of library following the same documentation. After a 21 hours run on a second-use FLO-MIN106 flow cell with the GridIONxs apparatus, 449K reads were obtained, with 27.39 kb read length basecalled N50. The basecalling was performed on the fly with the high-accuracy basecalling model from Guppy (version 3.2.10).

Sequence assembly

The STK005 (PacBio RSII reads) were assembled using RS_HGAP3 program with parameters: minimum subread length 5000 bp, minimum polymerase quality 0.8, Minimum polymerase read length 10,000 bp, 5 Mbp genome size, and other parameters as default settings. The three resulting contigs, of size 4,802,499, 103,465, and 8138 bp, were circularized with Circlator (Hunt et al. 2015) with CCS generated with parameters minimum full passes 3, minimum predicted accuracy 95, minimum read length of insert 1000 bp, and the other parameters as default. The smallest contig was a subset of the chromosome contig.

PacBio polishing for STK005 was done by running the RS_Reseq program twice, with Minimum subread length 500, minimum polymerase read quality 75 and then 80, and 1500 bp for Minimum polymerase read length. The final assembly is composed of a 4,796,208 bp chromosome and a 93,938 bp plasmid.

The STK005 Nanopore reads were corrected and filtered using Canu (Koren et al. 2017), the output corrected reads were then trimmed and assembled with Canu (version 1.4). A first correction step was run with parameters minReadLength = 10000, stopOnReadQuality = false and other as default. Then trimming and assembly were run with genomeSize = 5 m and other parameters as default. Two contigs of length 162,546 and 4,863,042 bp corresponding to the plasmid and chromosome.

In parallel, the reads were mapped on the PacBio *de novo* assembly using minimap2 (version 2.17) with default parameters (Li 2018). In both cases, no inversion in between the Gifsy prophages was observed.

PacBio Sequel assembly of ER3625 used the RS_HGAP4 program with parameters: 5 Mb expected genome size, minimum mapped length 5 kb, minimum mapped concordance 70, and other parameters as default settings. The 2 contigs obtained were circularized with Circlator (version 1.5.5). Polishing was done by running the RS_Reseq program with minimum mapped length 500, yielded 19 corrections. The final assembly is composed of a 4,796,355 bp chromosome (coverage 541x) and a 93,939 bp plasmid (2247x coverage).

Modification motifs

DNA motifs and degree of modification were generated using InterPulse Duration (IPD) Ratios analyzed with RS_Modification_and_Motif_Analysis from PacBio as in (Flusberg et al. 2010; Clark et al. 2012).

Genome annotation

Genbank deposits STK005 (chromosome CP067397, plasmid CP067398) and ER3625 (chromosome CP067091 and plasmid CP067092) genomes were annotated with the Genbank PGAP pipeline (Tatusova et al. 2016; Haft et al. 2018). In-process analysis employed annotations in Geneious (2018 version in steps: Menu Align and Assemble; Align Whole Genomes; Tab Alignment View; Save as editable copy; Menu Annotate and Predict; Annotate from Database) using LT2 [NC_003197.2] and Abony [NZ_CP007534.1] and segment extractions from them as reference genomes in a local database folder.

Results and discussion

Strain engineering scars: *Salmonella enterica* subsp. *enterica* serovar Abony DNA in LB5000

The engineering history of LB5000 (Figure 1) includes two transductions from *Salmonella* serovar Abony strain 803 (Spicer and

Datta 1959) (see details and references in Supplementary Table S1). These two transductions left scars in the form of high-divergence segments in the genome surrounding the flagellar loci H1 and H2 (Figure 2A). In *S. enterica*, both H1 and H2 flagellar antigens are highly variable in wild lineages in the population; in addition, each individual lineage is capable of switching from expressing its H1 antigen to its H2 or the reverse, a phenomenon known as phase variation (Simon et al. 1980). This case of phase variation involves reversible site-specific rearrangement of regulatory signals that enable expression from only one of two distant loci. In LB5000, both of these loci derive from *S. enterica* sv Abony.

H1, phase 1 flagellin region transduction patch

The H1 phenotype depends on the allelic state of the gene *fliC*, within the H1 box of Figure 2A. This patch is the result of a first transduction mediated by P22: recipient strain LT2 SL270 replaced the H1-i 1.2 allele of LT2 with H1-b from *S. Abony* SW803 (Spicer and Datta 1959; Joys and Stocker 1969) (see Supplementary Table S1). Allele H1-b was bequeathed to LB5000 ancestor SD7. The highly divergent interval [LT2 locus_ID STM1946 (*uvrC*) to STM1961 (*fliD*)] includes two unshared IS insertions (IS3 in STK005, IS200 in LT2) and encompasses the *fliC* flagellin gene. In contrast, this region in LB5000 (JJB81_10090 to JJB81_10010; nt 2037077 -> 2052486) aligns with only three mismatches to *S. Abony* (SEEA0014_RS19195 to SEEA0014_RS19270; nt 3941779 -> 3957191); see below and Supplementary File S1.

H2, phase 2 flagellin region transduction patch

The H2 phenotype is determined by the allelic state of the gene *fliB*, a *fliC* homolog. When the adjacent regulatory gene *fliA* is expressed, it mediates repression of distant H1 expression; when the local DNA segment is inverted, *fliA* and *fliB* are not expressed, and repression of distant gene *fliC* is lifted. Wild type for this LT2 isolate is H2-1.2; *S. Abony* SW803 was H2-enx. H2-enx was bequeathed to LB5000 ancestor SD14.

These genetic changes are reflected in a second divergent region of the alignment (H2 black rectangle in Figure 2A; closer view in Figure 2B). Sequences present in LT2 but absent in STK005 include the Fels2 prophage (STM2694 to STM2740) and the flanking genes for STM2741 to STM2769. Present but divergent are locus_ID JJB81_13790 to JJB81_13790 when aligned with LT2 locus_ID STM2679 and STM3779. We infer that this divergent region represents genomic replacement of 106 kb of *S. sv Typhimurium* LT2 by 41 kb of *S. sv Abony* 803 during the second transduction (Figure 1A). Confirmation of the inference is illustrated in Figure 2C: 41 kb of our sequence is very similar (3 insertions, 14 mismatches) to a 41 kb segment of NZ_CP007534 [sequence of *S. enterica* sv Abony strain 0014 (NCTC 6017; ATCC BAA-2162) (Timme et al. 2013); see Supplementary File S1 sheet 4]. Both flanks of the long deletion show high divergence from LT2, and the *fliB* region is 100% identical to the sequenced *S. sv Abony* genome. Prophage Fels-2 is not present in Abony serovar isolates, consistent with the present configuration.

hin gene, STM2772, JJB81_13890

The Hin protein mediates inversion of a segment including its own gene and a promoter of *fliC* (H2, phase 2) and the *fliA* [repressor of *fliC* (H1, phase 1)] (Simon et al. 1980). The STK005 *hin* homolog (JJB81_13890) is within the H2 transduction patch. It is similar to the Abony SEEA0014-RS00045 gene except for the 3 single nucleotide insertions; these insertions are all shared among LT2, STK005, and ER3625. In addition to the closer identity to *S. enterica* sv Abony, the *hin* gene has suffered an inversion during the

descent of STK005 from LT2 (also found in ER3625). Considering the expected protein activity, this reorganization is evidence of expected function.

Mobile element movement

Surprisingly, three new insertions of IS200 were found in STK005. The extreme stability of insertion locations in *S. enterica enterica* serovar Typhimurium had led to the proposal that selection operated on the six insertion loci to enable formation of segmental chromosomal duplications (Haack and Roth 1995). Recent work has shown that lack of mobility in the natural state of IS200 results from post-transcriptional repression of TnpA translation (Ellis et al. 2015). In fact, the *tnpA* transcript is responsible for repression of numerous chromosomal genes as well (Ellis et al. 2017). Here, of the six IS200 LT2 copies, five were inherited by STK005 and one was lost by transductional replacement (in H1, see Figure 2A and Supplementary File S3 part III). STK005 gained three: one is a result of tandem duplication of the ancestral copy of *tnpA3* (middle blue star in Figure 2A), and two are at new sites (left and right blue stars in Figure 2A). Possibly, translation and action of TnpA was fostered by stressful steps in the pedigree: multiple heavy mutagenic treatments (both early and late steps) and F' transfer [late in the pedigree; see Supplementary Table S1 and e.g., Vandecraen et al. (2017)].

ER3625 engineering and variation from STK005

ER3625 was engineered in our laboratory to create a non-pathogenic recipient (Hoiseth and Stocker 1981) with no Mrr restriction, to pursue work on R-M systems in HGT using the system previously used in our laboratory (Kingston et al. 2015, 2017). Strain engineering required several steps with a variety of methods (Figure 1B).

For future work, we wish to verify that no off-target changes had occurred during λ -Red-mediated genome engineering or accompanying transduction. The *mrr* engineered locus carries the designed sequence (*npt* flanked by FRT sites replacing the gene). The *aroA* gene was deleted, but the scar of the FRT site recombination is aberrant. This DNA segment is annotated by NCBI pipeline as JJB80_05020 (DUF1317 domain-containing protein) and JJB80_05025 (DUF1382 protein family). No SNPs were identified in genes between STK005 and ER3625 for either the chromosome or the plasmid.

Structural complexity and intralinearage prophage recombination

Gifsy-1 and Gifsy-2 are two LT2 prophages located between 2,728,977 and 2,776,819 (STM2584 to STM2636) and between 1,098,231 and 1,143,702 (STM1005 to STM1056), respectively. These two related prophages are patchily similar to each other; while overall DNA identity is 44.1%, this ~42 kb region is interspersed with highly divergent segments and unrelated insertions (Figure 3). Of concern are two long regions of high identity: a ~12 kb stretch of >98% identity in the regions expressed early in lytic infection: regulators related to CI, CII, and Cro of λ , and replication; and an 8 kb patch of ~78% identity at the other end of prophages.

These long regions with very high identity are of concern for both technical and biological reasons. First, long repeats are an assembly challenge even with long-read technologies. Second, recombination within the host may generate new gene combinations that are then transmissible to new hosts on prophage induction. Extensive mosaicism among tailed phages is well known (Grose and Casjens 2014); and creation of a new mosaic

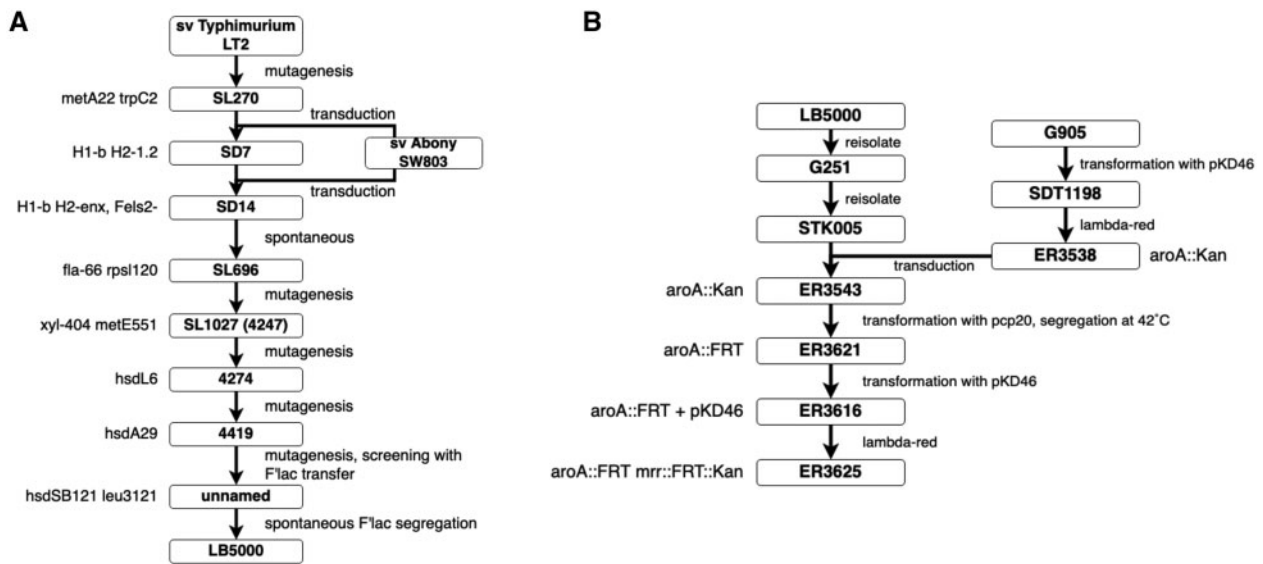


Figure 1 Strain pedigree charts. (A) LB5000 (B) ER3625. Strain names are boxed. To the right of each construction step (arrows) are the methods used; to the left of each box is the genotype designation assigned as a result of the step. In some cases, a genotype marker was discovered during later investigation. The different strain modifications were done by mutagenesis (random by use of NG), transduction (use of phages), or transformation (plasmids with induction). For more details and references on these steps, refer to Supplementary Table S1.

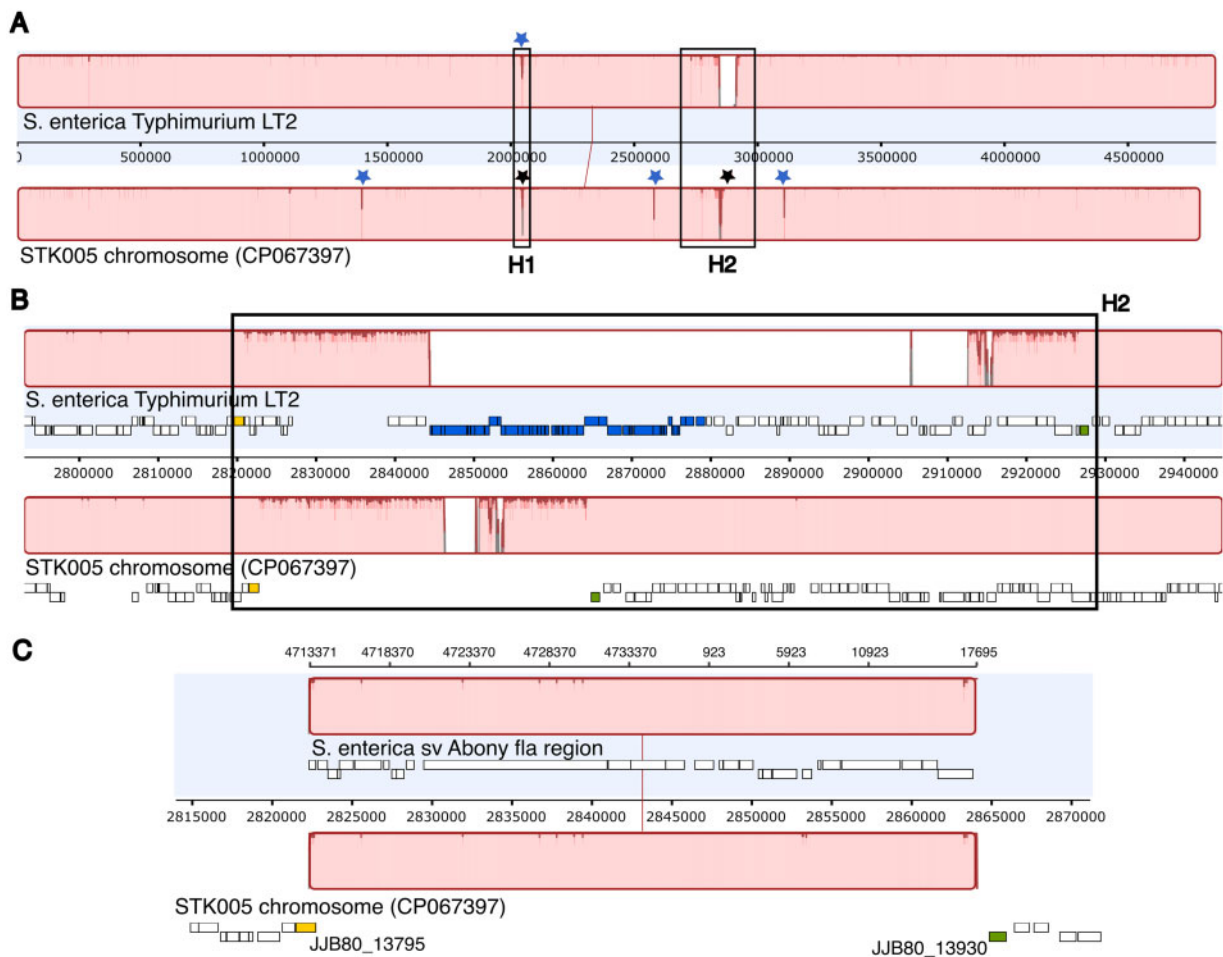


Figure 2 Recombination scars. Panel A is a Mauve alignment of the STK005 and LT2 chromosomes. Stars represent IS200 (blue) and IS3 (black) sequences not shared between the sequences. The flagellar regions H1 and H2 are boxed. Light pink color is for high similarity areas; the white area is a missing segment; and dark red lines dip where mismatches occur. Panel B is a close-up of the H2 region. Panel C is a Mauve alignment of *S. Abony* H2 (*fla*) region with STK005. In panel B and C, the yellow and green rectangles below the alignment are landmarks, respectively gene STM2679 (JJB80-13795) and STM2780 (JJB80-13930).

by recombinational marker rescue from defective prophages has been noted before (De Paepe et al. 2014). Evidence of recombination within a laboratory lineage was found in *E. coli* ER2796 (Anton et al. 2015).

Resolution of assembly errors

De novo assembly of the STK005 genome initially yielded (artificial) chromosome inversion. On examination of reads aligned with the two Gifsy regions, many poorly or partially mapped reads and an important number of SNPs were seen, suggesting imperfect mapping. A reassembly using only long CCS (>5 kb; 47.7x coverage) was then created with HGAP3, then circularized with Circlator. This assembly did not show a genomic inversion relative to LT2. Further polishing to resolve SNPs used RS_reseq with CCS > 500 bp (220.9x coverage). This time, few reads were mismatched at the Gifsy regions, confirming the absence of the inversion in STK005 genome.

In parallel, Oxford Nanopore was used to generate long reads that could cover 40–45 kb regions and reinforce confidence in the structural organization of the assembly. Reads >10 kb were used to generate a *de novo* assembly with Canu (version 1.4). This assembly agreed with the final PacBio assembly, i.e., without inversion of the genome. Mapping of the Nanopore reads on the PacBio assembly confirmed the structure.

Prophage recombination

This 12 kb region of near identity codes for proteins involved in homologous recombination, as well as regulation and replication. The recombination proteins are related to those of the *E. coli* Rac prophage RecET recombination cluster; also present are homologs of λ Orf and Rap, which also affect recombination. RecE and RecT provide homologous recombination systems that can substitute for the host RecABCD system (Lemire et al. 2008; Murphy 2016). Rap (originally named NinG) facilitates recombination near linear ends via the λ Red pathway (Tarkowski et al. 2002), acting as a structure-specific nuclease (Sharples et al. 2004).

Three segments within this long region posed interpretation problems: assembly error or recombination during the lineage leading to STK005. Our interpretation is that the reference LT2 sequence is in error in one case, and both reciprocal recombination and gene conversion have affected this region (see Figure 3 and Supplementary Files S2 and S4).

First is the appearance of a novel allele of Gifsy-1 *recET*, including a reading-frame-preserving insertion of 42 nt (14 aa) in the middle of the RecE CDS and 9 SNPs flanking it (Supplementary File S2 Embedded fig2). The Gifsy-1 allele of *recET* found here appears to be ancestral: it is also found in the Gifsy-1 sequence of

S. enterica sv Typhimurium strain 14028S (Genbank CP034479.1; Supplementary File S2 Embedded fig2). (Figueroa-Bossi et al. 2001; Lemire et al. 2007). Strain 14028S was used extensively in laboratory studies due to much more virulent properties in mice (e.g., Figueroa-Bossi et al. 1997). PCR tests of 19 LT2 isolates or derivatives confirm the presence of the insertion in all of them except strains specifically cured of Gifsy-1 (Supplementary File S4 Embedded fig1 and Embedded table1). A recent, independently determined LT2 sequence (CP014051) was obtained using PacBio and Illumina technology. This deposit agrees with our sequence of STK005 across this region.

Second, there has been a reciprocal exchange of the sequence between Gifsy-1 and Gifsy-2 prophages (Supplementary File S2 Embedded fig3A, 3B). This exchange includes the position of the 18-amino acid (51 bp) indel variation in CDSs encoding a DNA break and rejoin activity illustrated in Figure 3B above.

Third, gene conversion has resulted in loss of the original Gifsy-2 allele of *rap* (*ninG*), replaced by Gifsy-1 information without change in the original Gifsy-1 (Supplementary File S2 Embedded fig4A, 4B).

Finding evidence of recombination activity within the high-identity region is interesting. Though normally silent as prophages, both Gifsy-borne LT2 homologs of RecET can be activated by mutation to mediate recombination in vivo (Lemire et al. 2008). Action of these phage recombination systems may promote the mosaic genomic relationships among phages that has been found (Martinsohn et al. 2008). The two events found here are compatible with that role.

Recorded genotype and point mutations

Metabolic markers

LB5000 was originally described as auxotrophic for methionine, tryptophan, and leucine with the genotype *metA22 metE551 trpC2 ilv-452*. We identified mutations likely responsible for the reported phenotypes; we also found mutations in other genes involved in amino acid synthesis, see Supplementary File S3.

Concerning methionine production, changes are found in *metA* (*metA22*) and *metE* (*metE551*). In *metA* (STM4182 in LT2 and JJB80_21020 in STK005), a 10 nt deletion at the protein C-terminus truncates it by 9 amino acids. A SNP in *metE* (G877A) leads to G293R protein substitution. The *trpC2* allele carries nt change G545A leading to G->D amino acid substitution. For isoleucine and leucine auxotrophies, SNPs were found in several genes. Coding changes were a G>A mutation at position 418 in *ilvC* (G140S amino acid change) and a C to T transition in *leuA* at position 424 (D142N).

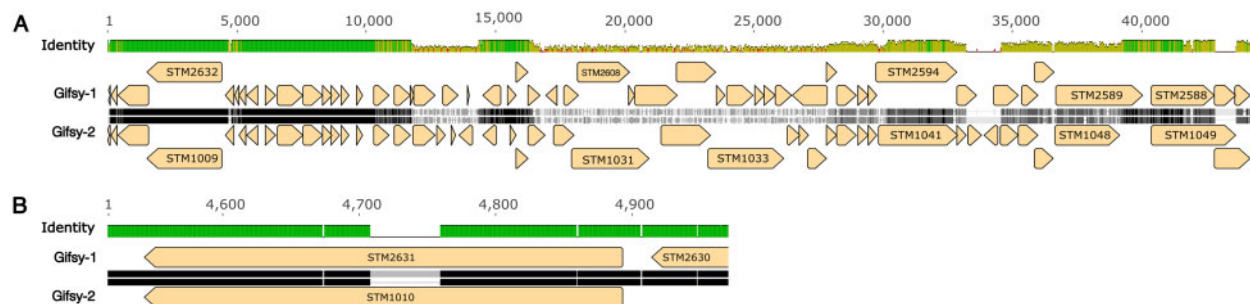


Figure 3 Pairwise alignment of the Gifsy-1 and Gifsy-2 prophages of LT2. Panel A: alignment of the whole prophages. Panel B: close-up of STM2631 and STM1010, revealing a 51 bp indel variation. The identity rows: green = identical bases; yellow = lower identity; red = very low local identity. The sequences Gifsy-1 and Gifsy-2 are in black, white, and gray: black for identical bases, white for SNPs, gray for insertion/deletion (indel) variation (corresponding gaps shown with a thin line).

Considering other amino acid pathways, *proB* and *proC* of proline synthesis have mutations respectively C > T at position 535 (P179S) and C > T at position 317 (A106V). However, our auxotrophy tests found that the strain requires methionine, valine/isoleucine and tryptophan/phenylalanine/tyrosine but can grow without addition of proline in the media (data not shown). The synthesis of proline may be mediated via *putA*, which has two silent mutations.

Genotype and phenotype of methyltransferases and their associates

DNA methylation motifs and level of methylation mediated by R-M systems and orphan MTases were analyzed with the PacBio RSII. At least six methylation motifs are known in ancestral strain *S. enterica* serovar Typhimurium LT2 (Roberts et al. 2015; REBASE Organism number 18099). Three motifs result from enterobacterial “orphan” methyltransferases [M; M.SenLT2 Dam, M.SenLT2 Dcm, and M.SenLT2IV; (Roberts et al. 2015)], and three are associated with RM phenomena (M.SenLT2I, M.SenLT2II, and M.SenLT2III). Two of these RM systems have been well characterized. Characterization began with a transfer isolate SD14 (Stocker lab strain SL1027). M.SenLT2I (known in the literature as the StyLT or StyLTI RM system, Colson and Colson 1971) is a Type III enzyme, and M.SenLT2II (StySB or StyLTII, Fuller-Pace et al. 1984) is Type I. The third system, RM.SenLT2III (StySA), is known to confer m6A modification (Hattman et al. 1976) but its protection/restriction mechanism of action remains unclear.

The LB5000 RM genotype is listed as *hsdL6 hsdSA29 hsdSB121* for StyLT, StySA, and StySB. We identified appropriate mutations consistent with the R^{M+} phenotype (Table 1 for M activity; Supplementary File S3 for genotype discussion). For StyLT and StySB, sense mutations leading to amino acid changes were identified in the endonuclease genes, as well as silent mutations. Two G->A changes (position 393 and 2749 of STM0358) are present in the *res* gene of StyLT, the second one leading to a D->N amino acid change in the C-terminal domain. For StySB, two G->A changes (position 2769 and 2473 of STM4526) are present in *hsdR*, one resulting in an R->C amino acid change in the C-terminal domain.

The LB5000 StySA R^{M+} phenotype was reported to be unstable (Tsai et al. 1989), and we found our STK005 isolate to be phenotypically StySA R^{M-}. The genetic position of StySA was mapped close to and counterclockwise of the *hsdSB* locus (Colson and Van Pel 1974); we identified a methyltransferase homolog (JJB81_22590) in an appropriate location, but found no coding change relative to STM4495. The genetic requirements for action of this system are under study.

In Table 1, we also find partial methylation for the conserved orphan methyltransferase M.SenLT2IV, as previously observed in three *Salmonella* species (Pirone-Davies et al. 2015). Methylation of m5C by M.Dcm was not determined here, due to technical limitations of Pacific Bioscience long-read sequencing in detection of m5C (Flusberg et al. 2010).

Data availability

Supplementary material is available at figshare: <https://doi.org/10.25387/g3.14815920>. Supplementary Table S1 “Pedigree LB5000 STK005” traces the descent of LB5000 and some relatives through the literature since 1959; with wild ancestor, donor strain, recipient strain, renaming, construction reference, and mode of construction. Supplementary File S1 “STK005 variants” contains 5 Excel spreadsheets documenting positions at which STK005 sequence differs from the genome sequences representing two ancestral components. For three genome segments derived from LT2, variant nucleotide positions of LT2 (NC_003197.2) are listed, with consequences for amino acid changes; for two short segments derived *S. enterica* serovar Abony SW803 (H1, 15.4 kb and H2, 41.7 kb), all annotated loci are listed as well, since few positions vary from the Abony reference. The reference segments H1 and H2 were extracted from *S. enterica* subsp. *enterica* serovar Abony str. 0014—NZ_CP007534.1: for transduction 1 of Supplementary Table S1, H1-nt 3956473-3942278 (reversed); for transduction 2, H2-nt 4713943-4725447 (catenated)1-18493 (the segment crosses the numerical origin, so is catenated). Supplementary File S2 “Analysis of 11 kb high-similarity region in the Gifsy prophages” describes and displays alignments (referred to as “Embedded figs”) that resolve allele configurations in Gifsy prophages. Supplementary File S3 “STK005 genotype inferences from sequence” describes the recorded genotype; explains assignment of flagellar genotype; accounts for IS element annotations in LT2 and STK005; describes the protocol used to generate the variant list found in Supplementary File S1; lists protein changes that account for the genotype; and points out some additional possible changes. Supplementary File S4 “PCR survey for 42 bp insertion in isolates of *S. enterica* sv Typhimurium str LT2” demonstrates the insertion is present in Gifsy-1 and conserved in most LT2 strains.

All the sequencing data can be found in the BioProject PRJNA605961. The genome sequences of STK005 and ER3625 are available from GenBank with the accession numbers (CP067397-CP067398) and (CP067091-CP067092), respectively. All the raw data are available under the SRA ID SRR13390529, (SRR13404095, SRR13404094, SRR13404093, SRR13404092, and SRR13404091) and

Table 1 DNA modification systems, motifs and degree of modification

System ^a	Motif ^b	MOD type ^c	% M ^d STK005	% M ^d ER3625
M.SenLT2I (StyLT, StyLTI)	5' CAGAG 3'	m6A	99.9	99.9
M.SenLT2II (StySB)	5' GAGNNNNNNRTAYG 3' 3' CTCNNNNNNYATRC 5'	m6A	99.1	98.3
M.SenLT2III (StySA)	5'-GATCAG-3'	m6A	100	100
M.SenLT2Dam	5' GATC 3'	m6A	0.3	0
M.SenLT2Dcm	5' CCWGG 3'	m6A	99.9	99.9
M.SenLT2IV	5'-ATGCAT-3'	(m5C)	(1.0)	(0.3)
		m6A	24.9	38.2

^a System: name used in REBASE for Organism number 18099 followed by (names most encountered in the literature)

^b Motif: recognition sequence with modified base in bold. Sites shown once are symmetric, or asymmetric sites methylated only on one strand; sites shown twice are asymmetric sites methylated on both strands.

^c Mod type: m6A = N6-methyladenine; m5C = 5 methylcytosine. Dcm (m5C) motif is inferred from the presence of the *dcm* gene; m5C is not efficiently detected by this approach.

^d %M: [# motif sites modified (IPD >2, 25X coverage)/motif occurrences] × 100.

SRR13787361 for ER3625, STK005 RSII, and STK005 Nanopore respectively. The methylation data are available as supplementary files on NCBI under the IDs SUPPF_0000003906 and SUPPF_0000003907. Both strains are available from New England Biolabs.

Acknowledgments

The authors thank Stanley Maloy and Anca Segall (San Diego State University) for *Salmonella* strains, phages and advice, and training in their use. We thank Chloé Baum and Giron Koetsier for technical assistance with the Oxford Nanopore sequencing and high molecular weight DNA extraction. They thank an anonymous reviewer for drawing attention to IS200 properties and lead references, making the paper stronger.

Funding

New England Biolabs provided support for J.Z., O.D., T.K., A.F., R.M., and E.A.R.

Conflicts of interest

None declared.

Literature cited

- Anton BP, Mongodin EF, Agrawal S, Fomenkov A, Byrd DR, et al. 2015. Complete genome sequence of ER2796, a DNA methyltransferase-deficient strain of *Escherichia coli* K-12. *PLoS One*. 10:e0127446.
- Azam AH, Tanji Y. 2019. Bacteriophage-host arm race: an update on the mechanism of phage resistance in bacteria and revenge of the phage with the perspective for phage therapy. *Appl Microbiol Biotechnol*. 103:2121–6611.
- Balabanov VP, Kotova VY, Kholodii GY, Mindlin SZ, Zavilgelsky GB. 2012. A novel gene, *ardD*, determines antirestriction activity of the non-conjugative transposon Tn5053 and is located antisense within the *tmiA* gene. *FEMS Microbiol Lett*. 337:55–60.
- Bullas LR, Colson C. 1975. DNA restriction and modification systems in *Salmonella*. III. SP, a *Salmonella potsdam* system allelic to the SB system in *Salmonella typhimurium*. *Mol Gen Genet*. 139:177–188.
- Bullas LR, Ryu JI. 1983. *Salmonella typhimurium* LT2 strains which are $r^+ m^+$ for all three chromosomally located systems of DNA restriction and modification. *J Bacteriol*. 156:471–474.
- Chaudhuri RR, Peters SE, Pleasance SJ, Northen H, Willers C, et al. 2009. Comprehensive identification of *Salmonella enterica* serovar Typhimurium genes required for infection of BALB/c mice. *PLoS Pathog*. 5:e1000529.
- Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, et al. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res*. 40:e29.
- Colson C, Colson AM. 1971. A new *Salmonella typhimurium* DNA host specificity. *J Gen Microbiol*. 69:345–351.
- Colson CC, Van Pel AA. 1974. DNA restriction and modification systems in *Salmonella*. I. SA and SB, two *Salmonella typhimurium* systems determined by genes with a chromosomal location comparable to that of the *Escherichia coli* *hsd* genes. *Mol Gen Genet*. 129:325–337.
- De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, et al. 2014. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet*. 10:e1004181.
- Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, et al. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*. 359:eaar4120.
- Ellis MJ, Trussler RS, Charles O, Haniford DB. 2017. A transposon-derived small RNA regulates gene expression in *Salmonella typhimurium*. *Nucleic Acids Res*. 45:5470–5486.
- Ellis MJ, Trussler RS, Haniford DB. 2015. A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition. *Nucleic Acids Res*. 43:6511–6527.
- Figueroa-Bossi N, Coissac EE, Netter PP, Bossi LL. 1997. Unsuspected prophage-like elements in *Salmonella typhimurium*. *Mol Microbiol*. 25:161–173.
- Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L. 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol Microbiol*. 39:260–271.
- Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, et al. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 7:461–465.
- Fomenkov A, Sun Z, Murray IA, Ruse C, McClung C, et al. 2020. Plasmid replication-associated single-strand-specific methyltransferases. *Nucleic Acids Res*. 48:12858–12873.
- Fuller-Pace FV, Bullas LR, Delius H, Murray NE. 1984. Genetic recombination can generate altered restriction specificity. *Proc Natl Acad Sci USA*. 81:6095–6099.
- González-Montes L, Del Campo I, Garcillán-Barcia MP, de la Cruz F, Moncalián G. 2020. ArdC, a ssDNA-binding protein with a metalloprotease domain, overpasses the recipient *hsdRMS* restriction system broadening conjugation host range. *PLoS Genet*. 16:e1008750.
- Griewisch KF, Pierce JG, Elfenbein JR. 2020. Genetic determinants of *Salmonella* resistance to the biofilm-inhibitory effects of a synthetic 4-oxazolidinone analog. *Appl Environ Microbiol*. 86:e01120-20.
- 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.
- Haack KR, Roth JR. 1995. Recombination between chromosomal IS200 elements supports frequent duplication formation in *Salmonella typhimurium*. *Genetics*. 141:1245–1252.
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvermin V, et al. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res*. 46:D851–D860.
- Hattman S, Schlagman S, Goldstein L, Frohlich M. 1976. *Salmonella typhimurium* SA host specificity system is based on deoxyribonucleic acid-adenine methylation. *J Bacteriol*. 127:211–217.
- Hoiseith SK, Stocker BA. 1981. Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature*. 291:238–239.
- Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, et al. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol*. 16:2601.
- Jones-Carson J, Yahashiri A, Kim J-S, Liu L, Fitzsimmons LF, et al. 2020. Nitric oxide disrupts bacterial cytokinesis by poisoning purine metabolism. *Sci Adv*. 6:eaaz0260.
- Joys TM, Stocker BA. 1965. Complementation of non-flagellate *Salmonella* mutants. *J Gen Microbiol*. 41:47–55.

- Joys TM, Stocker BA. 1969. Recombination in *H1*, the gene determining the flagellar antigen-i of *Salmonella typhimurium*; mapping of *H1* and *fla* mutations. *J Gen Microbiol.* 58:267–275.
- Kingston AW, Ponkratz C, Raleigh EA. 2017. Rpn (YhgA-Like) proteins of *Escherichia coli* K-12 and their contribution to RecA-independent horizontal transfer. *J Bacteriol.* 199:e00787-00716-00717.
- Kingston AW, Roussel-Rossin C, Dupont C, Raleigh EA. 2015. Novel *recA*-Independent horizontal gene transfer in *Escherichia coli* K-12. *PLoS One.* 10:e0130813.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722–736.
- Lemire S, Figueroa-Bossi N, Bossi L. 2007. Prophage contribution to *Salmonella* virulence and diversity. In: HSM Hensel, editor. *Horizontal Gene Transfer in the Evolution of Bacterial Pathogenesis.* Cambridge, UK: Cambridge University Press. p. 159–191.
- Lemire S, Figueroa-Bossi N, Bossi L. 2008. A singular case of prophage complementation in mutational activation of *recET* orthologs in *Salmonella enterica* serovar Typhimurium. *J Bacteriol.* 190: 6857–6866.
- Lerche M, Sandhu H, Flöckner L, Högbom M, Rapp M. 2017. Structure and cooperativity of the cytosolic domain of the CorA Mg²⁺ channel from *Escherichia coli*. *Structure.* 25:1175–1186.e1174.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 34:3094–3100.
- Martinsohn JT, Radman M, Petit MA. 2008. The λ Red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet.* 4: e1000065.
- McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, et al. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature.* 413:852–856.
- Murphy K. C. 2016. λ Recombination and Recombineering. *EcoSal Plus* 7. 10.1128/ecosalplus.esp-0011-2015.
- Murray IA, Morgan RD, Luyten Y, Fomenkov A, Correa IR, Jr, et al. 2018. The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* 46:840–848.
- Ornellas EP, Stocker BAD. 1974. Relation of lipopolysaccharide character to P1 sensitivity in *Salmonella typhimurium*. *Virology.* 60: 491–502.
- Pirone-Davies C, Hoffmann M, Roberts RJ, Muruvanda T, Timme RE, et al. 2015. Genome-wide methylation patterns in *Salmonella enterica* Subsp. *enterica* Serovars. *PLoS One.* 10:e0123639.
- Piya D, Vara L, Russell WK, Young R, Gill JJ. 2017. The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. *Mol Microbiol.* 105:399–412.
- Roberts GA, Stephanou AS, Kanwar N, Dawson A, Cooper LP, et al. 2012. Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Res.* 40:8129–8143.
- Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 43:D298–D299.
- Sanderson K. E. 1987. F-Mediated Conjugation, F + Strains, and Hfr Strains of *Salmonella typhimurium* and *Salmonella abony*. In: Neidhardt F. C., III Roy Curtiss, Ingraham John L, Lin E C, Low K Brooks, et al. (eds), *Escherichia coli and Salmonella: cellular and molecular biology.* Washington, D.C.: ASM Press. pp. 2406–2412.
- Sanderson KE, Roth JR. 1988. Linkage map of *Salmonella typhimurium*, edition VII. *Microbiol Rev.* 52:485–532.
- Schmieger H. 1999. Molecular survey of the *Salmonella* phage typing system of Anderson. *J Bacteriol.* 181:1630–1635.
- Segall AM, Roth JR. 1989. Recombination between homologies in direct and inverse orientation in the chromosome of *Salmonella*: intervals which are nonpermissive for inversion formation. *Genetics.* 122:737–747.
- Serfiotis-Mitsa D, Herbert AP, Roberts GA, Soares DC, White JH, et al. 2010. The structure of the KlcA and ArdB proteins reveals a novel fold and antirestriction activity against Type I DNA restriction systems in vivo but not in vitro. *Nucleic Acids Res.* 38:1723–1737.
- Sharples GJ, Curtis FA, McGlynn P, Bolt EL. 2004. Holliday junction binding and resolution by the Rap structure-specific endonuclease of phage λ . *J Mol Biol.* 340:739–751.
- Simon M, Zieg J, Silverman M, Mandel G, Doolittle R. 1980. Phase variation: evolution of a controlling element. *Science.* 209:1370–1374.
- Spicer CC, Datta N. 1959. Reversion of transduced antigenic characters in *Salmonella typhimurium*. *J Gen Microbiol.* 20:136–143.
- Stocker B, Males BM, Takano W. 1980. *Salmonella typhimurium* mutants of RfaH⁻ phenotype: genetics and antibiotic sensitivities. *J Gen Microbiol.* 116:17–24.
- Tarkowski TA, Mooney D, Thomason LC, Stahl FW. 2002. Gene products encoded in the *ninR* region of phage λ participate in Red-mediated recombination. *Genes Cells.* 7:351–363.
- Tatusova T, DiCuccio M, Badretdin A, Chetvermin V, Nawrocki EP, et al. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44:6614–6624.
- Timme RE, Pettengill JB, Allard MW, Strain E, Barrangou R, et al. 2013. Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. *Genome Biol Evol.* 5:2109–2123.
- Tsai SP, Hartin RJ, Ryu J. 1989. Transformation in restriction-deficient *Salmonella typhimurium* LT2. *J Gen Microbiol.* 135:2561–2567.
- Vandecraen J, Chandler M, Aertsen A, Houdt RV. 2017. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol.* 43:709–730.
- Yoon W, Yoo Y, Chae YS, Kee SH, Kim BM. 2018. Therapeutic advantage of genetically engineered *Salmonella typhimurium* carrying short hairpin RNA against inhibin alpha subunit in cancer treatment. *Ann Oncol.* 29:2010–2017.
- Zavilgelsky GB, Rastorguev SM. 2009. Antirestriction proteins ArdA and Ocr as efficient inhibitors of Type I restriction-modification enzymes. *Mol Biol.* 43:241–248.
- Zaworski J, McClung C, Ruse C, Weigele PR, Hendrix RW, et al. 2021. Genome analysis of *Salmonella enterica* serovar Typhimurium bacteriophage L, indicator for StySA (StyLT2III) restriction-modification system action. *G3 (Bethesda).* 11.

Communicating editor: B. J. Andrews