Diverse Durham collection phages demonstrate complex BREX defence responses

Abigail Kelly[a,*] Sam C. Went[a,*] Giuseppina Mariano[b], Liam P. Shaw[a,c], David M. Picton[a], Samuel J. Duffner[a], Isabel Coates[a], Ryan Herdman-Grant[a], Julia Gordeeva[d], Alena Drobiazko[d], Artem Isaev[d], Yan-Jiun Lee[e], Yvette Luyten[e], Richard D. Morgan[e], Peter Weigele[e], Konstantin Severinov[f], Nicolas Wenner[g], Jay C. D. Hinton[g], Tim R. Blower[a,§]

[a]Department of Biosciences, Durham University, Stockton Road, Durham, DH1 3LE, UK.

[b]Microbes in Health and Disease Theme, Newcastle University Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

[c]Department of Biology, University of Oxford, 11a Mansfield Road, Oxford, OX1 3SZ, UK.

[d]Skolkovo Institute of Science and Technology, Moscow 143028, Russia.

[e]New England Biolabs, 240 County Road, Ipswich, MA 01938, USA.

[f]Waksman Institute of Microbiology, Piscataway, NJ 08854, USA.

[g]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, L69 7ZB, UK.

[*]These authors contributed equally to this work.

[§]To whom correspondence may be addressed. Email: timothy.blower@durham.ac.uk, tel: +44(0)1913343923.

Abstract

Bacteriophages (phages) outnumber bacteria ten-to-one and cause infections at a rate of 10^{25} per second. The ability of phages to reduce bacterial populations make them attractive alternative antibacterials for use in combatting the rise in antimicrobial resistance. This effort may be hindered due to bacterial defences such as Bacteriophage Exclusion (BREX) that have arisen from the constant evolutionary battle between bacteria and phages. For phages to be widely accepted as therapeutics in Western medicine, more must be understood about bacteria-phage interactions and the outcomes of bacterial phage defence. Here, we present the annotated genomes of twelve novel bacteriophage species isolated from water sources in Durham, UK, during undergraduate practical classes. The collection includes diverse species from across known phylogenetic groups. Comparative analyses of two novel phages from the collection suggests they may be founding members of a new genus. Using this Durham phage collection we determined that particular BREX defence systems were likely to confer a varied degree of resistance against an invading phage. We concluded that the number of BREX target motifs encoded in the phage genome was not proportional to the degree of susceptibility.

Importance

Bacteriophages have long been the source of tools for biotechnology that are in everyday use in molecular biology research laboratories worldwide. Phages make attractive new targets for the development of novel antimicrobials. Whilst the number of phage genome depositions has increased in recent years, the expected bacteriophage diversity remains under-represented. Here we demonstrate how undergraduates can contribute to the identification of novel phages, and that a single City in England can provide ample phage diversity and the opportunity to find novel technologies. Moreover, we demonstrate that the interactions and intricacies of the interplay

between bacterial phage defence systems such as BREX and phages are more complex than originally thought. Further work will be required in the field before the dynamic interactions between phages and bacterial defence systems are fully understood and integrated with novel phage therapies.

Keywords

Bacteriophage, phylogenetics, autographiviridae, phage defence, BREX

Introduction

The interplay between bacteria and their viral predators, bacteriophages (phages), generates huge selection pressure driven by an estimated infection rate of 10^{25} per second (1). This constant challenge-and-response has forced the evolution of a broad diversity of bacterial anti-phage defence strategies including; restriction-modification (RM) systems (2), abortive infection systems (3–5) and CRISPR-*cas* (6). The recent increase in the availability of high-quality sequencing data has allowed a "guilt-by-association" approach to identifying defence islands. This has led to the discovery of a plethora of new anti-phage systems including; viperins (7), CBASS (8), PARIS (9) amongst others (10–12), many of which utilise conserved protein domains (13). The clustering of defence systems into islands also implied a means of co-regulation, which was demonstrated by the discovery of the widespread BrxR WYL-domain transcriptional repressors (14–16).

Analysis of genes linked to *pglZ* from the phage growth limitation defence system (17) identified Bacteriophage Exclusion (BREX) systems, which are encoded within 10% of bacterial and archaeal genomes (18). Together with *gmrS*/*gmrD* that encode a type IV restriction enzyme (19), BREX genes form one of the most common defence island pairings (20, 21). We have recently demonstrated that BREX and a GmrSD homologue, BrxU, both produced from a defence island encoded on a multidrug-resistant plasmid of *Escherichia fergusonii*, provide complementary phage defence characteristics (22). Of the six BREX sub-types, type I BREX contains six conserved genes; *brxA*, *brxB*, *brxC*, *pglX*, *pglZ* and *brxL* (18). BrxA is a DNA-binding protein (23), BrxL is a DNA-stimulated AAA+ ATPase (24), whilst the methyltransferase PglX methylates non-palindromic 6 bp sequences (BREX sites) on the N6 adenine at the fifth position of the motif, distinguishing host from foreign DNA (18, 22, 25, 26). Though reminiscent of restriction-modification, the mechanism of BREX activity is currently unknown.

Phage collections can be used to aid mechanistic studies of defence systems (27). They allow a consistent approach that will aid comparisons between homologues of defence systems, and between different types of defence systems. They also allow interrogation of defence mechanisms in reverse, using spontaneous resistant mutants derived from the phage collection to pinpoint key viral factors (28–30). Currently, we have characterised only a tiny fraction of the phage species representing the estimated $>10^{30}$ phage particles on Earth (31, 32). As phages have historically been exploited in a range of industries, for example in food production (33–35), as a source of molecular biology tools such as T7 RNA polymerase and T4 DNA ligase, and in the production of monoclonal antibodies (36), it stands to reason that continued exploration of phage sequence space will yield more tools. Due to the increasing prevalence of antimicrobial resistance (AMR) phages are also once again gaining traction as a viable therapeutic route in Western medicine, building upon 70 years of therapeutic use in countries such as Poland, Georgia and Russia (37, 38). Before phage therapy can become a widely used treatment option more must be learnt about bacterial responses to phages, and about any pitfalls the continuous arms-race between phages and bacteria may pose in clinical outcomes (37, 38). All of these aims rely on the availability of characterised phage collections and sequence data.

Here, we characterise a collection of *Escherichia coli* phages isolated during a final year undergraduate practical workshop module that was inspired by a range of similar taught courses, such as the SEA-PHAGES program (39). Students in this class isolated phages from waterways around Durham, UK, for testing against BREX phage defence systems. We present morphological, genomic and phylogenetic characterisation of a set of 12 new phages. Notably, two phages have low similarity to any other species in the NCBI database and fulfil ICTV criteria (40, 41) for deposition as related members of the same species, within a putative novel genus. We then use this Durham

collection to test three BREX homologues. The BREX responses proved to be varied, with scale of response having little correlation to number of BREX motifs, and in multiple cases differing BREX homologues having strikingly different impacts on the same phage. These mechanistic outcomes highlight the utility of the Collection and set-up future studies into the BREX response and other bacteria-phage interactions.

Results

Diverse coliphages exist in the waterways around Durham City

During 2016 and 2017, samples of water from natural sources around Durham City (UK) were collected as part of an undergraduate practical course. Abridged details are available in **Table 1**, with full collection details available in **Table S1**. These samples were enriched for phages able to infect *Escherichia coli* DH5α, as previously described (22). The resulting plaques went through three rounds of plaque purification, and the final set of 30 coliphages were made into lysates for further study. Characterisation of this initial "Durham collection" demonstrated how the phages responded to a phage defence island encoded on a multidrug-resistant plasmid, pEFER, from *Escherichia fergusonii* ATCC35469 (22). The pEFER defence island encodes a Bacteriophage Exclusion (BREX) system and a GmrSD family type IV restriction enzyme, BrxU. Through selective gene deletion followed by Efficiency of Plating (EOP) assays, 13 of the phages were shown to be sensitive to BrxU (22). Here, we sequenced the genomes of the 17 phages that were not BrxU sensitive (reserving the other 13 for a future study into DNA modification preference by BrxU). The genomes were sequenced using either Illumina Hiseq or Novaseq 6000, or Oxford Nanopore MinION sequencing. Having analysed the genomic data, it was apparent that the 17 phages included some redundancy. As a result, duplicates were removed, leaving a final set of 12 phages for downstream characterisation. As part of removing redundancy the phage "PATM", observed to be the same as phage "Pau", was excluded. This was interesting as Pau was isolated in 2016 from the River Wear, then PATM was isolated again from the River Wear a year later, suggesting that Pau/PATM perhaps represent a stable part of the microbial community, as has been described for other sites (42). We used CheckV v1.0.1 (43) to assess the quality of the remaining 12 phage genomes using the full pipeline. All 12 phage genomes showed no contamination and had high completeness scores (range: 95.3-100%, n=7 with 100% completeness; see **Table S1**). They were all high-quality as per the Minimum Information about an Uncultivated Virus Genome (MIUViG) standard (44). The internal CheckV quality of the genomes was

at least 'high-quality' (n=8) with some assessed as 'complete' (n=4: BB1, CS16, Pau, Sip). These 12 genome sequences were annotated and submitted to Genbank. Whilst each phage was given its own unique scientific name on submission to ENA, we use the abbreviated names throughout this manuscript for ease (**Table 1** and **Table S2**).

The obtained genomes were used for comparison against known phages in the database and to examine relatedness within the collection through generation of a phylogenetic tree (**Fig. 1**). This subset of the Durham Collection appears diverse, and includes representatives of the *Demerecviridae* (BB1), *Drexlerviridae* (Sip), *Schitoviridae* (Jura), and *Straboviridae* (TB34) families, the *Vequintavirinae* (Alma), *Stephanstirmvirinae* (Pau) and *Ounavirinae* (Trib and Baz) subfamilies, and the *Dhillonvirus* (CS16 and Mav) genus, as per the current ICTV taxonomy release (https://ictv.global/taxonomy). Intriguingly, Mak and Bam could not be classified at the genus level; this is explored later.

Snippy (https://github.com/tseemann/snippy) was used to probe for differences between the three pairs of closely related Durham phages, specifically CS16 and Mav, Mak and Bam, and Trib and Baz (**Fig. S1** and **Table S3**). CS16 and Mav are very similar and share 99.96% nucleotide identity over 100% of the genome (BLAST-N), with six single nucleotide variations identified by Snippy (four within hypothetical genes, one in the putative major tail protein and one intergenic SNP) (**Fig. S1** and **Table S3**). Bam and Mak could not be readily classified into existing families, but again are very similar at the nucleotide level, with 99.99% identity over 100% coverage. According to recent guidelines (45), two phages are considered the same species if they are more than 95% identical at the genome level; therefore Mak and Bam can be considered the same species. Snippy identified 18 single nucleotide variations between Mak and Bam (**Fig. S1** and **Table S3**), included within genes for the

type II holin, DNA-directed RNA polymerase, DNA ligase, DNA polymerase and three hypothetical genes (**Fig. S1** and **Table S3**). Prior phylogenetic analysis suggested that Trib and Baz belong to the same genus, Felixounavirus (**Fig. 1**). Indeed, at the nucleotide level the phages appear to be very similar (99.07% identity over 95% coverage). Nevertheless, Snippy identified 1466 single nucleotide variations between Trib and Baz, distributed throughout the genomes (**Table S2**).

Transmission electron microscopy was performed on representative members of each of the remaining nine phage groups (**Fig. 2**). Alma, Baz, Pau and TB34 were visualised with rigid tails and baseplates suggesting a myoviral-like structure, Jura and Mak have short tail structures with overall podoviral morphology, and BB1, Mav and Sip were observed to have long flexible tails indicative of a siphoviral morphology (**Fig. 2**). All phages were imaged with full capsids, except for Sip that could only be visualised with an empty capsid (**Fig. 2**). Curiously, BB1, was shown to produce a large range of tail-lengths, which included vastly extended and flexible tails (**Fig. S2**). This may indicate a mutation within the tail assembly apparatus, such as the tape-measure protein, that allows these diverse structures to form. Collectively, these data show genomic and morphological characterisation of our first cohort of coliphages in the Durham collection.

The Durham coliphages represent species across the collected bacteriophage taxonomy

Next, we explored phylogenetic relationships between the Durham phages and existing viral genomes in NCBI. First, we used mash screen (mash v2.3) (46), to compare the Durham phages against all NCBI RefSeq genomes (using a compressed sketch database downloaded from https://gembox.cbcb.umd.edu/mash/refseq.genomes.k21s1000.msh on 23rd February 2023). The top hits from the mash screen (**Table S4**) were all part of the family *Caudoviricetes*, which includes

tailed bacteriophages. We therefore downloaded all (n=29,188) NCBI Genbank genomes within *Caudoviricetes* as of 1st March 2023 (see Materials and Methods). We used fastANI v1.33 (fragment length 3,000 bp, minimum genome fraction 0.05) (47) to calculate approximate average nucleotide identity (ANI) of the Durham phage collection against all these genomes. The results confirmed that the majority of the phages (10 out of 12, excluding Mak and Bam) could be classified to at least genus level (**Table S5**).

To further contextualise the Durham phages relative to *Escherichia* phages in particular, we first generated a subset of our NCBI dataset for the *Caudoviricetes* genomes to include only those isolated from *Escherichia* (see Materials and Methods), resulting in 2,772 genomes. We then used VIPtree (48) to build phylogenies based on proteomes. Because VIPtree online allows inputs of ≤200 genomes, we used an all-against-all approach to compare all *Escherichia*-*Caudoviricetes* genomes to each other using fastANI v1.33 and then used the resulting similarity matrix to subsample down to *Caudoviricetes* genomes that were classified to genus level. We subsequently selected one representative genome for each genus ensuring the chosen genomes were maximally representative of NCBI diversity (**Table S6**), fetching the taxonomic information for these representatives from NCBI using the rentrez v1.2.3 package in R v4.1.2 (49). We then manually added in the 12 Durham phages, and phages T1-7 plus Lambda as well-known examples of coliphages. The resulting phylogenetic tree for this curated set of coliphage genomes demonstrates that the Durham coliphages include representatives covering a wide breadth of known coliphage diversity (**Fig. 3**). Intriguingly, throughout these analyses, phages Bam and Mak, whilst similar to each other, had no sufficiently close relative and could not be assigned to a specific genus (**Fig. 3**).

Escherichia phages Mak and Bam likely represent a novel bacteriophage genus

As there was no initial classification of Mak and Bam, we looked more deeply into potential known relatives. Mak and Bam were isolated in the same year, from sites approximately 800 m away from each other along the River Wear (Durham, UK). The genome sequences for both phages were used in an initial BLAST-N search to find the most similar deposited relative. For both Mak and Bam, *Salmonella* phage Rostov-1 (Genbank ID: MG957431.1) was identified as the top-scoring hit though this was poor, with a shared identity of 78% across only 7% of the genomes. Mak and Bam each have genome lengths of 40,137 bp, 47% GC content and differ by 19 SNPs (**Fig. S1B** and **Table S3**). Each was annotated as encoding 48 and 54 putative proteins, respectively, and no tRNAs were identified in either genome. A circularly permuted representation of the Mak genome (**Fig. 4**) shows how hypothesised structural and nucleotide metabolism genes tend to cluster together, respectively, and all but one predicted ORF is encoded on the positive strand. The genome of Bam was re-arranged to match the same starting point as Mak (**Fig. S3**). Automated annotation identified an additional five putative ORFs on the negative strand of Bam (**Fig. S3**). Whilst the majority of annotated putative proteins encoded by Mak and Bam do share sufficient similarity with known proteins to suggest function, there are exceptions (**Fig. 4** and **Fig. S3**). Based on the respective automatically annotated set of ORFs for each phage genome, 11 of 48 (23%) of predicted Mak ORF products and 19 of 54 (35%) of predicted Bam ORF product could not be assigned a likely protein family by InterProScan. For these outliers, there were no proteins with significant similarity within the NR database used for BLAST-P searches, nor could any domain families be assigned using InterProScan. This lack of similarity to proteins in published databases has been observed elsewhere (50). It may be that this does indeed indicate novel proteins, though there remains the possibility that these annotated coding sequences could perhaps not be actively expressed.

229 It was noted that Mak and Bam both encode a T7-like RNA polymerase (e.g. Mak CDS 35693-38377
230 bp), which would suggest that they both are members of the *Autographiviridae* family (50). A
231 curated subset of representatives from *Autographiviridae* was generated, consisting of 110 genomes
232 (**Table S7**). We then again used VIPtree (48) to generate phylogenies built on proteomes of these
233 representatives, alongside Mak and Bam (**Fig. 5**). The resulting tree demonstrates that within the
234 *Autographiviridae* family Mak and Bam cluster separately from the other genera and the closest
235 genus is represented by *Chatterjeevirus* (**Fig. 5**). As a result, we chose to look more closely into
236 subfamily relationships within *Autographiviridae*.

237

238 Using mash (46) and fastANI (47) the closest NCBI genomes to both Mak and Bam were within the
239 subfamily *Studiervirinae*, genus *Chatterjeevirus* (for example, NC_013651 which had an estimated
240 ANI of ~77% across 15% of the genome). We therefore downloaded all Genbank genomes within the
241 subfamily *Studiervirinae* (9th March 2023, n=865) to contextualise Mak and Bam within the
242 *Studiervirinae*. ICTV 2021 v3 lists 24 viral genera within *Studiervirinae*, of which 21 have
243 representative genomes in Genbank. We applied the same 'all-against-all' subsampling approach
244 previously used to select our sub-sets of representative genomes and analysed them together with
245 Mak and Bam in VIPtree online (**Fig. 6A**). We also performed VIPtree analysis with all 27 available
246 *Chatterjeevirus* NCBI Genbank genomes (**Fig. 6B**). This confirmed that Mak and Bam represent novel
247 phage diversity even when compared directly against their closest known relatives: they cluster
248 away from known viral genera and they cluster separately from known species of the *Chatterjeevirus*
249 genus (**Figs. 5** and **6**). In our attempt to establish the taxonomy of Mak and Bam, it became clear that
250 these two phages cluster separately from any known viral genus and species. Bam and Mak showed
251 no significant sequence identity to any phage genome through BLASTn and we could not find any
252 close relative either through PhageClouds (51) or mash (46). Given that 70% sequence identity has
253 been defined as the cut-off threshold for the definition of distinct genera (45), we postulate that

Mak and Bam are the first representatives of a new genus and are themselves members of the same new species. We humbly suggest that “Palatinivirus”, might be an appropriate Genus name, in reference to the history of County Durham.

BREX phage defence does not correlate to number of BREX motifs

Having established a characterised collection of phages, we returned to our initial motivation for isolating novel phages; investigating phage defence mechanisms. BREX systems have been shown to perform N6-adenine methylation at the 5th position of non-palindromic 6 bp motifs (18, 22, 25, 26). This distinguishes the host methylated DNA from incoming non-methylated DNA. Having obtained genomic data for our phages, we could now analyse the genomes for the presence of target BREX motifs. We chose to study three BREX systems as a means to investigate; (i) how the number and distribution of BREX motifs in a phage genome might impact the BREX response; and (ii) whether different BREX systems respond in the same way to the same phage. We selected three BREX systems for which the BREX motif has already been mapped, and that have been shown to be active in *E. coli*, and so can be tested against our phages. These are the BREX defence island from *E. fergusonii* plasmid pEFER (22), the BREX system from *E. coli* HS (25), and the BREX defence island from a sub-Saharan epidemic strain of *Salmonella enterica* serovar Typhimurium D23580 that is a close homologue to a recently reconstructed lab *Salmonella* BREX system (26, 52) (**Fig. 7A**). Notably, the pEFER defence island encodes BREX and BrxU, a type IV enzyme, but our test phages were specifically selected as they are not targeted by BrxU, so the presence of BrxU does not interfere with the experiment (22) (**Fig. 7A**). The *E. coli* BREX system has no known additional defence system genes (**Fig. 7A**). Both the pEFER and *E. coli* systems were sub-cloned onto vectors for use in *E. coli* DH5α (22, 25). The *S*. Typhimurium BREX defence island encodes a BREX system and a putative PARIS phage defence system (9) (**Fig. 7A**). We sub-cloned the *S*. Typhimurium BREX system again for

use in *E. coli* DH5α and deleted the two PARIS genes, *ariA* and *ariB*, to ensure that our experiments remain focussed on BREX phage defence (**Fig. 7A**).

The pEFER BREX system ($BREX_{Eferg}$) recognises GCTA**A**T motifs (bold underlining indicates the N6-modified adenine). The *E. coli* BREX system ($BREX_{Ecoli}$) recognises GGTA**A**G motifs and the *S.* Typhimurium BREX system ($BREX_{Sty}$) recognises GATC**A**G motifs. We mapped the number and positions of these three BREX motifs in each of our Durham phages (**Fig. 7B**). In the three cases where there are closely related phages (CS16/Mac, Mak/Bam, Trib/Baz), the genomes were aligned at the same origin to allow comparison. This analysis showed great variation in the number of each motif across the phages, from a high of 120 (the $BREX_{Sty}$ motif in TB34) and even zero in some cases (the $BREX_{Sty}$ motif in phages Trib, Baz and Jura) (**Fig. 7B**). When BREX motifs were present, they appeared to be spread throughout the phage genome without obvious clustering or strand preference (**Fig. 7B** and **Fig. S4**).

To investigate diversity amongst BREX responses we performed EOP assays for all phages against each of the three BREX systems (**Fig. 7C** and **Table S8**). All three BREX systems were shown to provide active phage defence against select phages (**Fig. 7C**). Phages Trib, Pau, CS16 and Sip were reduced for EOP by cells expressing $BREX_{Eferg}$ (**Fig. 7C**), as observed previously (22). Phages TB34, Trib, Mak, Bam, CS16, Mav and Sip were reduced for EOP by cells expressing $BREX_{Ecoli}$ (**Fig. 7C**). Phages TB34, Alma, PATM, BB1, CS16, Mav and Sip were reduced for EOP by cells expressing $BREX_{Sty}$ (**Fig. 7C**). With all three BREX systems, the scale of phage defence varied from low protection (10-100-fold), to high protection (EOPs of $<10^{-6}$) (**Fig. 7C**). In no case was there a correlation between the number of BREX motifs in the phage genome and the scale of BREX phage defence. For instance, examining $BREX_{Eferg}$, Trib (44 motifs) and Sip (24 motifs) are similarly impacted 10- to 100-fold, but

CS16 (23 motifs) is impacted 100- to 1000-fold and Pau (82 motifs) has an EOP of $<10^{-6}$ (**Figs. 7B and C**). A clearer example showing how number of motifs does not correlate with BREX phage defence is $BREX_{Ecoli}$; Bam and CS16 are the most impacted and have 41 and 32 motifs, respectively, whilst TB34 has only a 10- to 100-fold reduction and encodes 59 motifs (**Figs. 7B and C**). This indicates other factors are involved in determining scale of BREX defence beyond simply number of BREX motifs.

It might be expected that phages without BREX motifs are not targeted, and that was the case for Trib, Baz and Jura against $BREX_{Sty}$ (**Figs. 7B and C**). Interestingly, Mak and Bam have one $BREX_{Sty}$ motif but that was insufficient to induce BREX defence (**Figs. 7B and C**). Nevertheless, drawing conclusions about numbers of BREX motifs and impact of phage defence from only those phages that are reduced for EOP ignores the many phages with BREX motifs that were not targeted. We know that phages can encode anti-phage defence genes (53). Examining $BREX_{Eferg}$ in isolation, we would therefore be led to conclude that TB34, Baz, Alma, BB1, Jura, Mak, Bam and Mav all encode anti-BREX genes. We performed searches in all 12 phage genomes for genes encoding homologues of the known anti-BREX protein Ocr (53), and a range of other known phage encoded anti-defence factors (**Table S9**). Unexpectedly, we did not find any obvious candidates, but of course that does not rule out the possibility for novel BREX inhibitors. This suggests there may be further, yet to be identified, anti-BREX mechanisms at work in these phages.

It is also important to consider BREX responses *across* the three systems, against a single phage. For instance, with Pau, the genome encodes a similar number of each BREX motif, but only $BREX_{Eferg}$ provides phage defence against Pau (**Figs. 7B and C**). This implies that Pau may encode an as yet unidentified anti-BREX gene that can specifically inhibit $BREX_{Ecoli}$ and $BREX_{Sty}$, but not $BREX_{Eferg}$. Alternatively, Pau may not encode an anti-BREX gene, but there could be something fundamental

about the BREX mechanism that allows targeting of Pau by one BREX system, but not the other two. These possibilities can be examined using the pairs of closely related phages. CS16 and Mav have only six SNPs between them but vary in response to $BREX_{Eferg}$ and $BREX_{Ecoli}$ (**Fig. 7C**). Crucially, CS16 and Mav both have some susceptibility to $BREX_{Ecoli}$ and are equally susceptible to $BREX_{Sty}$, so it is not likely to be mutation of a single anti-BREX gene that makes the difference. In these cases, the number of BREX motifs stayed the same. These data suggest there is a phenotypic change as a result of the SNPs, dictating the BREX response. Without further molecular mechanistic data, outside the remit of this study, it is not yet possible to hypothesise what this change might be. However, having identified these changes it may be possible to now pursue mechanistic models through isolation and characterisation of spontaneously BREX-resistant phages.

These EOP data collected for our phages tested against three BREX systems have shown a wide and complex diversity of responses, informing us that the BREX mechanism is contextual and complex. The scale of BREX response does not correlate to number of BREX motifs. Different BREX systems can respond very differently to the same BREX-susceptible phage. BREX systems can either provide differing levels of protection, or successfully target a phage resistant to another BREX homologue, though sharing similar numbers of cognate BREX motifs.

Discussion

Here, we have successfully sequenced and phylogenetically categorised 12 *Escherichia coli* phages that were isolated for use in characterisation of phage defence systems such as BREX. We then examined the morphological and genetic differences between each of our phages, and between our phages and wider *E. coli* phages. We discovered that phages Mak and Bam are the first representatives of a new putative genus "Palatinivirus", and that Mak and Bam are close relatives of a novel species within this genus. Using the obtained genome sequences, we were able to map the distributions of motifs targeted by three different BREX systems and examine the response of each characterised phage against each BREX system.

Phylogenetic analyses demonstrated that the selected Durham phages represented several different families of previously described *Escherichia* phages, with CS16 and Mav, Bam and Mak, and Trib and Baz clustering in pairs (**Fig. 1**). It also showed that within one year there is not necessarily a large turnover of phages within the natural water systems of Durham, as markedly similar phages were isolated a year apart, and from water sources separated by distances of 1-2 kilometres (for instance Pau and PATM, or Jura, SPSP and AL25; see **Table S2**).

Recent studies have found bacteriophages that could not be classified at the genus level (27, 54, 55). Similarly, we were unable to categorise Mak and Bam at the genus level within the currently accepted taxonomy (45, 56). Whilst we identified that Bam and Mak are likely members of the subfamily *Studiervirinae*, they cluster away from other members of the *Chatterjeevirus* genus (**Fig. 6**). As per the guidelines for establishing new genera and species (40, 41), and recent precedent (57), we propose that Mak and Bam represent the founding members of a novel genus, "Palatinivirus".

We expected that the majority of phages encoding BREX motifs would be susceptible to the cognate BREX system. However, this was not the case (**Figs. 7B and 7C**), nor, as previously suggested (58, 59), did the number of BREX sites correlate with impact on EOP by BREX. We therefore explored the possibility that some of our phages may encode anti-BREX or other anti-defence proteins. Ocr is a known T7 phage DNA-mimic protein, that inhibits host restriction enzyme-based defence mechanisms, and BREX, allowing T7 to infect cells that would otherwise be resistant (53). We identified ten Ocr homologues using BLAST-P (**Table S9**) and used this mini-database to search for homologues within our phage genomes. We also looked for homologues of DarB, which provides resistance against type I restriction enzymes (60), in our phage genomes. None of our genomes encoded contained obvious Ocr or DarB-like proteins, nor any of the other anti-defence genes that were used for searches (**Table S9**). This search was, however, by no means exhaustive and further anti-BREX and anti-defence proteins will no doubt be identified. Having established this BREX response dataset, it will now be possible to explore the Durham collection for potential anti-defence genes.

Curated phage collections can provide a stable basis for long-term comparisons of experimental data and a platform for discovery (27). This initial deposition of phage genome data marks the beginning of the Durham phage collection. It was produced as part of our undergraduate teaching, and we hope to expand the program in future, mimicking previously established and highly successful public engagement programs such as SEA-PHAGES (39). Developing and expanding the Durham collection will provide insight into the local phage ecology, whilst allowing deeper understanding of phage-host interactions such as identification of anti-BREX factors and determination of the BREX mechanism of phage defence.

Materials and methods

Bacterial strains and culture conditions

E. coli strain DH5α was sourced from Invitrogen. All strains were grown at 37 °C, either on agar plates or shaking at 220 rpm for liquid cultures. Luria broth (LB) was used as the standard growth media for liquid cultures and was supplemented with 0.35% w/v or 1.5% w/v agar for semi-solid and solid agar plates, respectively. Growth was monitored using a spectrophotometer (WPA Biowave C08000) measuring optical density at 600 nm (OD_{600}). When necessary, growth media was supplemented with ampicillin (Ap, 100 µg/ml) chloramphenicol (Cm, 25 µg/ml), or kanamycin (Km, 50 µg/ml).

Use of environmental coliphages

The Durham collection of phages were initially harvested and isolated by undergraduates as previously described (22). More details are available in **Table S2**. To make lysates, 10 µl of phage dilution was mixed with 200 µl of *E. coli* DH5α overnight culture and mixed with 4 ml of sterile semi-solid "top" LB agar (0.35% agar) in a sterile plastic bijou. Samples were poured onto solid LB agar plates (1.5% agar) and incubated overnight at 37 °C. Plates showing a confluent lawn of plaques were chosen for lysate preparations and the semi-solid agar layer was scraped off into 3 ml of phage buffer. 500 µl of chloroform was added and samples were vigorously vortexed and incubated for 30 min at 4 °C. Samples were centrifuged at 4000 x g for 20 min at 4 °C and the supernatant was carefully transferred to a sterile glass bijou. 100 µl of chloroform was added and lysates were kept at 4°C for long term storage.

DNA isolation and manipulation

All genomic DNA extraction steps in this study were carried out using either a Zymo Miniprep Plus kit (Cambridge Biosciences) or a Monarch gDNA extraction kit (NEB). Primers for cloning protocols were designed using the Benchling cloning design software, available online (benchling.com). Primers were synthesised by IDT. More details of primers and plasmids are available in **Tables 2 and 3**, respectively.

For generation of wild-type of the pBREX$_{Sty}$, the BREX coding region was split into six fragments and primers were designed to amplify each using the Benchling assembly tool (www.benchling.com). Fragments were successfully amplified however, cloning attempts failed using these fragments directly. In attempt to improve efficiency, individual fragments were cloned into pUC19 donor plasmids. Fragments were cloned using a standard transformation method and plated on ampicillin plates for selection. Sequencing confirmed the successful cloning of fragments into donor plasmids. Full golden gate assembly was then attempted following the NEB protocol. Briefly, donor plasmids were added at 2:1 molar ratio to pGGA destination vector with assembly mix and GG reaction buffer. The reaction was cycled 30 times (37 °C, 5 min; 16 °C, 10 min) followed by a 55 °C, 10 min inactivation step. Resulting reaction mixes were transformed into DH5α and plated on Cm for selection. Sequencing confirmed correct assembly. The PARIS deletion was achieved by Gibson assembly of mutant fragments (**Table 2**).

Phage genomic DNA was extracted using phenol-chloroform extraction methods. Briefly, 450 µl of phage lysate was incubated with 4.5U DNase I (Thermo Scientific) and 2 µl 20mg/ml RNase I (NEB) at 37 ^{o}C for 30 min with shaking. Proteinase K was added (2.25 µl of 20 mg/ml stock (Thermo)) and 10% SDS to a final concentration of 0.5% before incubation at 37 ^{o}C 30 min. The sample was then

resuspended in 500 µl 24:24:1 (v/v) phenol:chloroform:isoamyl alcohol (PCI) before centrifugation at 4 °C 5 min 16,000 x *g*. The aqueous layer was transferred to a new tube, PCI treatment was repeated and the subsequent aqueous layer transferred to a new tube before treatment with 500 µl 24:1 (v/v) chloroform:isoamyl alcohol and centrifugation as above. The aqueous layer was transferred to a new tube and incubated with 45 µl 3 M sodium acetate pH 5.2 and 500 µl isopropanol for 15 min at room temperature. DNA was pelleted by centrifugation at 4 °C for 20 minutes at 16,000 x *g.* The pellet was washed twice with 70% ethanol before being left to soak overnight at 4 °C in EB (10 mM Tris-HCl 0.1 mM EDTA pH 8.5), TE (10 mM Tris-base 1 mM EDTA pH 8.0) or MilliQ dH_2O. The integrity of gDNA was assessed via 0.75% agarose gels and quantification was undertaken via NanoDrop One (ThermoFisher). In some cases, a Phage DNA Isolation kit (Norgen BioTek Corp) was used, as per manufacturer's instructions.

Electron microscopy

Charge-discharged copper grids were soaked in high-titre phage lysate for 45 sec before 3 successive washes in MilliQ followed by staining in 2% phosphotungstic acid. Excess stain was removed by blotting before imaging via transmission electron microscopy, using a Hitachi H7600 operated at 100 kV accelerating voltage equipped with a EMSIS Xarosa camera and assessing images using Radius.

Phage genome sequencing and annotation

Extracted phage gDNA was sequenced either via MinION Mk1C (Phages Alma and TB34), or by Illumina sequencing via MicrobesNG, using either the Illumina HiSeq (Baz, Pau and BB1) or Illumina Novaseq 6000 (remaining coliphages). Nanopore sequencing was carried out using a MinION Mk1C sequencer (Oxford Nanopore Ltd.) with R9.4.1 flow cells. Library preparation was carried out as per Oxford Nanopore protocols using the ligation sequencing kit (SQK-LSK109) and native genomic

barcoding expansion (EXP-NBD104). Following generation of raw sequencing data, basecalling was performed by the Guppy basecalling package (Oxford Nanopore Ltd.) using the high accuracy (hac) basecalling model either during sequencing or post sequencing and data was deconvoluted using the ont_fast5_api package (https://github.com/nanoporetech/ont_fast5_api). Phage genomes were then assembled *de novo* using the "Assembly tutorial" Jupiter notebooks workflow provided by EPI2ME Labs (https://labs.epi2me.io). Briefly, initial assembly was performed using Flye 2.8.1-b1676, with –genome-size set to 0.1m and –asm-coverage set to 100x to account for high coverage values (https://github.com/fenderglass/Flye), and final assemblies were polished with medaka 1.5.0 using medaka_consensus (https://github.com/nanoporetech/medaka). Optional settings were left as default, unless otherwise stated. Those phage genomes sequenced by Illumina sequencing were processed by MicrobesNG. MicrobesNG genomic DNA libraries were prepared using the Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: input DNA was increased 2-fold, and PCR elongation time was increased to 45 s. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland). Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina. Libraries were sequenced using Illumina sequencers (HiSeq/NovaSeq) and a 250 bp paired end protocol. Reads were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cutoff of Q15 (61). *De novo* assembly was performed on samples using SPAdes version 3.7 (62) with default settings, and contigs were annotated using Prokka 1.11 (63).

Genome assemblies were assessed using the Centre for Phage Technology's (CPT) Galaxy interface (64). Open reading frames, tRNA genes and terminators were identified using CPT's publicly available structural workflows (versions PAP 2017 Structural V8.9 to PAP structural v2022.02) available from https://cpt.tamu.edu/galaxy-pub/u/cpt-workflows/. Each ORF was then assessed via BLAST searches

using CPT's Functional workflows (versions v2020.02 to v2022.01) available from https://cpt.tamu.edu/galaxy-pub/u/cpt-workflows/. Hypothetical protein identities were assigned according to CDS coverage and E-values from either the workflow results or BLAST-P searches against the non-redundant database using default settings. For predicted proteins where no significant similarity was found, InterPro (65) was used to try and assign a protein family identity.

Annotated genomes were validated and submitted to ENA using the Webin command line submission interface (versions 3.0.1 through to 4.4.0). The annotated genome of phage BB1 was instead submitted to GenBank via Geneious (V2020.2.3) and the GenBank submission plug-in (Version 1.6.7). Accession numbers are available in **Table S2**.

For detection of BREX motifs in phage genomes we used a custom Python script to count motifs on the forward and reverse strands, available here: https://github.com/liampshaw/BREX-phage-motifs. Figures were plotted using R v4.1.2 and the following R packages: ggplot2 v3.3.6 and cowplot v1.1.1.

The circular representations of *Escherichia* phage vB_Eco_Mak and *Eschericihia* phage vB_Eco_Bam genomes were produced using Proksee (https://proksee.ca/), and manually edited in Inkscape 1.1.1.

Phage taxonomy determination and variant calling

VIPTree server (48) was used to determine the taxonomy of the isolated non-modified coliphages in the Durham collection, by comparing them with the closest relatives found using fastANI (**Table S5**). The Durham collection was first compared to a subset of *Caudoviricetes*. All (n=29,188) NCBI Genbank genomes within *Caudoviricetes* as of 1st March 2023 were extracted using the search term

514 "*Caudoviricetes*[Organism] AND srcdb_genbank[PROP] NOT wgs[PROP] NOT cellular
515 organisms[ORGN] NOT AC_000001:AC_999999[PACC]". This genome database was then reduced to
516 a representative set of 180 genomes based on fastANI scoring (**Table S6**). Mak and Bam were
517 compared to a subset of genomes belonging to the *Autographiviridae* family (**Table S7**), and
518 subsequently to subsets belonging to the *Studiervirinae* subfamily and the *Chatterjeevirus* genus. All
519 subsets were chosen using fastANI. VIPTree was run with the 'only query' option in all instances. The
520 phylogenetic relationship between the non-modified Durham coliphages was determined by
521 generating a subset tree on VIPTree. All obtained trees were annotated using iTOL (66). For those
522 phages that were very close at a taxonomic level, we performed variant calling by using Snippy
523 v4.6.0 (https://github.com/tseemann/snippy). Comparisons were performed using complete
524 genomes and the "--ctgs" option. Graphic representation of Snippy results was obtained by drawing
525 the reference genomes using GenomeDiagram from the biopython project (67).

526

527 Assessment of potential anti-defence homologues

528 BLAST-P, with default settings, was used to identify ten *Escherichia* phage T7 Ocr homologues
529 (UniprotKB: P03775.1), ten *Escherichia* phage T4 Arn homologues (Uniprot: KBP39510), two *Vibrio*
530 phage ICP1 OrbA homologues (NCBI: YP_004250966.1), ten *Erwinia* phage FBB1 homologues
531 (A0A868BQY3), ten *Bacillus* phage BSP38 Apyc homologues (Uniprot: A0A345MJY6) and nine
532 *Escherichia* phage HY01 IPI homologues (Uniprot: A0A0A0PSJ3) (**Table S8**). Representative amino
533 acid sequences, including those with high and low percentage coverage or identity across phages
534 infecting various bacteria were selected. Each mini- database was then used to probe each Durham
535 phage genome for homologues, using TBLASTX with default settings and genetic code 11. For DarA
536 homologue searches, *Escherichia phage P1* DarA protein (YP_006494.1) was used in a TBLASTN
537 search against the Durham phage genomes using default settings.

538

Efficiency of plating (EOP) assays

E. coli DH5α were transformed with pTRB507 (empty vector control for the *E. fergusonii* pEFER and *Salmonella* BREX systems)) (22), pBrxXL$_{Eferg}$ (BREX$_{Eferg}$) (22), pBrxXL$_{Sty}$-Δ*ariA*Δ*ariB* (BREX$_{Sty}$), pBR322 (empty vector control for *E. coli* BREX system) or pBREX-AL (BREX$_{Ecoli}$) (25). Serial dilutions of all phages were produced in phage buffer (10 mM Tris pH 7.4, 10 mM $MgSO_4$, 0.01% (v/v) gelatin). For experiments with BREX$_{Eferg}$ or BREX$_{Ecoli}$ 200 µl of overnight culture and 10 µl of phage dilution were added to 3 ml 0.35% LB-agar. For experiments with BREX$_{Sty}$ overnight cultures were used to inoculate a second overnight culture, then 400 µl of this culture was mixed with 10 µl of phage dilution before mixing with 3 ml of 0.35% LB-agar. Plates were incubated overnight at 37 ^{o}C before plaque forming units (pfu) were counted on each plate. The EOP was calculated by dividing the pfu of the test strain by the pfu of the parental strain. Data shown are the mean and the standard deviation of at least three biological and technical replicates.

Data Availability

Annotated phage genomes are available via Genbank via the accession numbers listed in **Table S1**. All other data needed to evaluate the conclusions in the paper are present in the paper and/or Supplementary Data. Scripts and data for bioinformatic analysis are available on github (https://github.com/liampshaw/BREX-phage-motifs).

Acknowledgements

This work was supported by a Lister Institute Prize Fellowship to A.K., D.M.P. and T.R.B., an Engineering and Physical Sciences Research Council Molecular Sciences for Medicine Centre for Doctoral Training studentship [grant number EP/S022791/1] to S.C.W., a Wellcome Trust Sir Henry Wellcome Fellowship [grant number 218622/Z/19/Z] to G.M., a Wellcome Trust Sir Henry Wellcome Postdoctoral Fellowship [grant number 220422/Z/20/Z] to L.P.S., a Biotechnology and Biological Sciences Research Council Newcastle-Liverpool-Durham Doctoral Training Partnership studentship [grant number BB/M011186/1] to D.M.P., a Royal Society Research Grant [grant number RGS\R1\201330] to T.R.B., a Royal Society International Exchange Grant [grant number IEC\R2\202085] to T.R.B., funding from the Biophysical Sciences Institute at Durham University to T.R.B., an RFBR grant Ko_A_(21-54-10001) to A.I., A.D., J.G. and K.S., an RSF grant (22-14-00004) to A.I., A.D., J.G. and K.S., and a grant from the Ministry of Science and Higher Education of the Russian Federation (075-10-2021-114) to A.I., A.D., J.G. and K.S. This work was supported in part by a Wellcome Trust Senior Investigator award [grant number 106914/Z/15/Z] to J.C.D.H. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. Illumina HiSeq sequencing at MicrobesNG was supported by the BBSRC (Grant number BB/L024209/1).

We thank Durham University undergraduates from the 2016 and 2017 Microbiology Workshop who aided in original isolation of the phages in this study, and helped with characterisation including imaging the phages by electron microscopy; Ruth Bentley, Zoe Wong, Felicity Liu, Jae Seon Hong, Peter Little, Christopher Mapp, Amy Fenwick, Emily Parker, Visitation Lesterpearson, Georgia Pittam, Willow Fox, Catherine Sutherland, Alex Bradbury, Alexis Beney, Benjamin Jones, Katherine Unthank, Matthew Clark, Sara Algebra, Jodie Wells, Phillip Grove, Anona Galbraith, Soufaine Taleb, Josh Harling-Lee, Orlagh Anderson, Bindhu Unni, Emily Puumala, Cameron Gaul, Daniel Bateson, David Berry, Lewis Goddard, Conall Jager, Stefan Kecojevic, Izaak Beck, William Hayes, Katherine Reid and Alice Martin. Thank you to Christine Richardson, Chieko Itakura and Martin Goldberg for assistance with acquiring EM images. Thank you to Northumbrian Water for allowing access and helping us to obtain effluent samples.

References

1. Lima-Mendez G, Toussaint A, Leplae R. 2007. Analysis of the phage sequence space: the benefit of structured information. Virology2007/05/08. 365:241–249.

2. Tock MR, Dryden DTF. 2005. The biology of restriction and anti-restriction. Curr Opin Microbiol. Elsevier Current Trends https://doi.org/10.1016/j.mib.2005.06.003.

3. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GP. 2009. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A2009/01/07. 106:894–899.

4. Blower TR, Fineran PC, Johnson MJ, Toth IK, Humphreys DP, Salmond GP. 2009. Mutagenesis and functional characterisation of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of Erwinia. J Bacteriol2009/07/28. 191:6029–6039.

5. Kelly A, Arrowsmith TJ, Went SC, Blower TR. 2023. Toxin–antitoxin systems as mediators of phage defence and the implications for abortive infection. Curr Opin Microbiol 73:102293.

6. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–12.

7. Bernheim A, Millman A, Ofir G, Meitav G, Avraham C, Shomar H, Rosenberg MM, Tal N, Melamed S, Amitai G, Sorek R. 2021. Prokaryotic viperins produce diverse antiviral molecules. Nature 589:120–124.

8. Lau RK, Ye Q, Birkholz EA, Berg KR, Patel L, Mathews IT, Watrous JD, Ego K, Whiteley AT, Lowey B, Mekalanos JJ, Kranzusch PJ, Jain M, Pogliano J, Corbett KD. 2020. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Mol Cell 77:723-733.e6.

9. Rousset F, Depardieu F, Miele S, Dowding J, Laval A-L, Lieberman E, Garry D, Rocha EPC, Bernheim A, Bikard D. 2022. Phages and their satellites encode hotspots of antiviral systems. Cell Host Microbe 30:740-753.e5.

10. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R. 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science (80-) 359:eaar4120.

11. Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT. 2022. A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. Nat Microbiol 7:1568–1579.

12. Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voichek M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R. 2020. Bacterial Retrons Function In Anti-Phage Defense. Cell 183:1551-1561.e12.

13. Mariano G, Blower TR. 2023. Conserved domains can be found across distinct phage defence systems. Mol Microbiol https://doi.org/10.1111/mmi.15047.

14. Picton DM, Harling-Lee JD, Duffner SJ, Went SC, Morgan RD, Hinton JCD, Blower TR. 2022. A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands. Nucleic Acids Res 50:5191–5207.

15. Luyten YA, Hausman DE, Young JC, Doyle LA, Higashi KM, Ubilla-Rodriguez NC, Lambert AR, Arroyo CS, Forsberg KJ, Morgan RD, Stoddard BL, Kaiser BK. 2022. Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system. Nucleic Acids Res 50:5171–5190.

16. Blankenchip CL, Nguyen J V, Lau RK, Ye Q, Gu Y, Corbett KD. 2022. Control of bacterial immune signaling by a WYL domain transcription factor. Nucleic Acids Res 50:5239–5250.

17. Hoskisson PA, Sumby P, Smith MCM. 2015. The phage growth limitation system in Streptomyces coelicolor A(3)2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity. Virology 477:100–109.

18. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J 34:169–83.

19. Machnicka MA, Kaminska KH, Dunin-Horkawicz S, Bujnicki JM. 2015. Phylogenomics and sequence-structure-function relationships in the GmrSD family of Type IV restriction enzymes. BMC Bioinformatics 16:336.

20. Makarova KS, Wolf YI, Koonin E V. 2013. Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res 41:4360–77.

21. Makarova KS, Wolf YI, Snir S, Koonin E V. 2011. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J Bacteriol 193:6039–56.

22. Picton DM, Luyten YA, Morgan RD, Nelson A, Smith DL, Dryden DTF, Hinton JCD, Blower TR. 2021. The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses. Nucleic Acids Res 49:11257–11273.

23. Beck IN, Picton DM, Blower TR. 2022. Crystal structure of the BREX phage defence protein BrxA. Curr Res Struct Biol 4:211–219.

24. Shen BW, Doyle LA, Werther R, Westburg AA, Bies DP, Walter SI, Luyten YA, Morgan RD, Stoddard BL, Kaiser BK. 2023. Structure, substrate binding and activity of a unique AAA+ protein: the BrxL phage restriction factor. Nucleic Acids Res https://doi.org/10.1093/nar/gkad083.

25. Gordeeva J, Morozova N, Sierro N, Isaev A, Sinkunas T, Tsvetkova K, Matlashov M, Truncaite

L, Morgan RD, Ivanov N V, Siksnys V, Zeng L, Severinov K. 2019. BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site. Nucleic Acids Res 47:253–265.

26. Zaworski J, Dagva O, Brandt J, Baum C, Ettwiller L, Fomenkov A, Raleigh EA. 2022. Reassembling a cannon in the DNA defense arsenal: Genetics of StySA, a BREX phage exclusion system in Salmonella lab strains. PLOS Genet 18:e1009943.

27. Maffei E, Shaidullina A, Burkolter M, Heyer Y, Estermann F, Druelle V, Sauer P, Willi L, Michaelis S, Hilbi H, Thaler DS, Harms A. 2021. Systematic exploration of Escherichia coli phage–host interactions with the BASEL phage collection. PLOS Biol 19:e3001424.

28. Blower TR, Short FL, Fineran PC, Salmond GP. 2012. Viral molecular mimicry circumvents abortive infection and suppresses bacterial suicide to make hosts permissive for replication. Bacteriophage2013/06/07. 2:234–238.

29. Blower TR, Evans TJ, Przybilski R, Fineran PC, Salmond GP. 2012. Viral Evasion of a Bacterial Suicide System by RNA-Based Molecular Mimicry Enables Infectious Altruism. PLoS Genet2012/10/31. 8:e1003023.

30. Blower TR, Chai R, Przybilski R, Chindhy S, Fang X, Kidman SE, Tan H, Luisi BF, Fineran PC, Salmond GPC. 2017. Evolution of Pectobacterium Bacteriophage ΦM1 To Escape Two Bifunctional Type III Toxin-Antitoxin and Abortive Infection Systems through Mutations in a Single Viral Gene. Appl Environ Microbiol 83:e03229-16.

31. Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brussow H. 2004. Phage-host interaction: an ecological perspective. J Bacteriol2004/06/04. 186:3677–3686.

32. Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69–114.

33. Endersen L, Coffey A. 2020. The use of bacteriophages for food safety. Curr Opin Food Sci 36:1–8.

34. Rogovski P, Cadamuro RD, da Silva R, de Souza EB, Bonatto C, Viancelli A, Michelon W, Elmahdy EM, Treichel H, Rodríguez-Lázaro D, Fongaro G. 2021. Uses of Bacteriophages as Bacterial Control Tools and Environmental Safety Indicators. Front Microbiol 12:793135.

35. Gutiérrez D, Rodríguez-Rubio L, Martínez B, Rodríguez A, García P. 2016. Bacteriophages as Weapons Against Bacterial Biofilms in the Food Industry. Front Microbiol 7:825.

36. Alfaleh MA, Alsaab HO, Mahmoud AB, Alkayyal AA, Jones ML, Mahler SM, Hashem AM. 2020. Phage Display Derived Monoclonal Antibodies: From Bench to Bedside. Front Immunol 11:1986.

37. Venturini C, Petrovic Fabijan A, Fajardo Lubian A, Barbirz S, Iredell J. 2022. Biological foundations of successful bacteriophage therapy. EMBO Mol Med 14:e12435.

38. Górski A, Międzybrodzki R, Węgrzyn G, Jończyk-Matysiak E, Borysowski J, Weber-Dąbrowska B. 2020. Phage therapy: Current status and perspectives. Med Res Rev 40:459–463.

39. Hanauer DI, Graham MJ, SEA-PHAGES B, Betancur L, Bobrownicki A, Cresawn SG, Garlena RA, Jacobs-Sera D, Kaufmann N, Pope WH, Russell DA, Jacobs WR, Sivanathan V, Asai DJ, Hatfull GF. 2017. An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. Proc Natl Acad Sci U S A 114:13531–13536.

40. Turner D, Adriaenssens EM, Tolstoy I, Kropinski AM. 2021. Phage Annotation Guide: Guidelines for Assembly and High-Quality Annotation. https://home.liebertpub.com/phage 2:170–182.

41. Turner D, Kropinski AM, Adriaenssens EM. 2021. A Roadmap for Genome-Based Phage

703 Taxonomy. Viruses 13:506.

704 42. Bartlau N, Wichels A, Krohne G, Adriaenssens EM, Heins A, Fuchs BM, Amann R, Moraru C.
705 2022. Highly diverse flavobacterial phages isolated from North Sea spring blooms. ISME J
706 16:555–568.

707 43. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses
708 the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol
709 39:578–585.

710 44. Roux S, Adriaenssens EM, Dutilh BE, Koonin E V, Kropinski AM, Krupovic M, Kuhn JH, Lavigne
711 R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR,
712 Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz
713 P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi
714 IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F,
715 Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B,
716 Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T,
717 Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloe-Fadrosh EA.
718 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nat Biotechnol
719 37:29–37.

720 45. Turner D, Kropinski AM, Adriaenssens EM. 2021. A Roadmap for Genome-Based Phage
721 Taxonomy. Viruses 13:506.

722 46. Ondov BD, Starrett GJ, Sappington A, Kostic A, Koren S, Buck CB, Phillippy AM. 2019. Mash
723 Screen: high-throughput sequence containment estimation for genome discovery. Genome
724 Biol 20:232.

725 47. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI
726 analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.

48. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. 2017. ViPTree: the viral proteomic tree server. Bioinformatics 33:2379–2380.

49. Winter DJ. 2017. rentrez: An R package for the NCBI eUtils API https://doi.org/10.7287/peerj.preprints.3179v2.

50. Bujak K, Decewicz P, Kitowicz M, Radlinska M. 2022. Characterization of Three Novel Virulent Aeromonas Phages Provides Insights into the Diversity of the Autographiviridae Family. Viruses 14.

51. Rangel-Pineros G, Millard A, Michniewski S, Scanlan D, Sirén K, Reyes A, Petersen B, Clokie MRJ, Sicheritz-Pontén T. 2021. From Trees to Clouds: PhageClouds for Fast Comparison of ~640,000 Phage Genomic Sequences and Host-Centric Visualization Using Genomic Network Graphs. PHAGE (New Rochelle, NY) 2:194–203.

52. Rodwell E V., Wenner N, Pulford C V., Cai Y, Bowers-Barnard A, Beckett A, Rigby J, Picton DM, Blower TR, Feasey NA, Hinton JCD, Perez-Sepulveda BM. 2021. Isolation and Characterisation of Bacteriophages with Activity against Invasive Non-Typhoidal Salmonella Causing Bloodstream Infection in Malawi. Viruses 2021, Vol 13, Page 478 13:478.

53. Isaev A, Drobiazko A, Sierro N, Gordeeva J, Yosef I, Qimron U, Ivanov N V, Severinov K. 2020. Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. Nucleic Acids Res 48:5397–5406.

54. Korf IHE, Meier-Kolthoff JP, Adriaenssens EM, Kropinski AM, Nimtz M, Rohde M, van Raaij MJ, Wittmann J. 2019. Still Something to Discover: Novel Insights intoEscherichia coli Phage Diversity and Taxonomy. Viruses 11.

55. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, Stekel DJ, Hobman J, Jones MA, Millard A. 2021. INfrastructure for a PHAge REference Database: Identification of Large-Scale

Biases in the Current Collection of Cultured Phage Genomes. PHAGE 2:214–223.

56. Simmonds P, Adriaenssens EM, Zerbini FM, Abrescia NGA, Aiewsakun P, Alfenas-Zerbini P, Bao Y, Barylski J, Drosten C, Duffy S, Duprex WP, Dutilh BE, Elena SF, García ML, Junglen S, Katzourakis A, Koonin E V., Krupovic M, Kuhn JH, Lambert AJ, Lefkowitz EJ, Łobocka M, Lood C, Mahony J, Meier-Kolthoff JP, Mushegian AR, Oksanen HM, Poranen MM, Reyes-Muñoz A, Robertson DL, Roux S, Rubino L, Sabanadzovic S, Siddell S, Skern T, Smith DB, Sullivan MB, Suzuki N, Turner D, Van Doorslaer K, Vandamme A-M, Varsani A, Vasilakis N. 2023. Four principles to establish a universal virus taxonomy. PLOS Biol 21:e3001922.

57. Kazantseva OA, Piligrimova EG, Shadrin AM. 2022. Novel Bacillus-Infecting Bacteriophage B13—The Founding Member of the Proposed New Genus Bunatrivirus. Viruses 14:2300.

58. Hui W, Zhang W, Kwok LY, Zhang H, Kong J, Suna T. 2019. A novel bacteriophage exclusion (BREX) system encoded by the pglX gene in Lactobacillus casei Zhang. Appl Environ Microbiol 85.

59. Hui W, Zhang W, Li J, Kwok LY, Zhang H, Kong J, Sun T. 2022. Functional analysis of the second methyltransferase in the bacteriophage exclusion system of Lactobacillus casei Zhang. J Dairy Sci 105:2049–2057.

60. Iida S, Streiff MB, Bickle TA, Arber W. 1987. Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. Virology 157:156–66.

61. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

62. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell

sequencing. J Comput Biol 19:455–77.

63. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

64. Ramsey J, Rasche H, Maughmer C, Criscione A, Mijalis E, Liu M, Hu JC, Young R, Gill JJ. 2020. Galaxy and Apollo as a biologist-friendly interface for high-quality cooperative phage genome annotation. PLOS Comput Biol 16:e1008214.

65. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. 2023. InterPro in 2022. Nucleic Acids Res 51:D418–D427.

66. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.

67. Pritchard L, White JA, Birch PRJ, Toth IK. 2006. GenomeDiagram: a python package for the visualization of large-scale genomic data. Bioinformatics 22:616–7.

Figure Legends

Fig. 1. The Durham collection phages represent diverse species. Phylogenetic analysis of the relatedness of the selected Durham phages shows the phages belong to multiple independent species. Mak and Bam could not be classified at this point. Scale represents the genomic distance scores (SG) calculated by tBLASTx.

Fig. 2. Electron microscopy of representative Durham phages. Transmission electron micrographs of representative phages from each group. TB34, Baz, Alma and Pau represent families with myoviral morphologies, having contractile tails. Baseplate and tail fibres are visible for TB34. BB1 samples showed great variation in phage particles, with either filled heads and short tails of commonly expected length or empty phage heads connected to vastly extended tails (**Fig. S2**). Jura and Mak have short tails and represent podoviral families, with tail fibres clearly visible for Jura. BB1, Mav and Sip represent groups with siphoviral morphology, having flexible tails. All Sip particles visualised had empty capsids. Scalebars represent 100 nm.

Fig. 3. Durham phages are spread throughout class *Caudoviricetes*. The Durham phages were compared to representative *Caudoviricetes* phages, the T phages and Lambda. The Durham phages are spread across the diversity of *Caudoviricetes*, with Trib and Baz, CS16 and Mav and Mak and Bam clustering together. Again, Mak and Bam were unlike any of the sampled representative genomes. The tree scale reported refers to the branch length metadata of the tree. Similarly, an internal scale of the tree, based on branch length values, is also shown. Coloured strips indicate the taxonomy of each branch to family level.

Fig. 4. Circular representation of the Mak genome (A), with predicted CDS annotations (B). The genome of Mak is approximately 40 Kb, with 47% GC content, and with no predicted tRNA genes. Predicted CDS were assigned a function based on similarity to homologues in the reference database. NSPP, No Significant Phage Protein.

Fig. 5. Mak and Bam are distinct from other *Autographviridae*. VIPTree was used to build a phylogenetic tree including Mak and Bam, based on protein homology with selected members of *Autographiviridae*. Mak and Bam cluster together, but away from other sub-families within *Autographiviridae*. An internal scale of the tree, based on branch length values, is also shown. Coloured strips indicate the taxonomy of each branch to sub-family level.

Fig. 6. Mak and Bam are distinct within their closest subfamily and Genus. VIPTree was used again to try and determine the relationship of Mak and Bam to other members of their expected (A) subfamily, *Studiervirinae*, and (B) genus, *Chatterjeevirus*. An internal scale of the tree, based on branch length values, is also shown.

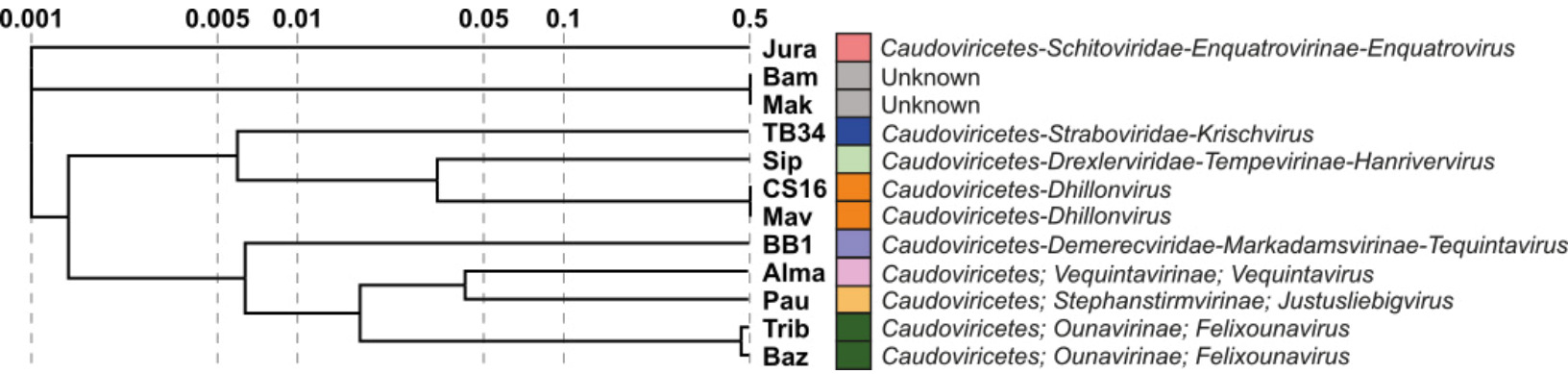
Fig. 7. Durham collection phages show diverse responses to BREX systems. (A) BREX loci used in this study. (B) Locations of the respective target BREX sites for $BREX_{Eferg}$ (GCTA**A**T), $BREX_{Ecoli}$ (GGTA**A**G) and $BREX_{Sty}$-Δ*ariA*Δ*ariB* (GATC**A**G) in the phage genomes, combining locations on the forward and reverse strands. For a plot showing separate strands, see **Fig. S4**. Numbers give the total number of hits for the respective motif in each phage genome. (C) EOPs of the 12 Durham phages in this study against the three BREX systems. For full data see **Table S7**.

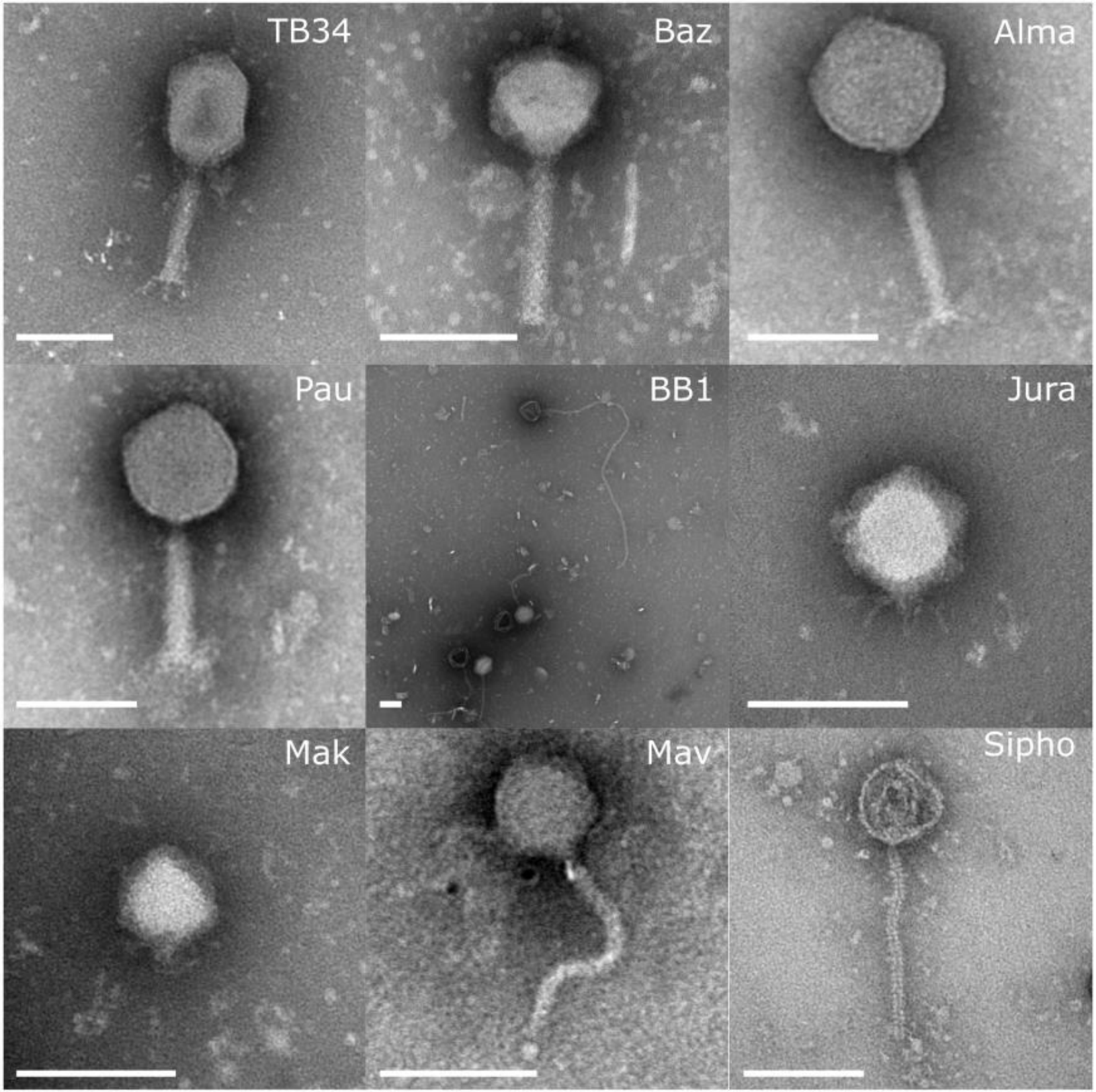
Table Legends

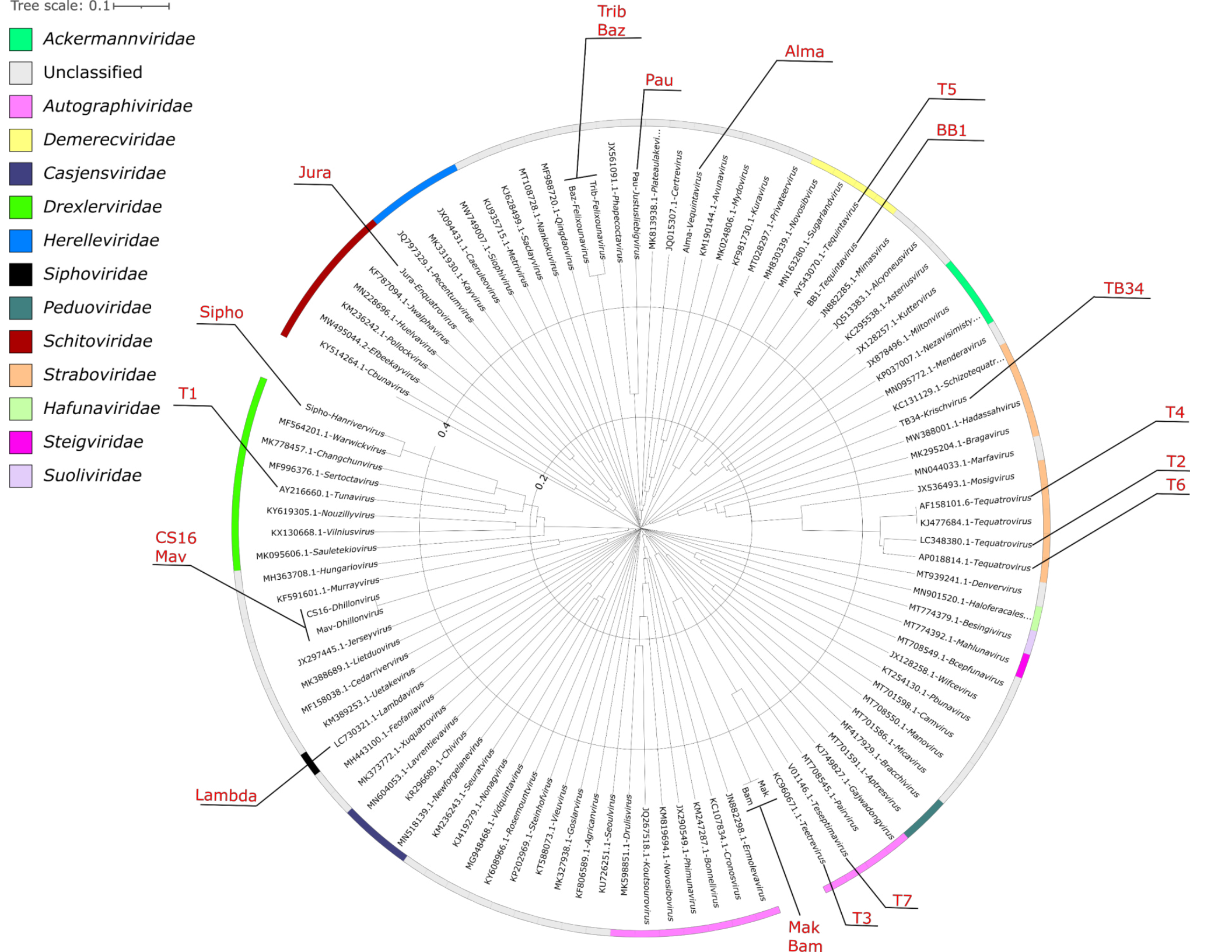
Table 1. Summary information for 12 Durham collection phages.

Table 2. Primers used in this study

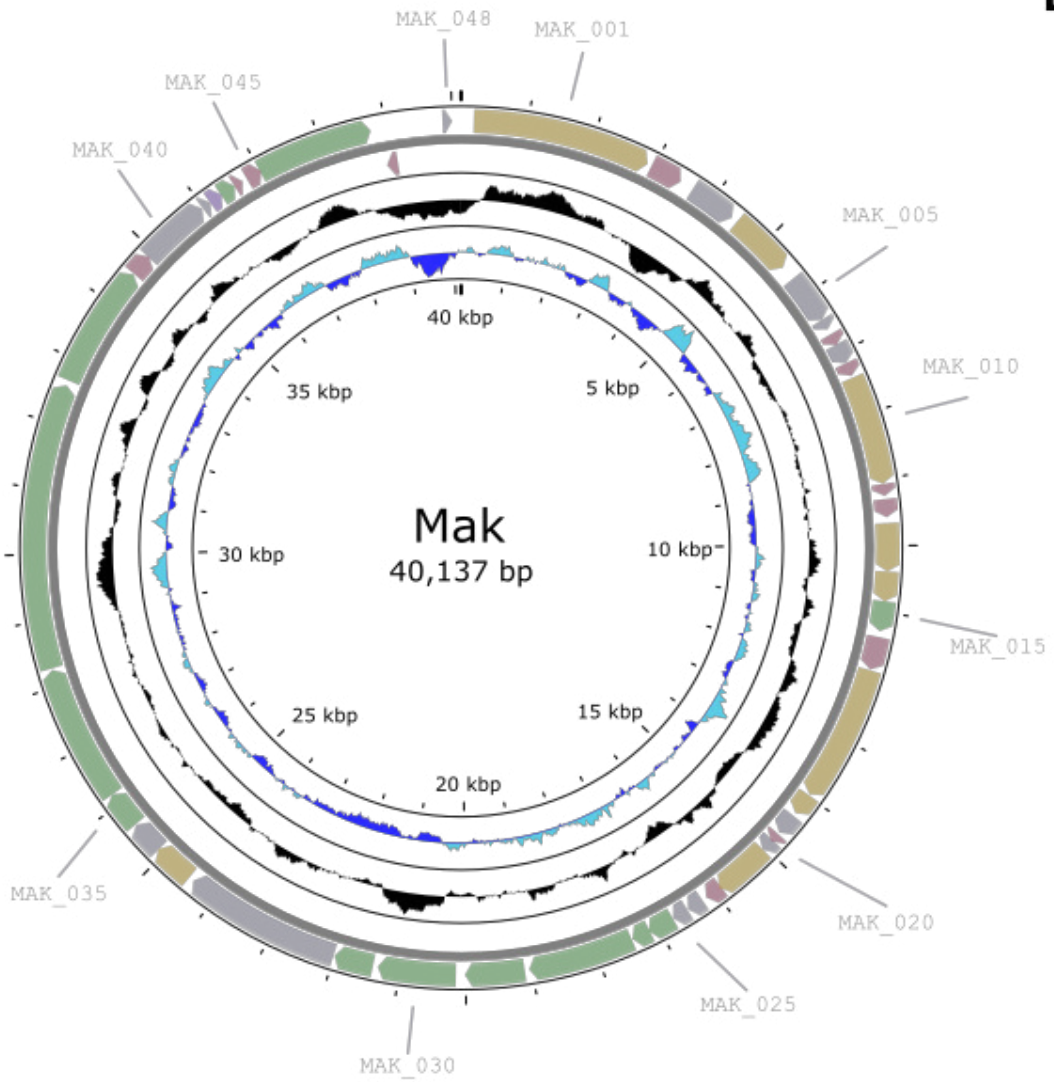
Table 3. Plasmids used in this study


0.001
0.005
0.01
0.05
0.1
0.5
Jura
Bam
Mak
TB34
Sip
CS16
Mav
BB1
Alma
Pau
Trib
Baz
Caudoviricetes-Schitoviridae-Enquatrovirinae-Enquatrovirus
Unknown
Unknown
Caudoviricetes-Straboviridae-Krischvirus
Caudoviricetes-Drexlerviridae-Tempevirinae-Hanrivervirus
Caudoviricetes-Dhillonvirus
Caudoviricetes-Dhillonvirus
Caudoviricetes-Demerecviridae-Markadamsvirinae-Tequintavirus
Caudoviricetes; Vequintavirinae; Vequintavirus
Caudoviricetes; Stephanstirmvirinae; Justusliebigvirus
Caudoviricetes; Ounavirinae; Felixounavirus
Caudoviricetes; Ounavirinae; Felixounavirus


TB34
Baz
Alma
Pau
BB1
Jura
Mak
Mav
Sipho


Tree scale: 0.1
Ackermannviridae
Unclassified
Autographiviridae
Demerecviridae
Casjensviridae
Drexlerviridae
Herelleviridae
Siphoviridae
Peduoviridae
Schitoviridae
Straboviridae
Hafunaviridae
Steigviridae
Suoliviridae
Trib
Baz
Pau
Alma
T5
BB1
Jura
Sipho
T1
TB34
T4
T2
T6
CS16
Mav
Lambda
T7
T3
Mak
Bam
0.4
0.2
Trib-Baz-Felixounavirus
Pau-Justusliebigvirus
Alma-Vequintavirus
BB1-Tequintavirus
Jura-Enquatrovirus
Sipho-Hanrivervirus
T34-Krischvirus
Tequatrovirus
Mak
Bam

A



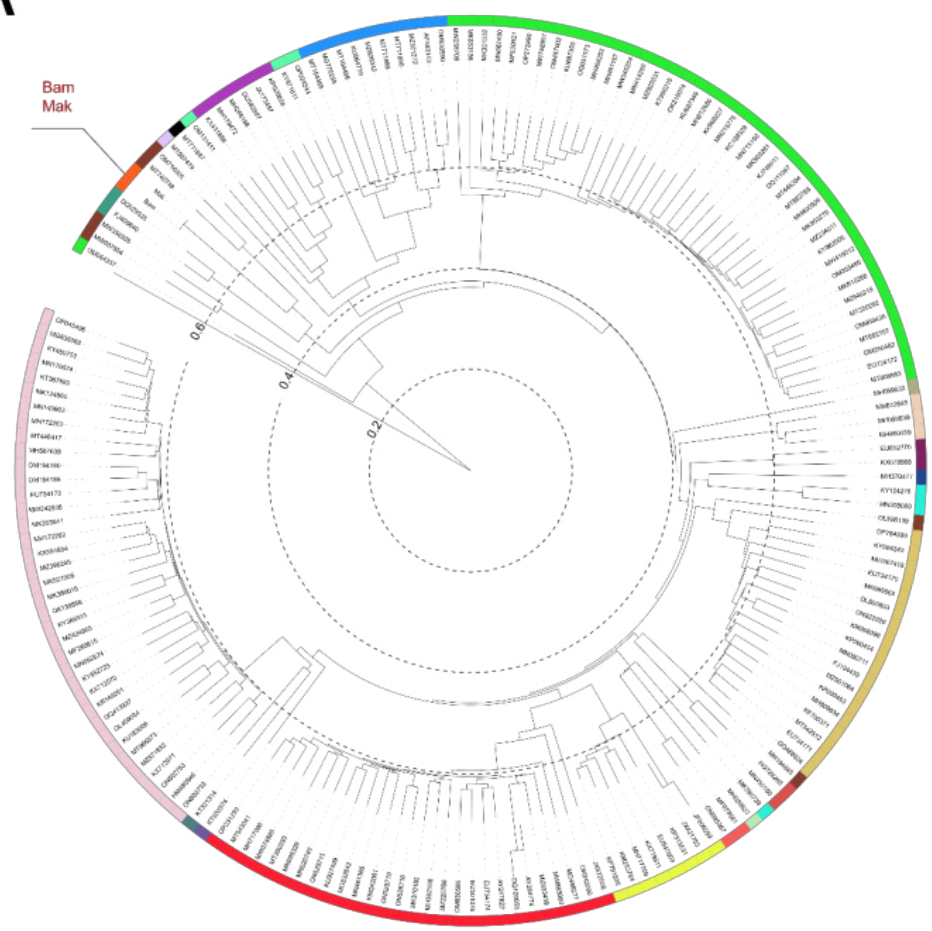
GC Content
GC Skew+
GC Skew-
Structural
Nucleic acid metabolism
Holin
Unknown function
Unknown function-NSPP

B

CDS	Proposed identity	Categorised as	Notes
Mak_001	T3/ T7 like RNA polymerase	Nucleic acid metabolism	
Mak_002	Hypothetical protein	Unknown function	NSPP
Mak_003	Hypothetical protein	Unknown function	
Mak_004	ATP-dependent ligase	Nucleic acid metabolism	
Mak_005	Hypothetical protein	Unknown function	
Mak_006	Hypothetical protein	Unknown function	
Mak_007	Hypothetical protein	Unknown function	
Mak_008	Hypothetical protein	Unknown function	
Mak_009	Hypothetical protein	Unknown function	NSPP
Mak_010	primase/helicase protein	Nucleic acid metabolism	
Mak_011	Hypothetical protein	Unknown function	NSPP
Mak_012	Hypothetical protein	Unknown function	NSPP
Mak_013	ssDNA-binding protein	Nucleic acid metabolism	
Mak_014	Putative endonuclease	Nucleic acid metabolism	
Mak_015	Lysozyme	Structural	
Mak_016	Hypothetical protein	Unknown function	NSPP
Mak_017	DNA polymerase	Nucleic acid metabolism	
Mak_018	Endonuclease	Nucleic acid metabolism	
Mak_019	Hypothetical protein	Unknown function	
Mak_020	Hypothetical protein	Unknown function	NSPP
Mak_021	Hypothetical protein	Unknown function	
Mak_022	Exonuclease	Nucleic acid metabolism	
Mak_023	Hypothetical protein	Unknown function	NSPP
Mak_024	Hypothetical protein	Unknown function	
Mak_025	Hypothetical protein	Unknown function	
Mak_026	Putative host specificity protein A	Unknown function	
Mak_027	Putative tail assembly protein	Structural	
Mak_028	Collar protein	Structural	
Mak_029	Caspid assembly protein	Structural	
Mak_030	Major caspid protein	Structural	
Mak_031	Tail tubular protein A	Structural	
Mak_032	Tail tubular protein B	Structural	
Mak_033	Putative deoxyribonucleoside kinase	Nucleic acid metabolism	
Mak_034	Hypothetical protein	Unknown function	
Mak_035	Internal virion protein B	Structural	
Mak_036	Internal virion protein C	Structural	
Mak_037	Internal virion protein D	Structural	
Mak_038	Tail fiber protein	Structural	
Mak_039	Hypothetical protein	Unknown function	NSPP
Mak_040	Hypothetical protein	Unknown function	
Mak_041	Hypothetical protein	Unknown function	
Mak_042	Type II holin	Holin	
Mak_043	Predicted DNA packaging protein A or DNA maturase A	Structural	
Mak_044	Hypothetical protein	Unknown function	NSPP
Mak_045	Hypothetical protein	Unknown function	NSPP
Mak_046	DNA Packaging protein B	Structural	
Mak_047	Hypothetical protein	Unknown function	NSPP
Mak_048	Hypothetical protein	Unknown function	

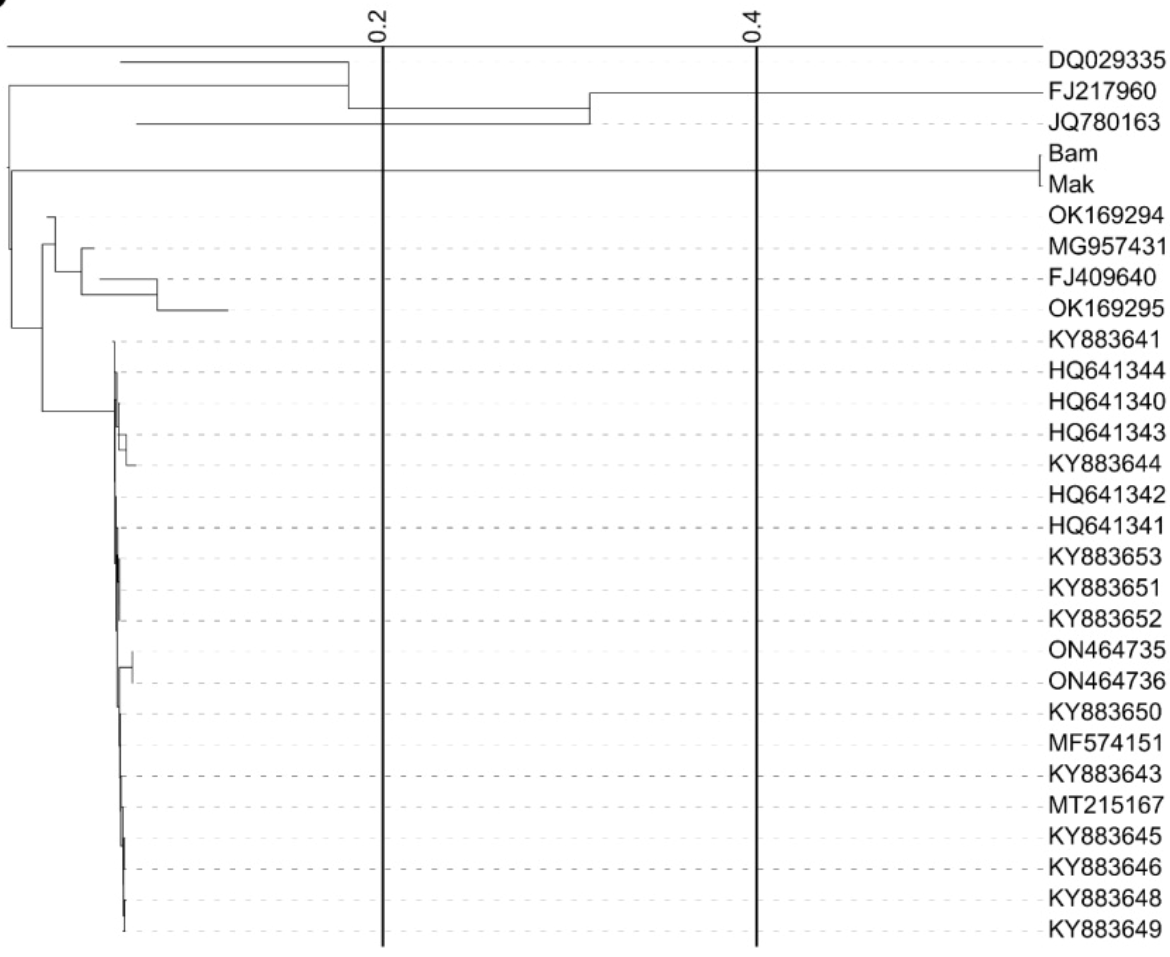

Tree scale: 0.1
Unclassified
Okabevirinae
Beijerinckvirinae
Melnykvirinae
Studiervirinae
Colwellvirinae
Molineuxvirinae
Slopekvirinae
Corkvirinae
Krylovirinae
Mak
Bam
0.4
0.2
MF621978.1-Lullwatervirus
KT381879.1-Percyvirus
MH684921.1-Scottvirus
KX660669.1-Aqualcavirus
KF626667.1-Pradovirus
KY981272.1-Kalppathivirus
MT740728.1-Anchaingvirus
KY316062.1-Sukuvirus
KY464836.1-Ampunavirus
MG652450.1-Higashivirus
KC853746.1-Mguuvirus
MG775261.1-Pollyceevirus
JN991020.1-Bifseptvirus
KX066068.1-Napahaivirus
JN882298.1-Ermolevavirus
KX756572.1-Wanjuvirus
KC107834.1-Cronosvirus
KY318515.1-Pokrovskaiavirus
MG878892.2-Panjvirus
KT184661.1-Aghbyvirus
JN651747.1-Aerosvirus
KT949345.1-Ahphunavirus
MF403005.1-Atuphduovirus
JX483874.1-Cuernavacavirus
KX397280.1-Kafavirus
MG592574.1-Cyclitvirus
MF278336.1-Foturvirus
KY065149.1-Tawavirus
MK562503.1-Bucovirus
GQ413938.2-Drullisvirus
JQ267518.1-Koutsourovirus
MK387869.1-Kakivirus
KF319020.1-Novosibovirus
FJ896200.1-Maculvirus
KX278418.2-Kotilavirus
KT240186.1-Stompvirus
JX290549.1-Phimunavirus
KU310944.1-Kantovirus
KJ473423.1-Daemvirus
HQ186308.1-Friunavirus
KF669656.1-Pettyvirus
MG845683.1-Kirikabuvirus
EU056923.1-Phikmvvirus
MW627366.1-Tunggulvirus
KY117485.1-Ayakvirus
KM247287.1-Bonnellivirus
KJ502657.1-Gutovirus
KF322026.1-Kaohsiungvirus
MG545917.1-Trungvirus
KY626176.1-Ampunavirus
MG018929.2-Uliginvirus
KJ749827.1-Gajwadongvirus
KM819695.1-Acadevirus
AY288927.2-Zindervirus
KM370384.1-Tuodvirus
JQ837901.1-Axomammavirus
EF160123.1-Eracentumvirus
KT852574.1-Apdecimavirus
EU547803.1-Teetrevirus
EU734173.1-Przondovirus
AY247822.1-Teseptimavirus
KT321314.1-Eapunavirus
EU734170.1-Berlinvirus
MH059637.1-Jarilovirus
MF979561.1-Unyawovirus
HQ728265.1-Elunavirus
MH059639.1-Ningirsuvirus
EU652770.3-Minipunavirus
KY124276.3-Pektosvirus
MH370477.2-Foetvirus
MH059632.2-Aarhusvirus
DQ111067.1-Kayfunavirus
MF034659.1-Wuhanvirus
MT711887.2-Troedvirus
GU583987.1-Pifdecavirus
KY971611.1-Pijolavirus
AF493143.1-Ghunavirus
MT597419.1-Warsawvirus
DQ029335.1-Chatterjeevirus
MF893341.1-Stompelvirus
KP343639.1-Serkorvirus
KR153873.1-Piedvirus
MT375523.1-Fussvirus
KC465900.1-Stopavirus
KT997876.1-Stupnyavirus
KT997847.1-Chosvirus
GU943031.1-Jalkavirus
MH598801.1-Votkovvirus
KC465901.1-Pelagivirus
MF979559.1-Gyeongsanvirus
JX483873.1-Paadamvirus
KM199771.1-Pairvirus
AY939843.2-Tiamatvirus
GU943068.1-Pagavirus
GU943073.1-Powvirus
GU943054.1-Pekhitvirus
HQ317388.1-Kajamvirus
MT424636.1-Ashivirus
HQ317389.1-Sednavirus
MK016662.1-Qadamvirus
HM559717.1-Poseidonvirus
HQ633062.1-Linvirus
GU071102.1-Cheungvirus
GU943065.1-Podivirus
GU071100.1-Tangaroavirus
HQ316584.1-Banchanvirus
EF372997.1-Voetvirus
AF338467.2-Tiilvirus
KC310806.1-Kembevirus
KJ183191.1-Daemvirus

A


Bam
Mak
0.2
0.4
0.6


Legend
Aarhusvirus
Apdecimavirus
Berlinvirus
Chatterjeevirus
Eapunavirus
Elunavirus
Foetvirus
Ghunavirus
Jarilovirus
Kayfunavirus
Minipunavirus
New Genus
Ningirsuvirus
Pektosvirus
Pifdecavirus
Pijolavirus
Przondovirus
Teetrevirus
Teseptimavirus
Troedvirus
Unclassified
Unyawovirus
Warsawvirus
Tree scale: 0.1

B


0.2
0.4
DQ029335
FJ217960
JQ780163
Bam
Mak
OK169294
MG957431
FJ409640
OK169295
KY883641
HQ641344
HQ641340
HQ641343
KY883644
HQ641342
HQ641341
KY883653
KY883651
KY883652
ON464735
ON464736
KY883650
MF574151
KY883643
MT215167
KY883645
KY883646
KY883648
KY883649
Tree scale: 0.1

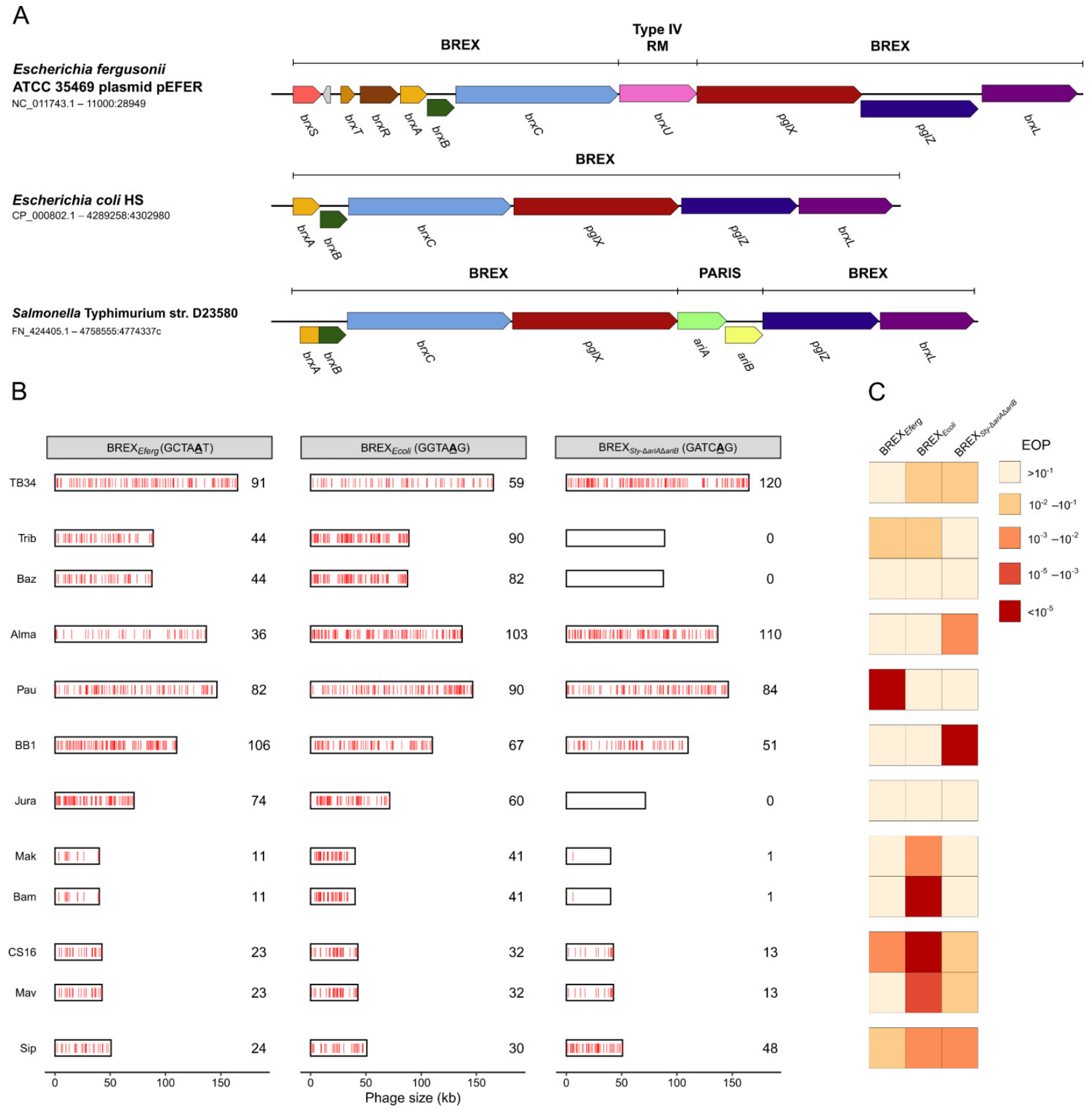

A
Escherichia fergusonii
ATCC 35469 plasmid pEFER
NC_011743.1 – 11000:28949
BREX
Type IV
RM
BREX
brxS
brxT
brxR
brxA
brxB
brxC
brxU
pglX
pglZ
brxL
Escherichia coli HS
CP_000802.1 – 4289258:4302980
BREX
brxA
brxB
brxC
pglX
pglZ
brxL
Salmonella Typhimurium str. D23580
FN_424405.1 – 4758555:4774337c
BREX
PARIS
BREX
brxA
brxB
brxC
pglX
ariA
ariB
pglZ
brxL
B
BREXEferg (GCTAAT)
BREXEcoli (GGTAAG)
BREXSty-ΔariAΔariB (GATCAG)
TB34 91 59 120
Trib 44 90 0
Baz 44 82 0
Alma 36 103 110
Pau 82 90 84
BB1 106 67 51
Jura 74 60 0
Mak 11 41 1
Bam 11 41 1
CS16 23 32 13
Mav 23 32 13
Sip 24 30 48
0 50 100 150
Phage size (kb)
C
BREXEferg
BREXEcoli
BREXSty-ΔariAΔariB
EOP
>10-1
10-2 –10-1
10-3 –10-2
10-5 –10-3
<10-5

Table 1. Summary information for 12 Durham collection phages.

Phage	AKA	Genome size (kb)	# ORFs	# tRNAs	% GC	Taxonomic group	GenBank Accession	Closest relative	% coverage	% identity	Genbank
Escherichia phage vB_Eco_TB34	TB34	165.22	288	0	40.5	*Krischvirus*	OX001802.1	*Escherichia* phage W115	96	98.31	ON286974.1
Escherichia phage vB_Eco_Tribble	Trib	88.85	129+4	26	39.0	*Felixonunavirus*	OX016465.1	*Escherichia* phage JLBYU32	93	97.12	OK272490.1
Escherichia phage Barry	Baz	87.69	126	26	38.9	*Felixonunavirus*	LR880803.1	*Escherichia* phage Warpig	95	97.51	MN850637.1
Escherichia phage vB_Eco_Alma	Alma	136.98	229	7	43.5	*Vequintavirus*	OV101294.1	*Escherichia* phage EdH4	98	97.62	MK327930.1
Escherichia phage Paula	Pau	146.55	248	16	37.5	*Justusliebigvirus*	LR865361.1	*Escherichia* phage Inny	99	98.42	MN850601.1
Escherichia phage BB1	BB1	110.10	158	22	38.9	*Tequintavirus*	MT843274.1	*Salmonella* phage vB_StyS-LmqsSP1	87	96.92	MT577844.1
Escherichia phage vB_Eco_Jura	Jura	71.49	77	5	41.3	*Enquatrovirus*	LR999871.1	*Escherichia* phage PMBT57	99	99.05	MG770228.1
Escherichia phage vB_Eco_Mak	Mak	40.14	48	0	47.0	*Unknown*	OX001577.1	*Vibrio* phage Rostov-1	7	78.47	MG957431.1
Escherichia phage vB_Eco_Bam	Bam	40.14	54	0	47.0	*Unknown*	OW991346.1	*Vibrio* phage Rostov-1	7	78.47	MG957431.1
Escherichia phage vB_Eco_CS16	CS16	42.66	60	0	54.5	*Dhillonvirus*	LR999870.1	*Escherichia* phage Slur05	93	96.91	LN881730.1
Escherichia phage vB_Eco_Maverick	Mav	42.62	57	0	54.5	*Dhillonvirus*	LR990702.1	*Escherichia* phage Sponge	96	93.05	MW749005.1
Escherichia phage vB_Eco_Sip	Sip	50.81	87	0	44.6	*Warwickvirus*	OU734268.1	*Escherichia* phage Tiwna	96	98.04	NC_054896.1

Table 2. Primers used in this study.

Primer	Sequence	Notes
Golden-gate assembly of *S. enterica* serovar Typhimurium BREX into pGGA		
TRB1367	tttTCTAGAGGTCTCCGGAGAATGGTTCATCTGGCGCT	*S. enterica* serovar Typhimurium BREX frag1 FWD
TRB1368	tttTCTAGAGGTCTCCTAACGAGGATTCAATGTCGC	*S. enterica* serovar Typhimurium BREX frag1 REV
TRB1369	tttTCTAGAGGTCTCGGTTAAAGAGCACAGCAATGAATATTG	*S. enterica* serovar Typhimurium BREX frag2 FWD
TRB1370	tttTCTAGAGGTCTCGAGgCCAAAAATATTATTTTCCAGAattag	*S. enterica* serovar Typhimurium BREX frag2 REV
TRB1371	tttTCTAGAGGTCTCGGcCTCGACATTGACGACCGTGCT	*S. enterica* serovar Typhimurium BREX frag3 FWD
TRB1372	tttTCTAGAGGTCTCGTTCATTAAATAATCTCCGGTGCATTGCCG	*S. enterica* serovar Typhimurium BREX frag3 REV
TRB1373	tttTCTAGAGGTCTCGTGAAAATAGCGGCGATTTATAC	*S. enterica* serovar Typhimurium BREX frag4 FWD
TRB1374	tttTCTAGAGGTCTCGGGTCACAATCCCATTTCATTCCA	*S. enterica* serovar Typhimurium BREX frag4 REV
TRB1375	tttTCTAGAGGTCTCGGACCTTGCAAAATCAGGAATTTATT	*S. enterica* serovar Typhimurium BREX frag5 FWD
TRB1376	tttTCTAGAGGTCTCGCACTTAAAAGAAATCATCCTGAAATGC	*S. enterica* serovar Typhimurium BREX frag5 REV
TRB1377	tttTCTAGAGGTCTCCAGTGAGGCGCTATGCAAAC	*S. enterica* serovar Typhimurium BREX frag6 FWD
TRB1378	tttTCTAGAGGTCTCCATGGAGGGAAACCAGGGGTTAC	*S. enterica* serovar Typhimurium BREX frag6 REV
TRB1200	aCGCGGTATCATTGCAGCACTGG	“Round-the-horn" mutagenesis FWD to remove BsaI site in Amp gene of pUC19
TRB1201	GACCCACGCTCACCGGCTCCAG	“Round-the-horn" mutagenesis FWD to remove BsaI site in Amp gene of pUC19
Gibson assembly primers for creation of pBrxXL$_{Sty}$ PARIS knockout (KO)		
TRB1738	ATGCACCGGAGATTATTTAAATCTGGAATGAAATGGGATT	pBrxXL$_{Sty}$ PARIS Gibson assembly KO FWD (1)
TRB1739	TTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCG	pBrxXL$_{Sty}$ PARIS Gibson assembly KO REV (1)
TRB1740	CGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAA	pBrxXL$_{Sty}$ PARIS Gibson assembly KO FWD (2)
TRB1741	TTCTTCGTTGGTCAGAAACACGTATTTGTCTTCAACACGT	pBrxXL$_{Sty}$ PARIS Gibson assembly KO REV (2)
TRB1742	ACGTGTTGAAGACAAATACGTGTTTCTGACCAACGAAGAA	pBrxXL$_{Sty}$ PARIS Gibson assembly KO FWD (3)
TRB1743	AATCCCATTTCATTCCAGATTTAAATAATCTCCGGTGCAT	pBrxXL$_{Sty}$ PARIS Gibson assembly KO REV (3)
Sequencing primers (ranges indicate position of primer in *Salmonella* BREX coding region)		
TRB710	CGTTACCTGGAACCATTCGT	152-171
TRB711	CCCTATGGATAGCTGGGATG	865-884
TRB712	GCAGGACGTGATGGGTTTTA	1551-1570
TRB713	GCCAATACGACGCGTTTAAG	2266-2285
TRB714	GTCTATCCGGACCAAAGGTG	2957-2976

TRB715	CGGCTGCATTTTAATTCGTT	3659-3678
TRB716	GCACAAACTATGGCGGAAAT	4362-4381
TRB717	ACGGATGCCGAGAAGAAGAT	5058-5077
TRB718	GATAACCCGACAGGCTTTGA	5751-5770
TRB719	GTCTCGACATTGACGACCG	6463-6481
TRB720	ATTACGATGGCACATTTGGG	7155-7174
TRB721	GGATTAATGTGCACTCCGGT	7863-7882
TRB722	AAATCTCGAATTTATCGCCG	8567-8586
TRB723	ATTGGCTGGGCACGGGTA	9261-9278
TRB724	GGCGGTGTTGTACTCATTGAT	9968-9988
TRB725	CCGGTTTTAATACTGCGTTTC	10651-10671
TRB726	CTTGAAAGGCCTGGTCACTG	11368-11387
TRB727	TTGCTGAATCTGCGTAATCG	12056-12075
TRB728	GCATAACACCATTGATGCCA	12751-12770
TRB729	CGAATTTCAATCGCCGTAAT	13461-13480
TRB730	GTATTTACCGCACGTACGCA	14154-14173
TRB731	AACCAGCGCGACGTTATC	14835-14870
TRB732	CGCGGTAGATATTCCGACTG	15566-15585

Table 3. Plasmids used in this study.

Plasmid	Notes	Primers used	Reference
pBrxXL$_{Sty}$	Full *S. enterica* serovar Typhimurium coding region in a pGGA vector backbone, created by golden gate assembly from genomic *S. enterica* serovar Typhimurium D23580 DNA.	TRB1367-TRB1378	This study
pBrxXL$_{Sty}$-*ΔariAΔariB*	Created by Gibson assembly from pBrxXL$_{Sty.}$	TRB1748-TRB1743	This study
pBrxXL-AL	*E. coli* BREX		Gordeeva *et al.* 2019
pBrxXL$_{Eferg}$	*E. fergusonii* BREX		Picton *et al.* 2021
pUC19	pUC19 with BsaI sites mutated out via “round-the-horn" mutagenesis	TRB1200, TRB1201	NEB
pTRB507	pGGA plasmid backbone containing 12400 – 14394 of the pEFER plasmid from *E. fergusonii,* used as a negative control.		Picton *et al.* 2021