



**HAL**  
open science

## **A Pair of DNA Glucosyltransferases Elevate Counter-defense in Bacteriophage T4**

Luis Ramirez-Chamorro, Frédéric Bonhomme, Anton Lukas Ipsen Wolff, François Lecointe, Marcel Hollenstein, Mart Krupovic, Marianne De Paepe, Yuvaraj Bhoobalan-Chitty

### ► To cite this version:

Luis Ramirez-Chamorro, Frédéric Bonhomme, Anton Lukas Ipsen Wolff, François Lecointe, Marcel Hollenstein, et al.. A Pair of DNA Glucosyltransferases Elevate Counter-defense in Bacteriophage T4. 2026. <pasteur-05621909>

**HAL Id: pasteur-05621909**

**<https://pasteur.hal.science/pasteur-05621909v1>**

Preprint submitted on 13 May 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# A Pair of DNA Glucosyltransferases Elevate Counter-defense in Bacteriophage T4

Luis Ramirez-Chamorro<sup>1</sup>, Frédéric Bonhomme<sup>2</sup>, Anton Lukas Ipsen Wolff<sup>3</sup>, François Lecointe<sup>1</sup>, Marcel Hollenstein<sup>4</sup>, Mart Krupovic<sup>5</sup>, Marianne De Paepe<sup>1</sup>, Yuvaraj Bhoobalan-Chitty<sup>1,3,\*</sup>

<sup>1</sup> Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350 Jouy-en-Josas, France

<sup>2</sup> Institut Pasteur, Université Paris Cité, CNRS UMR3523, Department of Structural Biology and Chemistry, Unité de Chimie Biologique Epigénétique, 28, rue du Docteur Roux, 75724 Paris Cedex 15, France

<sup>3</sup> Department of Biology, University of Copenhagen, Copenhagen N, Denmark

<sup>4</sup> Institut Pasteur, Université Paris Cité, CNRS UMR3523, Department of Structural Biology and Chemistry, Laboratory for Bioorganic Chemistry of Nucleic Acids, 28, rue du Docteur Roux, 75724 Paris Cedex 15, France

<sup>5</sup> Institut Pasteur, Université Paris Cité, Cell Biology and Virology of Archaea Unit, Paris, France

\*Address correspondence to Yuvaraj Bhoobalan-Chitty, [yuvarajb@bio.ku.dk](mailto:yuvarajb@bio.ku.dk)

## ABSTRACT

Bacteriophages encode diverse pathways to modify their nucleobases. These modifications help phages to evade the host defense systems such as restriction-modification (RM), and type II and type V CRISPR-Cas systems. On the other hand, modifications can also serve as a target for other host defense systems, illustrating the complexity of the defense and counter-defense landscape. Bacteriophage T4 encodes two glucosyltransferases (GTs),  $\alpha$ -GT and  $\beta$ -GT, that post-replicatively add a glucose moiety to the hydroxymethylated deoxycytosines (5-hmC) on phage DNA in the  $\alpha$ - and  $\beta$ -conformation, respectively. Among all known phages, only six closely related phages encode both  $\alpha$ -GT and  $\beta$ -GT. Here, through biochemical and genetic analysis, we show that  $\beta$ -GT has higher catalytic activity, whereas  $\alpha$ -GT is more strongly expressed. During the T4 infection, these factors determine the contributions of both GTs, with  $\alpha$ -GT and  $\beta$ -GT contributing respectively to glucosylation of 66% and 33% of all 5-hmC. Encoding a single GTs is sufficient for T4 to overcome the *E. coli* type I and type IV RM systems, unless the glucosylation capacity decreases below the 80% threshold. However, when encountering a host encoding DNA glycosylase Brig1 in addition to type I and type IV RM systems, a second GT is necessary to enable Brig1 escapers to resist RM systems. These results demonstrate that encoding multiple GTs with redundant functionalities provides an evolutionary advantage when simultaneously confronted with multiple antiphage defense systems.

## INTRODUCTION

Bacteriophages, particularly members of the class *Caudoviricetes*, encode a variety of pre-replicative and post-replicative pathways to diversify their DNA nucleobase content beyond canonical composition<sup>1</sup>. Modifications range from methylation, the most prevalent modification among all domains of life, to complex modifications that result in hypermodified bases such as glucosylmethylcytosine (5-gmC), aminocarboxymethyladenine and queuosine-like 7-deazaguanine derivatives<sup>2,3</sup>. Methylation is a key component of the host-encoded RM systems that target viruses<sup>4</sup>. Similarly, hypermodification has been shown to be essential in phage T4 to evade host RM systems, as well as type II and type V CRISPR-Cas systems<sup>5-10</sup>. Additionally, modifications enable recognition of non-phage (host) DNA by phage proteins, thereby contributing to transcriptional repression<sup>11</sup> and cleavage of host DNA by phage endonucleases<sup>12,13</sup>.

Hypermodified DNA bases, in particular 5-gmC, were first identified among T-even phages<sup>14</sup>. Phage-encoded enzymes first modify the host nucleotide precursor pool, converting deoxycytidine diphosphate (dCDP) and deoxycytidine monophosphate (dCMP) into hydroxymethylated deoxycytidine triphosphate nucleotides (5-hmdC)<sup>15</sup>. During phage replication, the modified nucleotides are incorporated into the nascent DNA. Subsequently, the glucosyl group from uridine diphosphate glucose (UDP-glucose) is transferred onto the hydroxymethylated cytosine (5-hmC) by glucosyltransferase enzymes (GTs)<sup>16</sup>. In phage T4, all cytosines in DNA are exclusively mono-glucosylated by primary glucosyltransferases, either alpha-glucosyltransferase ( $\alpha$ -GT) or beta-glucosyltransferase ( $\beta$ -GT)<sup>17,18</sup>. Among these, 60% possess  $\alpha$ -glycosidic linkages (catalysed by  $\alpha$ -GT) while 40% have  $\beta$ -glycosidic linkages (catalysed by  $\beta$ -GT)<sup>19</sup>. In T4, it has been suggested that glucosylation by  $\alpha$ -GT is concomitant with replication and occurs upon association of the enzyme with the T4 replisome, whereas the remaining non-glucosylated 5-hmdC residues serve as substrates for  $\beta$ -GT<sup>20</sup>. In phages T2 and T6, that possess only the primary glucosyltransferase  $\alpha$ -GT,  $\beta$ -glucosyl-HMC- $\alpha$ -GT acts as a secondary glucosyltransferase, catalysing the addition of a glucosyl group onto an already  $\alpha$ -glucosylated cytosine, resulting in 5-gentiobiosyl-methylcytosine base<sup>17,18,21,22</sup>.

The diversification of modifications likely results from an evolutionary arms race between phages and their hosts. In the *E. coli*-T4 host-phage system, hydroxymethylation of cytosine protects the phage DNA from the widely conserved type I EcoKI RM system<sup>23</sup>. Subsequent glucosylation protects from the methylation specific type IV McrBC/McrA RM systems<sup>5,24</sup>. DNA glycosylases such as Brig1 and the prophage encoded type IV RM system GmrSD (glucose-modified restriction) are both capable of antiphage activity by targeting 5-gmC<sup>25-27</sup>. Brig1 acts specifically against  $\alpha$ -glucosylated cytosine, whereas GmrSD targets both stereoisomers of 5-gmC and the 5-gentiobiosyl-methylcytosine. The activity of GmrSD and its variants is inhibited by phage T4 proteins IPI, IPII and IPIII<sup>25,26,28</sup>. As an alternative to glucosylation, arabinosylation of the hydroxy deoxycytosine (5-hdC) in the phage DNA<sup>29,30</sup> also offers robust protection not only against type II RM and the DNA glycosylase Brig1<sup>31</sup>, but also against type I CRISPR-Cas. In addition to Brig1, antiviral glycosylases such as Dag1/Dag2 and Brig2, which cleave modified guanine and thymine bases, respectively, have also been characterized<sup>32</sup>.

Our study reveals that while most phages encode only one primary GT, a few closely related phages, including phage T4, encode both  $\alpha$ -GT and  $\beta$ -GT. The two phage GTs

appear to have diverged from a common cellular ancestor with the GT-B structural fold following a gene duplication event, resulting in paralogous enzymes that yield stereochemically different reaction products. We show that the balanced levels of  $\alpha$ - and  $\beta$ -glucosylation observed in wild-type phage is due to the opposing strengths of the two GTs, reflecting differences in both their enzymatic activity and gene expression. We demonstrate that each GT, when expressed alone, is sufficient to ensure more than 80% cytosine glucosylation, a level sufficient to provide complete protection against *E. coli* type IV restriction-modification systems. However, despite the apparent functional redundancy of the two GTs in protection against typical RM systems, we show that in the presence of the additional glycosylase-based defense system Brig1, the non-targeted GT serves as an efficient back-up allowing the survival of phage escaper mutants with inactivated target GT.

## RESULTS

### Co-occurrence of two primary glucosyltransferases is rare in phage genomes

Genetic redundancy is rare in phage genomes, due to a high constraint on maximal genome content<sup>33-35</sup>. Yet  $\beta$ -GT in phage T4 was previously described to have a redundant function with  $\alpha$ -GT, which is intriguing<sup>36</sup>. To investigate this phenomenon, we started by analysing the distribution of GTs (primary and secondary) in phage genomes. We first identified all viral homologs of T-even phage GTs in the RefSeq database by PSI-BLAST ([Table S1](#)). The hits were then associated to their respective phage genomes ([Figure 1A](#) and [Table S1](#)). Among the 111 phages encoding GT homologs, 56 phages carried two types of GT. In all cases, except phage T4 and Bacillus phage G, phage genomes contained a primary-GT alone or in association with a homolog of a secondary-GT. Bacillus phage G contains a second copy of  $\alpha$ -GT, whereas phage T4 is unique in encoding two primary GTs,  $\alpha$ -GT and  $\beta$ -GT.

To check if the co-occurrence of two primary-GTs in T4 is an exception, we expanded our analysis to all phage genomes in the non-redundant (nr) NCBI GenBank database. Among the 753 phages that carried at least one homolog of T-even phage GTs, only one encoded all three GT types, and five phages, including phage T4, encoded both primary GTs ([Figure 1B](#) and [Table S1](#)). The overall nucleotide identity between the six phage genomes ranged from 84% to 88%, with a coverage of 90-100% ([Figure 1C](#)), indicating that these phages are closely related and belong to the same *Tequatrovirus* genus. A whole-genome tree showed that the five phages encoding both primary GTs share a common ancestor ([Figure S1](#)). This analysis also confirmed that the vast majority of phage genomes carry a single primary GT together with a secondary GT. Surprisingly, several phages also encoded distantly related secondary GTs in the absence of any recognizable primary GT ([Figure S2](#)). Collectively, these findings indicate that the co-occurrence of  $\alpha$ -GT and  $\beta$ -GT is an uncommon event which may indicate either that it does not confer a selective advantage to the phage or that the acquisition of the second primary GT is a relatively recent event in phage evolution.

### Evolutionary relationship between $\alpha$ -GT, $\beta$ -GT and $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT

To investigate this hypothesis, we determined the evolutionary relationships between the three GT types. Homologs of the  $\alpha$ -GT,  $\beta$ -GT and  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT do not display any appreciable sequence similarity across the GT types ([Table S1](#)). However, structural comparison showed that  $\alpha$ -GT and  $\beta$ -GT adopt the same structural fold<sup>37,38</sup>, known as the GT-B<sup>39</sup> ([Figure 2](#)), suggesting that the two enzymes are evolutionarily related. By contrast, analysis of the predicted structural model of  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT from phage T6 generated with AlphaFold3 revealed that it belongs to an evolutionarily distinct superfamily of GTs that adopt the GT-A fold<sup>39</sup> ([Figure 2](#)). To gain further insights into their origins and evolution, we performed structural clustering of the phage GTs with representatives from diverse GT families using DALI<sup>40</sup>.  $\alpha$ -GTs and  $\beta$ -GTs represent two distinct GT families, GT63 and GT72, respectively, which are exclusively found in phages<sup>41</sup>. However, our analysis showed that in structural comparisons, the two families clustered together, forming sister groups in the structure-based dendrogram ([Figure 2](#)). Notably, among other GT-B superfamily members, the phage GTs were most closely related to the bacterial GT113-family GTs ([Figure 2](#)). These results suggest that  $\alpha$ -GT and  $\beta$ -GT have evolved following gene duplication from a common ancestor, which itself has originated from a cellular GT.  $\beta$ -

glucosyl-HMC- $\alpha/\beta$ -GT homologs from phages clustered together as a sister group to the GT2 family GTs of the GT-A fold superfamily, also suggesting cellular ancestry of the phage  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GTs ([Figure 2](#)).

Together, these results suggest that cellular GTs have been coopted for hypermodification of the phage DNA at least twice independently ( $\alpha$ -GT/ $\beta$ -GT and  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GTs, respectively). In the case of the primary  $\alpha$ -GT and  $\beta$ -GT, it is tempting to speculate that specificity towards the DNA substrate has evolved once, in the common ancestor of the two families, followed by gene duplication and diversification of the paralogs towards different stereospecificities.

### **Analysis of protein-protein interactions associated with $\alpha$ -GT and $\beta$ -GT**

Before addressing the potential advantage of encoding two GTs with different stereospecificities, we investigated the mechanisms behind the distribution of  $\alpha$ - and  $\beta$ -glycosidic modifications on cytosines in T4 DNA. Previously it has been estimated that  $\alpha$ -GT and  $\beta$ -GT contribute to 60% and 40% of glucosylation, respectively<sup>17,19</sup>. To explain that distribution, it has been previously hypothesized, based on *in vitro* results<sup>20</sup>, that T4  $\alpha$ -GT binds to gp45, the sliding clamp protein that forms the basis of the T4 replication machinery, and modifies the T4 DNA concomitantly with replication, with  $\beta$ -GT presumed to intervene later, on unmodified bases. To investigate this possibility, we analyzed the interaction between the sliding clamp protein and  $\alpha$ -GT or  $\beta$ -GT. C-terminally histidine tagged versions of  $\alpha$ -GT and  $\beta$ -GT were individually purified from *E. coli*. Based on the elution volume during size-exclusion chromatography it was estimated that both proteins exist as monomers ([Figure S3](#)). Tag-free gp45 was overexpressed in *E. coli* ([Figure S4](#)). Soluble fractions containing either tagged  $\alpha$ -GT or  $\beta$ -GT were combined with the fraction containing gp45 prior to immobilized metal affinity chromatography (IMAC) purification. Compared to the controls from which the gp45-containing fraction was omitted, no co-purification of protein matching the size of gp45 was observed in the eluted fractions ([Figure 3A](#)), contrary to what was previously observed<sup>20</sup>.

To identify any possible interactions *in vivo*, we engineered two phages encoding either  $\alpha$ -GT or  $\beta$ -GT, along with an in-frame coding sequence for the histidine-tag at their C-termini, T4  $\Delta\beta\text{-gt}/\alpha\text{-gt}_{Chis}$  and T4  $\Delta\alpha\text{-gt}/\beta\text{-gt}_{Chis}$  ([Figure S5](#)). The recombinant phages did not display any loss of virulence in *E. coli* MG1655, which encodes the type IV RM systems ([Figure S6](#)). The protein pull-down experiments were performed at least twice using IMAC purification and analyzed by mass spectrometry-based proteomics. Under *in vivo* conditions, neither  $\alpha$ -GT nor  $\beta$ -GT co-purified with the sliding clamp protein or any other phage protein ([Figure 3B](#)). Proteomic analysis of each pull-down sample further confirmed the lack of interaction between the two GTs and other phage proteins ([Table S2](#)).

Here, through *in vitro* and *in vivo* experiments, we show that neither  $\alpha$ -GT nor  $\beta$ -GT interact with the sliding clamp protein or other components of the replication machinery of phage T4, contrary to the published results. This suggests that  $\alpha$ -GT and  $\beta$ -GT compete directly for their substrate on T4 DNA.

### **Relative enzymatic activity and expression strength govern the distribution of glycosidic linkages**

In the absence of interaction with the replisome, this distribution could result from competition between the two enzymes, with  $\alpha$ -GT being either produced more, and/or enzymatically more active.

In order to identify the contributing factors underlying the specific glucosylation levels, we estimated the difference in enzymatic activity between the two GTs. Using the C-terminally histidine tagged GTs purified earlier, we performed a competition assay. The two GTs were mixed at different molar ratios and incubated with a 5-hmC containing PCR fragment. The contribution of each GT was estimated from the percentage 5-hmC residues that were either  $\alpha$ - or  $\beta$ -glucosylated. In the absence of  $\beta$ -GT,  $\alpha$ -GT contributed to the glucosylation of 64% of 5-hmC bases. Upon addition of  $\beta$ -GT, even at a concentration four times lower than  $\alpha$ -GT, 91% of 5-hmC were  $\beta$ -glucosylated, with the  $\alpha$ -GT contributing to only 0.65% of glucosylations ([Figure 4A](#)). These results demonstrate that under the *in vitro* conditions used,  $\beta$ -GT exhibits much higher enzymatic activity than  $\alpha$ -GT, despite contributing to only 40% of all glucosylations *in vivo*. This strongly suggests that the higher level of glucosylation by  $\alpha$ -GT in T4 phage DNA is not due to its higher enzymatic activity.

We then reasoned that the difference could stem from higher or earlier expression of  $\alpha$ -gt during the T4 lytic cycle. An analysis of previously published transcriptomics data<sup>42</sup> showed no significant differences in the expression levels of the two GTs, despite slight differences in the timing of expression ([Figure S7](#)). The limited difference observed could not conclusively explain the differences in glucosylation levels observed *in vivo*, especially given the significantly higher enzymatic activity of  $\beta$ -GT observed *in vitro*. To elucidate the role of expression strength, we set out to modify the transcription levels of  $\alpha$ -gt and  $\beta$ -gt. Due to the technical challenges associated with direct promoter modification on phage genome and lack of knowledge on the ensuing impact on expression strength, we opted instead to switch the positions of the two GTs. We constructed the mutant phage T4 $\nabla$ P $_{\alpha\text{-gt}}::\beta\text{-gt}\nabla$ P $_{\beta\text{-gt}}::\alpha\text{-gt}$ , encoding both GTs but under the control of each other's regulatory sequences, along with two additional phage mutants T4 $\nabla$ P $_{\beta\text{-gt}}::\alpha\text{-gt}$  and T4 $\nabla$ P $_{\alpha\text{-gt}}::\beta\text{-gt}$ , encoding either  $\alpha$ -GT or  $\beta$ -GT under the control of the promoter of the other GT ([Figure 4B](#), [Figure S8](#) and [Figure S9](#), see Methods section).

We then evaluated the impact of the switching of promoters on the relative levels of  $\alpha$ - and  $\beta$ -glucosylated cytosine residues by LC-MS/MS. In phage T4 $\nabla$ P $_{\beta\text{-gt}}::\alpha\text{-gt}$ , 75% of 5-hmC were glucosylated, exclusively by  $\alpha$ -GT ([Figure 4C](#) and [Figure 4E](#)). By contrast, in phage T4 $\nabla$ P $_{\alpha\text{-gt}}::\beta\text{-gt}$ , nearly all 5-hmC were  $\beta$ -glucosylated ([Figure 4C](#) and [Figure 4E](#)). In T4 $\nabla$ P $_{\alpha\text{-gt}}::\beta\text{-gt}\nabla$ P $_{\beta\text{-gt}}::\alpha\text{-gt}$ , 100% of 5-hmC were glucosylated, but only 12% were  $\alpha$ -glucosylated, with the remaining 87% being  $\beta$ -glucosylated ([Figure 4C](#) and [Figure 4E](#)). To interpret these results, we also determined the percentage of glucosylation in the wild-type T4, in the single *gt* deletion mutants that we previously constructed<sup>43</sup> and in a double *gt* deletion mutant that was generated in this study ([Figure S10](#)). In wild-type T4, 66% of all 5-hmC bases were  $\alpha$ -glucosylated and the remaining 33% were  $\beta$ -glucosylated ([Figure 4D](#) and [Figure 4E](#)), in close agreement with the previous estimations<sup>17,19</sup>. In the phage lacking both GTs, as expected, all cytosines were only hydroxymethylated. Glucosylation by  $\alpha$ -GT increased from 66% in the wild-type to 80% in T4 $\Delta\beta\text{-gt}$ . Similarly, in T4 $\Delta\alpha\text{-gt}$  the percentage of glucosylation by  $\beta$ -GT is increased, reaching 99%, while it is only 33% in the wild-type T4.

The increased  $\beta$ -glucosylation when  $\beta$ -*gt* is expressed from  $P_{\alpha\text{-gt}}$  compared to  $P_{\beta\text{-gt}}$  promoter (88% in phage  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$  as compared to 33% in wild-type T4) suggests that the expression strength of  $P_{\alpha\text{-gt}}$  is greater than that of  $P_{\beta\text{-gt}}$ . Further supporting this conclusion, upon expression from  $P_{\beta\text{-gt}}$ ,  $\alpha$ -GT was able to glucosylate only 75% of 5-hmdC as compared 80% when expressed from its native promoter in phage  $T4\Delta\beta\text{-gt}$ , a small but significant difference. Moreover, concomitant expression of  $\beta$ -*gt* from the  $P_{\alpha\text{-gt}}$  promoter in phage  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$  lowers the contribution of  $\alpha$ -GT by 7-fold.

Together, these results demonstrate that the *in vitro* enzymatic activity of  $\beta$ -GT is greater than that of  $\alpha$ -GT, confirmed *in vivo* by the complete glucosylation of 5-hmC by  $\beta$ -GT in  $T4\Delta\alpha\text{-gt}$  in comparison to only 80% glucosylation by  $\alpha$ -GT in  $T4\Delta\beta\text{-gt}$ . In phage T4, the higher enzymatic activity of  $\beta$ -GT is compensated by the stronger upstream regulatory sequence and the Shine-Dalgarno (SD) motif associated with  $\alpha$ -*gt*, resulting in  $\alpha$ -glucosylation being the predominant modification.

### Each individual GT is sufficient to provide wild-type level virulence

To investigate the potential selective advantage of encoding two primary GTs, as a first step, we characterized the virulence of T4 GT deletion mutants in relation to their percentage of cytosine glucosylation. To this end, we estimated the killing activity of the wild-type, single and double deletion mutant phages by determining their virulence index ( $V_i$ ), a recently proposed quantitative measure of the virulence of a phage against a given host <sup>44</sup>.

Except for the double deletion mutant,  $T4\Delta\alpha\text{-gt}\Delta\beta\text{-gt}$ , the WT and single deletion mutants were able to inhibit growth of the host at all MOIs (Figure 5A). In congruence,  $V_i$  estimated over 7 hours of growth post infection was similar between these phages, but the double mutant had a much lower  $V_i$  (Figure 5B and Figure S11).

To evaluate the impact of the glucosylation levels obtained with switched promoters on phage virulence, growth of the three mutant phages on *E. coli* MG1655 was compared to that of the wild-type phage (Figure 5C). While the mutant phages  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}$  and  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$  demonstrated the same levels of virulence as the wild-type phage, the mutant phage  $T4\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$  showed decreased virulence, evident at lower MOIs (Figure 5D and Figure S12). Spot assay of the mutant phages  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}$  and  $T4\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$  on *E. coli* DH10B (no RM systems) and *E. coli* MG1655 confirmed that the decrease in virulence of the latter mutant phage was specific to the presence of RM systems. In addition to the reduced plaque size, a 5-fold decrease in the efficiency of plating was observed with phage  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}$  in the presence of RM systems (Figure 5E).

In light of the glucosylation levels observed in the single *gt* deletion mutants, the decreased virulence observed in phage  $T4\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$  suggests that 80% glucosylation, observed in  $T4\Delta\beta\text{-gt}$ , is the absolute threshold below which phage DNA is susceptible to the activity of the *E. coli* type IV RM systems. Previous estimations of glucosylation level were performed using genomic DNA obtained from phages propagated in an RM-negative strain (*E. coli* DH10B). To determine if the presence of the RM system influences the glucosylation level, the analysis was repeated for phages T4,  $T4\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$ ,  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}$  and  $T4\Delta P_{\alpha\text{-gt}}::\beta\text{-gt}\Delta P_{\beta\text{-gt}}::\alpha\text{-gt}$ , after propagation in *E. coli* MG1655 (RM-positive) strain. Among the four phages, the

percentage of 5-hmC that were  $\alpha$ -glucosylated or  $\beta$ -glucosylated remained the same to that observed earlier ([Figure 5F](#) and [Figure 4E](#)). These results further support that the decreased virulence observed in phage T4 $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$  was indeed due to a decrease in the percentage of overall glucosylation.

Taken together, these results demonstrate that T4  $\alpha$ -GT and  $\beta$ -GT have evolved to be able to individually glucosylate at least 80% of all 5-hmC bases in the T4 genome, which is sufficient for complete protection from 5-hmC targeting type IV RM systems and is therefore not the explanation for the presence of two GTs. The reduced virulence of phage T4 $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$ , observed in RM-positive strain, reinforces our previous conclusion that the reduced  $\beta$ -glucosylation in wild-type T4 is due to the greater expression of  $P_{\alpha\text{-gt}}$  in relation to  $P_{\beta\text{-gt}}$ .

### **$\beta$ -GT is essential for phage T4 survival in wild-type *E. coli* encoding RM and Brig1 defense systems.**

Despite the redundancy ensuring complete protection against type IV RM systems in the event of loss of a GT, we reasoned that encoding two GTs could be beneficial against other defense systems targeting glycosylated DNA. A DNA glycosylase-based host defense system, Brig1, was recently shown to provide immunity against T-even phages<sup>27</sup> by generating abasic sites in DNA containing 5-gmC with  $\alpha$ -glycosidic linkages. The authors showed that phage T4 was able to produce escaper mutants in the presence of cosmid-borne Brig1, in the *E. coli* EC100 strain devoid of the type IV RM systems targeting 5-hmC ([Figure S13](#)).

To understand the influence of varying levels of  $\alpha$ -glucosylation on the anti-phage activity of Brig1, we determined the escaper frequency ( $\alpha$ -gt inactivating mutants) on a RM-negative strain across a range of phage mutants with varying levels of  $\alpha$ -glucosylation. Among phages T4 $\Delta\beta\text{-gt}$  and T4 $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$ , both lacking  $\beta$ -GT and having higher  $\alpha$ -glucosylation levels than wild-type T4, T4 $\Delta\beta\text{-gt}$  showed the lowest escaper frequency. In comparison, a 6-fold higher escaper frequency was observed with phage T4 $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$ , which could be related to its lower  $\alpha$ -glucosylation levels. The presence of  $\beta$ -GT, which consequently lowers  $\alpha$ -glucosylation levels, resulted in a further 2-fold higher frequency of escapers mutants in the wild-type T4 and T4 $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$  phages ([Figure 6A](#)). Hence, decrease in the  $\alpha$ -glucosylation levels increases the escaper frequency until saturation is reached at low  $\alpha$ -glucosylation levels. These findings demonstrate that in the absence of type IV RM systems, the competition between  $\beta$ -GT and  $\alpha$ -GT increases the ability of T4 to escape Brig1 by lowering the  $\alpha$ -glucosylation levels.

The formation of abasic sites during Brig1 activity results in escapers that are likely to contain single base pair deletions/insertions or substitution mutations, rather than large deletions typical of escapers generated through mini-homology mediated recombination<sup>45</sup>. To confirm this, we analyzed the  $\alpha\text{-gt}$  locus from escapers of Brig1 in comparison to escapers from CRISPR-Cas targeting of the same locus. Indeed, none of the 29 escapers from Brig1 targeting showed any difference in size, relative to the coding region from the wild-type phage, as opposed to the escapers of CRISPR-Cas targeting ([Figure S14](#) and [Table S3](#)).

Next, we investigated the role of T4  $\beta$ -GT in a scenario where the host encodes both DNA glycosylases and RM systems. In wild-type T4, Brig1 escaper frequency was

independent of the presence of RM systems ([Figure 6B](#)). On the other hand, phage T4 $\Delta\beta$ -*gt* was unable to generate escaper mutants in the presence of both RM and Brig1 ([Figure 6B](#)). To quantify the consequences of these results on a population level, we estimated the virulence index of phage T4 and T4 $\Delta\beta$ -*gt* in hosts encoding Brig1 either in the absence or presence of RM systems. T4 virulence was independent of the presence of RM systems but phage T4 $\Delta\beta$ -*gt* lacking  $\beta$ -GT was completely unable to propagate in the presence of both RM and Brig1 ([Figure 6C](#), [Figure S15](#) and [Figure 6D](#)). Based on these results we can conclude that in the presence of type IV RM systems,  $\beta$ -GT enables the survival of  $\alpha$ -GT inactivated Brig1 escapers by adequately glucosylating phage DNA.

Together, these results show that the antiphage activity of Brig1 and phage survival is partly dependent on the amount of  $\alpha$ -glucosylation within the phage DNA. We also show that the fitness of Brig1 escaper T4 mutants in a host strain encoding type I and type IV RM systems is entirely dependent on the phage encoding an alternative primary modification system such as  $\beta$ -*gt*. These results also demonstrate that in the presence of a single GT, Brig1 combined with RM systems can completely prevent emergence of phage escaper mutants.

## DISCUSSION

DNA hypermodification in phages is a general counter defense strategy against host defense systems whose mode of action is dependent on DNA sequence recognition, such as CRISPR-Cas and RM systems. Although it was known that bacteriophage T4 encodes two glucosyltransferases with similar functions, the mechanisms underlying their respective activities and the advantage of this functional redundancy had never been thoroughly investigated.

In phage T4, the cytosines are pre-replicatively hydroxymethylated and subsequently glucosylated to the fullest extent. We show that  $\alpha$ -GT and  $\beta$ -GT, respectively, contribute to 66% and 33% of the total glucosylation, consistent with the prior estimates of 60% and 40%, respectively<sup>17,19</sup>. Competition assay involving the two GTs demonstrated that, under *in vitro* conditions,  $\beta$ -GT exhibited greater enzymatic activity than  $\alpha$ -GT, which contrasts with its lower contribution to glucosylation observed *in vivo* in wild-type T4. It has been previously hypothesized that the higher *in vivo* contribution of  $\alpha$ -GT was due to a strong interaction between  $\alpha$ -GT and the T4 replication machinery. Yet, under our tested *in vitro* conditions, neither the  $\alpha$ -GT nor  $\beta$ -GT interacted with the T4 sliding clamp protein, T4 gp45, and metal-affinity purification using genomically tagged  $\alpha$ -GT or  $\beta$ -GT revealed no interacting proteins. Rather, thanks to experiments of promoter switching in T4 genome, we show that the proportion of cytosines undergoing  $\alpha$ - or  $\beta$ -glucosylation is significantly modulated by the transcriptional and translational activities associated with each GT, together with their enzymatic activity.

Despite strict size constraints on phage genomes, why do T4 and other closely related phages encode two primary GTs? We show that deleting either *gt* did not affect phage fitness, even in the presence of type IV RM systems targeting 5-hmC containing DNA. Moreover, the observation that, in the presence of  $\alpha$ -GT alone, 80% of 5-hmC glucosylation fully protect phages against *E. coli* type IV systems indicates that this degree of glucosylation is sufficient for protection. Interestingly, we observed that the mutant phage T4 $\nabla$ P $_{\beta\text{-gt}}::\alpha\text{-gt}$ , with  $\alpha\text{-gt}$  expressed from the  $\beta\text{-gt}$  promoter, makes smaller plaques and has a lower virulence index than other phages. The percentage of glucosylated 5-hmC was estimated to be 75%, a lower proportion of glucosylated bases as compared to the 80% glucosylation observed in single deletion genotypes. This suggest that 80% of glucosylation is the absolute minimal threshold necessary to protect against the *E. coli* McrBC/McrA RM system, an outcome that might be generalizable to other *E. coli* phages. In conclusion, in the wild-type phage, competition between the two enzymes ensures balanced distribution of the two types of linkages, but when expressed alone each enzyme utilizes either its expression strength or superior enzymatic activity to ensure adequate glucosylation and complete protection of phage DNA.

Here we also demonstrate that growth of phages encoding a single GT is entirely restricted in hosts encoding defense systems, such as Brig1, that target specific stereoisomers of 5-gmC, in addition to classical RM systems. On the contrary, phages encoding two different GTs can generate escape mutants at high frequency ( $1 \times 10^{-5}$ ). Indeed, in the presence of a single GT, mutations inactivating  $\alpha$ -GT leads to phage DNA targeting by type IV RM systems and are therefore lethal. By contrast, in the presence of two GTs with different stereospecificity, mutations in  $\alpha\text{-gt}$  are offset by  $\beta$ -glucosylation, ensuring survival of escape mutants with inactivating mutations in  $\alpha\text{-gt}$ .

Furthermore, among hosts encoding only the DNA glycosylase defense system, the presence of non-targeted GT(s) promotes a higher rate of escaper generation. Our results emphasize the benefits of GT diversity in phage genomes, and in the diversity of the types of modification on a more general level. These findings also underscore a notable drawback inherent to DNA glycosylase-based defense systems, such as Brig1, in that their activity invariably results in escaper phages in the absence of co-occurring defense systems. Analysis of Brig1 escapers across the  $\alpha$ -gt locus showed that the mutants comprised only point mutations (single base pair insertions or deletions and substitutions), with none of the mutant's showing larger deletions across or within  $\alpha$ -gt, in contrast to what is observed in escapers from CRISPR-Cas systems<sup>45</sup>. This constitutes an advantage for Brig1 escapers, as point mutations are reversible, offering the possibility to recover functional genes. Hence encoding two GTs whose products are stereoisomers confers an evolutionary advantage when faced with a stereoisomer specific host defense system.

The frequent association of  $\beta$ -glucosyl-HMC- $\alpha$ -GT homologs with both  $\alpha$ -GT and  $\beta$ -GT indicates that either the secondary GT does not differentiate between the two primary linkage types or that there are two distinct kinds of secondary GT:  $\beta$ -glucosyl-HMC- $\alpha$ -GT and  $\beta$ -glucosyl-HMC- $\beta$ -GT. Furthermore, the presence of solitary secondary GT homologs, independent of both  $\alpha$ -GT and  $\beta$ -GT, in 288 out of 753 phage genomes indicates that these phages potentially encode novel primary transferases which forms the basis for additional glucosyl modifications. The not so limited prevalence of  $\beta$ -GT and arabinosylation clusters<sup>31</sup> among phages, along with the identification of diverse DNA hypermodification enzymes from metagenomic DNA<sup>46</sup>, suggest that Brig1 is likely not the only DNA glycosylase based host anti-phage defense system. Similarly, it is also likely that the yet to be defined function of the secondary GTs lies in their ability to overcome DNA glycosylases that indiscriminately target primary GTs. To an extent, this has been demonstrated in Bas46 phage, where deletion of the secondary arabinosyltransferase Aat resulted in susceptibility to a plasmid encoded defense system<sup>31</sup>. Overall, only 6 out of the 753 phage genomes, including phage T4 and its close relatives, encoded both  $\alpha$ -GT and  $\beta$ -GT. The limited co-occurrence of the two primary GTs might indicate that the adaptation of DNA glycosylase defense system by the hosts and the counter measure of encoding multiple hypermodification enzymes by phages is a relatively recent innovation.

Encoding multiple hypermodification enzymes such as GTs along with dCTPase, thymidylate synthase, IPI, IPII and IPIII is a great example of continuous antagonistic evolution of defense and counter defense systems. Phages hydroxymethylate their cytosines in response to sequence-specific targeting by the host type I RM system<sup>23</sup>, which, in turn, makes them susceptible to the type IV McrBC RM system<sup>5,24</sup>. This vulnerability is then countered by sugar modification, such as glucosylation or arabinosylation, of the cytosines<sup>17,18,31</sup>. These modifications not only confer resistance to type IV RM systems but also enable the phage to evade targeting by certain types of CRISPR-Cas systems<sup>5-10,31</sup>. Other type IV RM systems such as GmrSD are capable of indiscriminately targeting the DNA modified by the primary GTs, eliciting the evolution of phage encoded cognate inhibitors, IPI, IPII and IPIII<sup>25,26,28</sup>. Subsequently, when hosts evolve stereoisomer-specific DNA glycosylases such as Brig1<sup>27,32</sup>, instead of evolving a cognate inhibitor, phage T4 has acquired a variant GT enzyme with different stereospecificity,  $\beta$ -gt, to diversify the distribution of the glycosidic linkages in its genome ([Figure 7](#)).

## **DATA AVAILABILITY**

Supporting data linked to the findings in this study can be found in the main article or the supplementary materials. Further data will be made available from the corresponding author upon request.

## **ACKNOWLEDGEMENTS**

The authors thank all members of the MuSE lab (Micalis), INRAE, Jouy-en-Josas for the insightful discussions; Safir Hien Phan (University of Copenhagen, Denmark) for her technical support; Luciano A. Marraffini (The Rockefeller University, USA) for providing the pBrig1 plasmid and *E. coli* EC100 strain. We would also like to thank the BSRC Mass Spectrometry & Proteomics Facility, University of St. Andrews and PAPPSO, INRAE for their help with proteomic analysis. The authors acknowledge a DIM1Health 2019 grant from the Région Île-de-France to the project EpiK for the LC-MS equipment. This work was supported by the ANR grant MUMI [ANR-20-CE12-0008-02] to M.D.P. and the Novo Nordisk Fonden Postdoctoral Fellowship in Bioscience and Basic Biomedicine Grant [NNF21OC0067491] to Y.B.-C.

## **AUTHOR CONTRIBUTIONS**

Conceptualization, Y.B.-C. and M.D.P.; investigation, Y.B.-C., L.R.-C., F.B. and A.L.I.W.; formal analysis, L.R.-C., M.K., F.B., Y.B.-C. and F.L.; validation, L.R.-C., F.B. and A.L.I.W.; visualization Y.B.-C., L.R.-C. and M.K.; funding acquisition, Y.B.-C., M.D.P. and M.H.; supervision, Y.B.-C., M.D.P. and M.H.; writing – original draft, Y.B.-C., M.K., L.R.-C., F.B. and M.D.P; writing – review and editing, M.D.P., F.L., M.K., Y.B.-C. and M.H.

## **DECLARATION OF COMPETING INTEREST**

The authors declare that they have no competing financial interests.

## REFERENCES:

1. Hutinet, G., Lee, Y.J., de Crecy-Lagard, V., and Weigele, P.R. (2021). Hypermodified DNA in Viruses of *E. coli* and *Salmonella*. *EcoSal Plus* 9, eESP00282019. 10.1128/ecosalplus.ESP-0028-2019.
2. Weigele, P., and Raleigh, E.A. (2016). Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chem Rev* 116, 12655-12687. 10.1021/acs.chemrev.6b00114.
3. Sood, A.J., Viner, C., and Hoffman, M.M. (2019). DNAmoD: the DNA modification database. *J Cheminform* 11, 30. 10.1186/s13321-019-0349-4.
4. Meselson, M., and Yuan, R. (1968). DNA restriction enzyme from *E. coli*. *Nature* 217, 1110-1114. 10.1038/2171110a0.
5. Revel, H.R., and Luria, S.E. (1970). DNA-glucosylation in T-even phage: genetic determination and role in phagehost interaction. *Annu Rev Genet* 4, 177-192. 10.1146/annurev.ge.04.120170.001141.
6. Huang, L.H., Farnet, C.M., Ehrlich, K.C., and Ehrlich, M. (1982). Digestion of highly modified bacteriophage DNA by restriction endonucleases. *Nucleic Acids Res* 10, 1579-1591. 10.1093/nar/10.5.1579.
7. Miller, P.B., Wakarchuk, W.W., and Warren, R.A. (1985). alpha-Putrescinylothymine and the sensitivity of bacteriophage phi W-14 DNA to restriction endonucleases. *Nucleic Acids Res* 13, 2559-2568. 10.1093/nar/13.7.2559.
8. Bryson, A.L., Hwang, Y., Sherrill-Mix, S., Wu, G.D., Lewis, J.D., Black, L., Clark, T.A., and Bushman, F.D. (2015). Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. *mBio* 6, e00648. 10.1128/mBio.00648-15.
9. Tsai, R., Correa, I.R., Xu, M.Y., and Xu, S.Y. (2017). Restriction and modification of deoxyarchaeosine (dG(+))-containing phage 9 g DNA. *Sci Rep* 7, 8348. 10.1038/s41598-017-08864-4.
10. Vlot, M., Houkes, J., Lochs, S.J.A., Swarts, D.C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J., et al. (2018). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. *Nucleic Acids Res* 46, 873-885. 10.1093/nar/gkx1264.
11. Kutter, E.M., Bradley, D., Schenck, R., Guttman, B.S., and Laiken, R. (1981). Bacteriophage T4 alc gene product: general inhibitor of transcription from cytosine-containing DNA. *J Virol* 40, 822-829. 10.1128/JVI.40.3.822-829.1981.
12. Sadowski, P.D., Warner, H.R., Hercules, K., Munro, J.L., Mendelsohn, S., and Wiberg, J.S. (1971). Mutants of bacteriophage T4 defective in the induction of T4 endonuclease II. *J Biol Chem* 246, 3431-3433.
13. Hercules, K., Munro, J.L., Mendelsohn, S., and Wiberg, J.S. (1971). Mutants in a nonessential gene of bacteriophage T4 which are defective in the degradation of *Escherichia coli* deoxyribonucleic acid. *J Virol* 7, 95-105. 10.1128/JVI.7.1.95-105.1971.
14. Sinsheimer, R.L. (1954). Nucleotides from T2r+ bacteriophage. *Science* 120, 551-553. 10.1126/science.120.3119.551.
15. Lamm, N., Wang, Y., Mathews, C.K., and Ruger, W. (1988). Deoxycytidylate hydroxymethylase gene of bacteriophage T4. Nucleotide sequence determination and over-expression of the gene. *Eur J Biochem* 172, 553-563. 10.1111/j.1432-1033.1988.tb13925.x.

16. Morera, S., Imberty, A., Aschke-Sonnenborn, U., Ruger, W., and Freemont, P.S. (1999). T4 phage beta-glucosyltransferase: substrate binding and proposed catalytic mechanism. *J Mol Biol* 292, 717-730. 10.1006/jmbi.1999.3094.
17. Lehman, I.R., and Pratt, E.A. (1960). On the Structure of the Glucosylated Hydroxymethylcytosine Nucleotides of Coliphages T2, T4, and T6. *Journal of Biological Chemistry* 235, 3254-3259. 10.1016/s0021-9258(20)81347-7.
18. Kornberg, S.R., Zimmerman, S.B., and Kornberg, A. (1961). Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected *Escherichia coli*. *J Biol Chem* 236, 1487-1493.
19. Flodman, K., Correa, I.R., Jr., Dai, N., Weigele, P., and Xu, S.Y. (2020). In vitro Type II Restriction of Bacteriophage DNA With Modified Pyrimidines. *Front Microbiol* 11, 604618. 10.3389/fmicb.2020.604618.
20. Sommer, N., Depping, R., Piotrowski, M., and Ruger, W. (2004). Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism. *Biochem Biophys Res Commun* 323, 809-815. 10.1016/j.bbrc.2004.08.170.
21. Kuno, S., and Lehman, I.R. (1962). Gentiobiose, a constituent of deoxyribonucleic acid from coliphage T6. *J Biol Chem* 237, 1266-1270.
22. Lunt, M.R., and Newton, E.A. (1965). Glucosylated Nucleotide Sequences from T-Even Bacteriophage Deoxyribonucleic Acids. *Biochem J* 95, 717-723. 10.1042/bj0950717.
23. Loenen, W.A., Dryden, D.T., Raleigh, E.A., and Wilson, G.G. (2014). Type I restriction enzymes and their relatives. *Nucleic Acids Res* 42, 20-44. 10.1093/nar/gkt847.
24. Dila, D., Sutherland, E., Moran, L., Slatko, B., and Raleigh, E.A. (1990). Genetic and sequence organization of the *mcrBC* locus of *Escherichia coli* K-12. *J Bacteriol* 172, 4888-4900. 10.1128/jb.172.9.4888-4900.1990.
25. Bair, C.L., and Black, L.W. (2007). A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J Mol Biol* 366, 768-778. 10.1016/j.jmb.2006.11.051.
26. He, X., Hull, V., Thomas, J.A., Fu, X., Gidwani, S., Gupta, Y.K., Black, L.W., and Xu, S.Y. (2015). Expression and purification of a single-chain Type IV restriction enzyme Eco94GmrSD and determination of its substrate preference. *Sci Rep* 5, 9747. 10.1038/srep09747.
27. Hossain, A.A., Pigli, Y.Z., Baca, C.F., Heissel, S., Thomas, A., Libis, V.K., Burian, J., Chappie, J.S., Brady, S.F., Rice, P.A., and Marraffini, L.A. (2024). DNA glycosylases provide antiviral defence in prokaryotes. *Nature* 629, 410-416. 10.1038/s41586-024-07329-9.
28. Silas, S., Carion, H., Makarova, K.S., Sanchez Godinez, D., Haniyur, S., Volino, L., Yee, W.X., Fossati, A., Swaney, D., Bocek, M., et al. (2025). Anti-restriction functions of injected phage proteins revealed by peeling back layers of bacterial immunity. *Nat Commun* 16, 7828. 10.1038/s41467-025-63056-3.
29. Thomas, J.A., Orwenyo, J., Wang, L.X., and Black, L.W. (2018). The Odd "RB" Phage-Identification of Arabinosylation as a New Epigenetic Modification of DNA in T4-Like Phage RB69. *Viruses* 10. 10.3390/v10060313.
30. Borges, A.L., Dutton, R. J., McDaniel, E. A., Radkov, A., Reiter, T., & Weiss, E. C. P. (2023). Isolation of a phage with an arabinosylated genome from a cheese microbial community. *Arcadia Science*. 10.57844/arcadia-743p-ty94.
31. Mahler, M., Cui, L., Smith, L.M., Wandera, K.G., Dietrich, O., Mayo-Munoz, D., Balamkundu, S., Lee, M.E., Ye, H., Liu, C.F., et al. (2025). Phage arabinosyl-

- hydroxy-cytosine DNA modifications result in distinct evasion and sensitivity responses to phage defense systems. *Cell Host Microbe* 33, 1173-1190 e1179. 10.1016/j.chom.2025.06.005.
32. Getz, L.J., Qian, A.L., Liu, Y.V., Fairburn, S.R., Butt, M.S., Lee, Y.-J., Weigele, P.J., and Maxwell, K.L. (2025). Antiviral defence is a conserved function of diverse DNA glycosylases. *bioRxiv*. 10.1101/2025.10.29.685425.
  33. Glazko, G., Makarenkov, V., Liu, J., and Mushegian, A. (2007). Evolutionary history of bacteriophages with double-stranded DNA genomes. *Biol Direct* 2, 36. 10.1186/1745-6150-2-36.
  34. Kristensen, D.M., Cai, X., and Mushegian, A. (2011). Evolutionarily conserved orthologous families in phages are relatively rare in their prokaryotic hosts. *J Bacteriol* 193, 1806-1814. 10.1128/JB.01311-10.
  35. Nurmammedov, E., Castelnovo, M., Medina, E., Catalano, C.E., and Evilevitch, A. (2012). Challenging packaging limits and infectivity of phage lambda. *J Mol Biol* 415, 263-273. 10.1016/j.jmb.2011.11.015.
  36. Georgopoulos, C.P. (1967). Isolation and preliminary characterization of T4 mutants with nonglycosylated DNA. *Biochem Biophys Res Commun* 28, 179-184. 10.1016/0006-291x(67)90426-3.
  37. Morera, S., Lariviere, L., Kurzeck, J., Aschke-Sonnenborn, U., Freemont, P.S., Janin, J., and Ruger, W. (2001). High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding. *J Mol Biol* 311, 569-577. 10.1006/jmbi.2001.4905.
  38. Lariviere, L., Sommer, N., and Morera, S. (2005). Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase. *J Mol Biol* 352, 139-150. 10.1016/j.jmb.2005.07.007.
  39. Lairson, L.L., Henrissat, B., Davies, G.J., and Withers, S.G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* 77, 521-555. 10.1146/annurev.biochem.76.061005.092322.
  40. Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Res* 50, W210-W215. 10.1093/nar/gkac387.
  41. Drula, E., Garron, M.L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 50, D571-D577. 10.1093/nar/gkab1045.
  42. Wolfram-Schauerte, M., Pozhydaieva, N., Viering, M., Glatzer, T., and Hofer, K. (2022). Integrated Omics Reveal Time-Resolved Insights into T4 Phage Infection of *E. coli* on Proteome and Transcriptome Levels. *Viruses* 14. 10.3390/v14112502.
  43. Bhoobalan-Chitty, Y., Stouf, M., and De Paepe, M. (2024). Genetic manipulation of bacteriophage T4 utilizing the CRISPR-Cas13b system. *Front Genome Ed* 6, 1495968. 10.3389/fgeed.2024.1495968.
  44. Storms, Z.J., Teel, M.R., Mercurio, K., and Sauvageau, D. (2020). The Virulence Index: A Metric for Quantitative Analysis of Phage Virulence. *Phage (New Rochelle)* 1, 27-36. 10.1089/phage.2019.0001.
  45. Wu, X., Zhu, J., Tao, P., and Rao, V.B. (2021). Bacteriophage T4 Escapes CRISPR Attack by Minihomology Recombination and Repair. *mBio* 12, e0136121. 10.1128/mBio.01361-21.
  46. Pyle, J.D., Lund, S.R., O'Toole, K.H., and Saleh, L. (2024). Virus-encoded glycosyltransferases hypermodify DNA with diverse glycans. *Cell Rep* 43, 114631. 10.1016/j.celrep.2024.114631.

47. Gibson, D.G. (2009). Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* 37, 6984-6990. 10.1093/nar/gkp687.
48. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402. 10.1093/nar/25.17.3389.
49. Meier-Kolthoff, J.P., and Goker, M. (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* 33, 3396-3404. 10.1093/bioinformatics/btx440.
50. Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., and Goker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14, 60. 10.1186/1471-2105-14-60.
51. Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinformatics* 69, e96. 10.1002/cpbi.96.
52. Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., and Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* 47, W260-W265. 10.1093/nar/gkz303.
53. Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44, W242-245. 10.1093/nar/gkw290.
54. Moraru, C., Varsani, A., and Kropinski, A.M. (2020). VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses* 12. 10.3390/v12111268.
55. Vallat, B., Rose, Y., Piehl, D.W., Duarte, J.M., Bittrich, S., Bi, C., Segura, J., Zalevsky, A., Sekharan, M.R., Webb, B.M., et al. (2025). RCSB Protein Data Bank: Delivering integrative structures alongside experimental structures and computed structure models. *Nucleic Acids Res*. 10.1093/nar/gkaf1187.
56. Kim, R.S., Levy Karin, E., Mirdita, M., Chikhi, R., and Steinegger, M. (2025). BFVD-a large repository of predicted viral protein structures. *Nucleic Acids Res* 53, D340-D347. 10.1093/nar/gkae1119.
57. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493-500. 10.1038/s41586-024-07487-w.
58. Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem Sci* 20, 478-480. 10.1016/s0968-0004(00)89105-7.
59. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin, T.E. (2023). UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* 32, e4792. 10.1002/pro.4792.

## METHODS

### Growth conditions, PCR, plasmids and strains

*Escherichia coli* MG1655 was utilized for the growth curve assays, involving the estimation of virulence index, and *E. coli* DH10B was utilized in genome editing experiments and plaque assays. All *E. coli* cultures were grown in Luria-Bertani (LB) medium, incubated at 37°C with shaking at 200 rpm.

Oligonucleotides and PCR fragments were fused together either by Overlap Extension-PCR or by Gibson Assembly (NEB #E2611S)<sup>47</sup> and cloned into the plasmid pBAD24. In case of the positional swapping between  $\alpha$ -gt and  $\beta$ -gt, the coding sequence of  $\beta$ -GT was first replaced with a pseudo-gene whose sequence includes appropriate protospacers for targeting by the CRISPR-Cas13b system. A pBAD24 based plasmid was constructed with the coding sequence of  $\alpha$ -GT fused to the homologous arms that correspond to the sequences flanking the coding sequence of  $\beta$ -GT. Similarly, a plasmid containing the coding sequence of  $\beta$ -GT flanked by the upstream and downstream sequence of  $\alpha$ -GT was also constructed. A similar strategy was used to insert glucosyltransferases with in-frame histidine coding sequence.

Hydroxymethylated deoxycytosines (hmdC) containing DNA fragment was prepared by PCR amplification of DNA fragment using the primers T4 DNAPol For and T4 DNAPol Rev with T4 DNA as template. In the PCR preparation, the dNTP mixture was substituted with an alternative mixture that was prepared by mixing the individual nucleotides, dATP, dGTP and dTTP along with 5-Hydroxymethyl-dCTP (Jena Bioscience, #NU-932S).

Spacers corresponding to target protospacers in  $\alpha$ -gt,  $\beta$ -gt and pseudo-gene were introduced into the plasmid pBZCas13b (Addgene #103986), encoding Cas13b, as described earlier. Briefly, overlapping oligonucleotides with 5' and 3' extensions were pooled together at 1  $\mu$ M final concentration, along with PCR reaction buffer. The mixture was incubated at 95°C for 10 minutes and gradually cooled to room temperature to facilitate annealing. The annealed oligonucleotides were ligated into the pBZCas13b vector linearized with the BsaI-HF (NEB #R3733S).

For heterologous expression of individual proteins, the coding sequences were cloned into the pET-28a(+) vector while retaining an in-frame C-terminal histidine sequence (in the case of  $\alpha$ -gt and  $\beta$ -gt) or without an in-frame histidine coding sequence (T4 gp45).

All primers, plasmids and strains used in this study are listed in [Table S4](#).

### Bioinformatic analysis

Homologs of the three glucosyltransferases, NP\_049673.1 ( $\alpha$ -GT, pfam11440), NP\_049658.1 ( $\beta$ -GT, pfam09198) and YP\_010067197.1 ( $\beta$ -glucosyl-HMC- $\alpha$ -GT, pfam20691), encoded among viruses (taxid = 10239), in the RefSeq database were identified using Position Specific Iterative BLAST (PSI-BLAST)<sup>48</sup>, with an expect threshold = 0.005, over three iterations. The protein ids of the homologs of all three glucosyltransferases were combined and organized according to their respective phage accession numbers, which was then utilized to visualize their distribution. A similar analysis was carried out against all viruses in the non-redundant (nr) NCBI GenBank database.

VICTOR (Virus Classification and Tree Building Online Resource) <sup>49</sup> was used to construct the phage phylogenetic tree based on phages retrieved from the RefSeq database that encode at least one GT and belong to the genus *Tequatrovirus*. Representative phages from diverse genera that also encode at least one GT were also included. Pairwise comparisons of the phage nucleotide sequences were conducted using the Genome-BLAST Distance Phylogeny method (including 100 pseudo-bootstrap replicates each) with the D0 formula for nucleotides <sup>50</sup>. The tree was visualized with ggtree <sup>51</sup>. To construct a phylogenetic tree of T6  $\beta$ -glucosyl-HMC- $\alpha$ -GT, protein homologs were retrieved from RefSeq database during the earlier PSI-BLAST and submitted to the phylogenetic analysis tool NGphylogeny.fr with default settings <sup>52</sup>. The Newick tree generated by PhyML was visualized on iTOL <sup>53</sup>.

For comparison of the six phage genomes, the complete nucleotide genome sequences of *Escherichia* phage 236Ecol005PP (PP434423.1), *Escherichia* phage FL12 (PP400786.1), *Escherichia* phage Pu-Krd-SF1 (PQ249395.1), *Escherichia* phage T4 (NC\_000866.4), *Enterobacteria* phage RB55 (KM607002.1) and *Escherichia* phage vB-Eco-KMB26 (OR539218.1) were downloaded in fasta format and combined into a single file. The nucleotide sequences were then analyzed using VIRIDIC WEB <sup>54</sup>, to calculate the nucleotide similarity between each of the six genomes, with default BLASTN parameters ('-word\_size 7 -reward 2 -penalty -3 -gapopen 5 -gapextend2').

### Three-dimensional structure prediction and analysis

PDB identifiers for representatives of different GT families were retrieved from the CAZy database <sup>41</sup>. All viral and cellular protein structures were downloaded from the PDB database <sup>55</sup>. Structural models of phage proteins for which experimentally determined structures were not available were either downloaded from the BFVD database <sup>56</sup> or modelled using AlphaFold3 <sup>57</sup>. Structural similarities between cellular and viral proteins were evaluated based on the DALI Z score, which is a measure of the quality of the structural alignment. Z scores above 2, i.e., two SDs above expected, are usually considered significant <sup>58</sup>. Structural similarity matrix from all-against-all structure comparisons as well as corresponding dendrograms were obtained using the latest release of the DALI server <sup>40</sup>. Structures were visualized using the University of California, San Francisco (UCSF) ChimeraX v1.9 <sup>59</sup>.

### Protein purification

200 ml cultures of *E. coli* BL21(DE3) strains (Novagen) carrying pET-28a(+) <sub>$\alpha$</sub> -*gtchis* or pET-28a(+) <sub>$\beta$</sub> -*gtchis* were grown in LB medium at 37°C, 200rpm until OD<sub>600</sub> = 0.6-0.8 was reached. Cultures were then incubated at 30°C for 3 hours and protein expression induced by the addition of IPTG to a final concentration 0.5 mM. Cells were collected by centrifugation at 6300 x g for 10 minutes at 16°C. The pellets were resuspended in lysis buffer (25 mM Tris-Cl, 200 mM NaCl, 5% glycerol and 30 mM Imidazole, pH = 7.5) and stored at -16°C until further processing. The thawed mixture was lysed by homogenization (STANSTED model SPCH-10 from homogenizing systems, UK) and sonication (30 cycles, 3 seconds ON and 3 seconds OFF). The cell debris was removed by centrifugation at 12000 x g for 30 minutes at 4°C and the supernatant was filtered with a 0.45  $\mu$ M filter. Proteins were bound to histrap columns

(HisTrap™ High Performance column, Cytiva), equilibrated with the lysis buffer. After washing with 40 column volumes of the lysis buffer including 30 mM Imidazole, the proteins were eluted in a buffer including 500 mM Imidazole. The eluted fractions, containing the protein of interest, were pooled together and concentrated using Pierce™ Protein Concentrators PES 10K MWCO (ThermoFisher SCIENTIFIC, #88516). The concentrated fractions were loaded on a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in 25 mM Tris-Cl, 200 mM NaCl, 5% glycerol, pH = 7.5 (SEC buffer).

### **Analysis of $\alpha$ -GT-gp45 and $\beta$ -GT-gp45 interactions, *in vitro***

For the analysis of *in vitro* protein-protein interaction, induction of pET-28a(+)\_gp45, along with pET-28a(+)\_ $\alpha$ -gtchis or pET-28a(+)\_ $\beta$ -gtchis was performed as described above. The filtered supernatants containing tag-free gp45 was split into two fractions and mixed with supernatants containing  $\alpha$ -GT chis or  $\beta$ -GT chis. The supernatants were then combined with Ni-NTA Agarose (Qiagen, #30210), equilibrated with the lysis buffer. The mixtures were incubated at 4°C for 60 minutes, washed with lysis buffer several times and eluted with a buffer containing 500 mM Imidazole. Filtered supernatants of the individual proteins were also treated similarly and purified, as controls.

### ***In vivo* pull-down with endogenously tagged $\alpha$ -GT and $\beta$ -GT**

The pull-down assay was carried out as described previously<sup>20</sup>, with minor modifications. *E. coli* MG1655 cultures in exponential phase (OD<sub>600</sub> = 0.5-0.6) were infected with mutant phages, that encoded a histidine tagged  $\alpha$ -GT or  $\beta$ -GT, at an MOI of 3. The infected culture was then incubated for 15 min at 37 °C, chilled on ice and centrifuged at 6000 xg for 10 min at 4 °C. Pellets were conserved at -20 °C until further use. Pellets were resuspended in a lysis solution (50 mM Tris-HCl pH = 7.5, 250 mM NaCl) mixed with cComplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Sigma #11836170001) and sonicated at 200 W 5 cycles, 10 sec on, 10 sec off per cycle (Sonics cat. VC505). 1  $\mu$ L of Turbo DNase I (Invitrogen #AM2238) was added to the lysate, which was then centrifuged at 10 000 xg for 10 min. Supernatants were filtered through 0.20  $\mu$ m PVDF (Fisher # 097203) and incubated overnight at 4 °C in contact with Ni resin (NEB #S1428S). The protein bound resins were washed thrice with 20 mM Tris-HCl pH = 7.5, 150 mM NaCl and 30 mM Imidazole. Finally, the bound proteins were eluted with 20 mM Tris-HCl pH = 7.5, 150 mM NaCl and 500 mM Imidazole.

### **Competition assay**

Substrate DNA containing 5-hmdC was amplified using the primers T4 DNapol For and T4 DNapol Rev as described earlier. The glucosylation reaction was carried out in NEB buffer 4 and included 740 ng of DNA (equivalent to 400 pmol of 5-hmdC, 16 pmol/ $\mu$ l final concentration), Uridine-5'-diphosphoglucose disodium salt (MP Biomedicals, LLC, Cat no. 101208) to a final concentration of 2  $\mu$ M, along with the appropriate amounts of GT, to a final concentration of 1.5 pmol/ $\mu$ l. The reactions were incubated at 37°C for 60 minutes, following which the products were directly digested using the Nucleoside digestion mix and quantified by LC-MS/MS (described below).

## Phage preparation and plaque assay

Phage stocks were prepared as lysates by infection of *E. coli* DH10B cultures in exponential phase (OD<sub>600</sub> 0.1-0.2) in LB with 10 mM MgSO<sub>4</sub> and 5 mM CaCl<sub>2</sub>. After the infected cultures reached an OD<sub>600</sub> lower than 0.1, they were centrifuged and filtered. The stocks were titrated and stored at 4 °C until use.

For plaque assay, preheated top agar (0.4% LB-agar) was mixed with 100 µl of overnight culture and phage preparation. The mixture was then poured onto preheated LB agar plates and incubated overnight at 37 °C. The plaques were counted and the plaque-forming unit per ml (PFU/ml) was calculated.

## Genome editing of phage T4

Insertions and deletions in bacteriophage T4 were performed as described earlier<sup>43</sup>. Briefly, overnight cultures of the recombinant strains, *E. coli* carrying plasmids constructed for deletions or insertions, were diluted and incubated at 37°C until an OD<sub>600</sub> between 0.15-0.25 was reached. The cultures were then infected with phage T4 preparations at an MOI of 0.05 and incubated at 37°C until complete lysis was observed (1-2 hours post infection). Serial dilutions of the supernatants were spotted on LB plates with the top layer containing the selection strain, *E. coli* DH10B with spacer containing pBZCas13b plasmid, to target the non-recombinant phage T4. The overnight incubated plates were then screened for the presence of the appropriate mutation by PCR.

Although replacement of the  $\alpha$ -GT coding sequence with  $\beta$ -*gt*, resulting in a recombinant phage with two copies of  $\beta$ -*gt*, is possible to construct using CRISPR-Cas based genome editing, subsequent replacement of the native  $\beta$ -GT coding sequence with  $\alpha$ -*gt* becomes unachievable, as the CRISPR spacer would target both  $\beta$ -*gt* copies. Instead, we developed an alternative strategy in which we replaced the  $\beta$ -GT coding sequence with a pseudo-open reading frame (pseudo-ORF) in wild-type T4 phage, to construct phage T4 LySE. The pseudo-ORF carried three protospacers, designed for efficient targeting by the CRISPR-Cas13b system (Figure S8A). Utilizing spacers designed for targeting  $\alpha$ -*gt* transcripts, we introduced the  $\beta$ -GT coding sequence downstream of the  $\alpha$ -*gt* promoter (referred to as T4 $\nabla$ P $_{\alpha\text{-}gt}::\beta\text{-}gt$ ). Subsequently, this recombinant phage was further engineered to replace the pseudo-ORF with the  $\alpha$ -GT coding sequence downstream of the *dCMP*/ $\beta$ -*gt* promoter (referred to as T4 $\nabla$ P $_{\alpha\text{-}gt}::\beta\text{-}gt$  P $_{\beta\text{-}gt}::\alpha\text{-}gt$ ) (Figure 4B, Figure S8B). Another phage with single replacements was also constructed, referred to as T4 $\nabla$ P $_{\beta\text{-}gt}::\alpha\text{-}gt$ , in addition to T4 $\nabla$ P $_{\alpha\text{-}gt}::\beta\text{-}gt$ .

## Quantification of nucleosides by LC-MS/MS

Total phage genomic DNA was extracted as described earlier<sup>43</sup>. Phage suspension was mixed with PEG and NaCl to a final concentration of 10% and 0.5 M respectively and incubated overnight at 4°C. Phages were pelleted by centrifugation at 12,000 rpm for 30 minutes at 4°C and resuspended in SM buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM MgSO<sub>4</sub>). Total DNA was extracted using Phenol-Chloroform followed by ethanol precipitation. 1 µg of total DNA was digested in 20 µl reaction volume using

the NEB Nucleoside digestion mix (#M0649S) according to the manufacturer's instructions.

Analysis of global levels of dC, 5mdC, 5hmdC, 5( $\alpha$ )-gmdC and 5( $\beta$ )-gmdC were performed on a Q Exactive Orbitrap Mass Spectrometer (ThermoFisher Scientific) equipped with an electrospray ionization source (H-ESI II Probe) coupled with an Ultimate 3000 RS HPLC (ThermoFisher Scientific). Digested DNA was injected into ThermoFisher Hypersil GOLD aQ chromatography column (100 mm \* 2.1 mm, 1.9  $\mu$ m particle size) heated at 30°C. The flow rate was set at 0.3 mL/min and run with an isocratic eluent of 1% ACN in water with 0.1% formic acid for 10 min. Parent ions were fragmented in positive ion mode with 10% normalized collision energy in parallel-reaction monitoring (PRM) mode. MS2 resolution was 17,500 with an AGC target of 2e5, a maximum injection time of 50ms, and an isolation window of 1.0 m/z. The inclusion list contained the following masses: dC (228.1), 5mdC (242.1), 5hmdC (258.1) and 5( $\alpha$ )-gmdC/5( $\beta$ )-gmdC (420.2). Extracted ion chromatograms of base fragments ( $\pm$ 5 ppm) were used for detection and quantification (112.0506 Da for dC; 126.0662 Da for 5mdC; 142.0609 Da for 5hmdC; 304.1133 Da for 5( $\alpha$ )-gmdC/5( $\beta$ )-gmdC). 5( $\alpha$ )-gmdC and 5( $\beta$ )-gmdC were distinguished by their retention times. Calibration curves were previously generated using synthetic standards (Clinisciences, France) in the ranges and 0.01–1 pmole.

## Virulence index

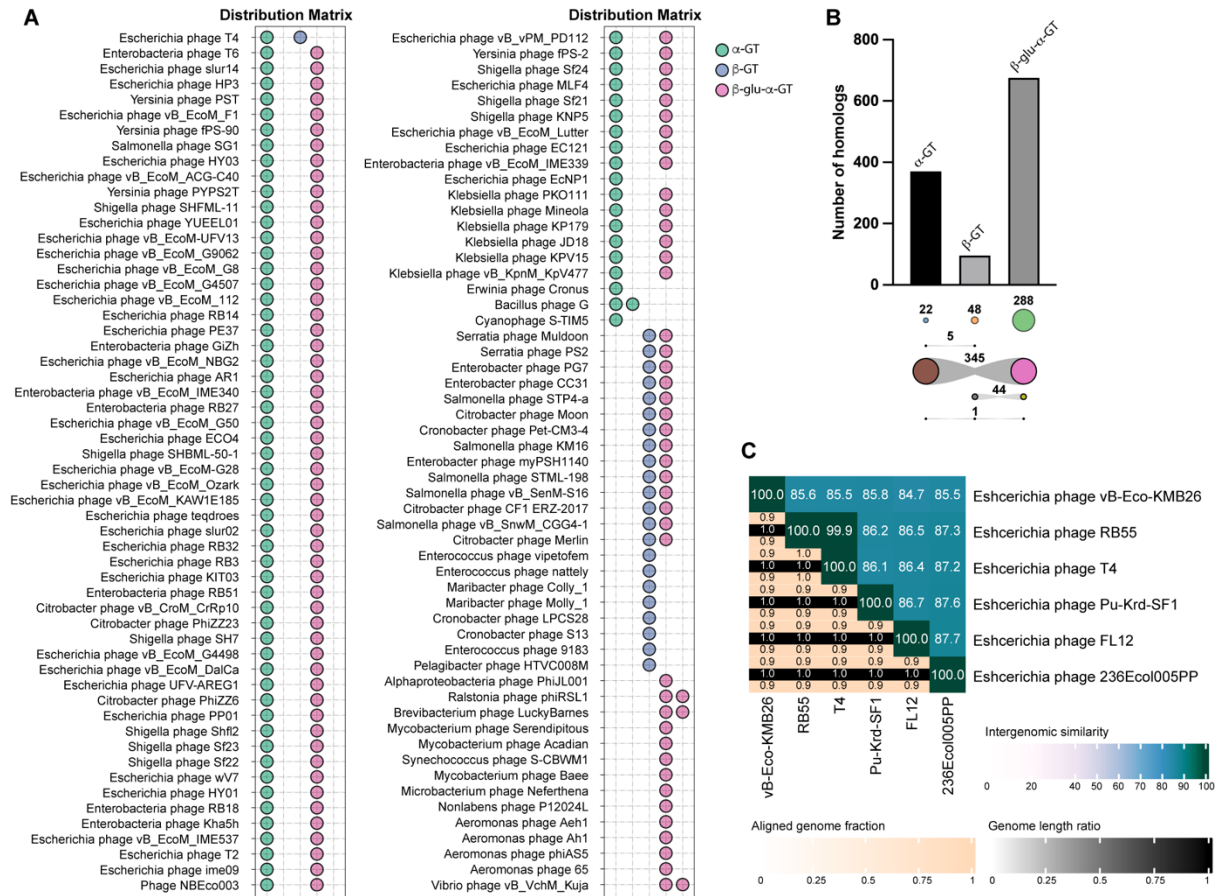
Virulence index was calculated as previously reported <sup>44</sup>. Briefly, 300  $\mu$ L of overnight culture of *E. coli* MG1655, encoding McrBC/McrA RM systems targeting 5-hmC DNA, was used to inoculate 30 mL of LB, which was then incubated for 1 hour at 37 °C while being agitated at 200 rpm. Once the cultures reached OD<sub>600</sub> between 0.15-0.25, 100  $\mu$ L was distributed in a 96-well plate. In parallel, pre-titrated phage lysates were serially diluted and 100  $\mu$ L of each dilution was added in to each well, covering a range of multiplicity of infections (MOIs) from 10<sup>2</sup> to 10<sup>-6</sup>. The Optical Density (OD<sub>600</sub>) of the cultures was measured every five minutes over 16 hours to obtain bacterial reduction curves.

The limit of integration was set at 7 hours post infection, since beyond this time point re-growth of phage-resistant mutants was observed. Area under the curve (AUC) of the bacterial reduction ones were obtained through the function MESS::auc depicted as local virulence curves (AUC vs log<sub>10</sub>(MOI)). The virulence index of each phage was obtained by integrating the virulence curves and normalizing. Virulence index values were compared by Kruskal Wallis Pairwise comparison and Dunn's all-pairs test.

## Efficiency of plating assay

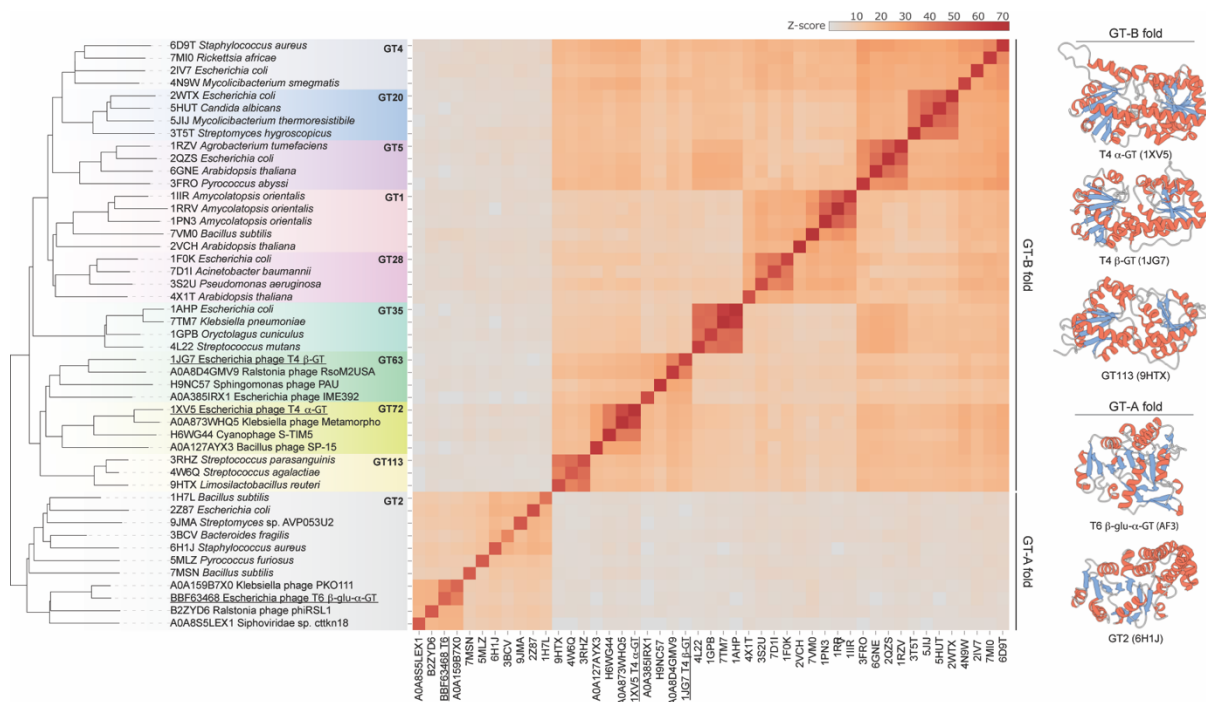
Infectivity of phages were estimated through efficiency of plating (EOP) assays. Ten-fold serial dilutions were prepared from a single stock of the phages. 10  $\mu$ l of phage dilution and 100  $\mu$ l of overnight bacterial culture was mixed with preheated top agar (0.4% LB-agar) and poured on LB agar plates. After overnight incubation at 37 °C, the number of plaques was counted and the PFU/ml was determined for both the test and control strains. EOP was calculated as the ratio of their PFU/ml values.

## MAIN FIGURES



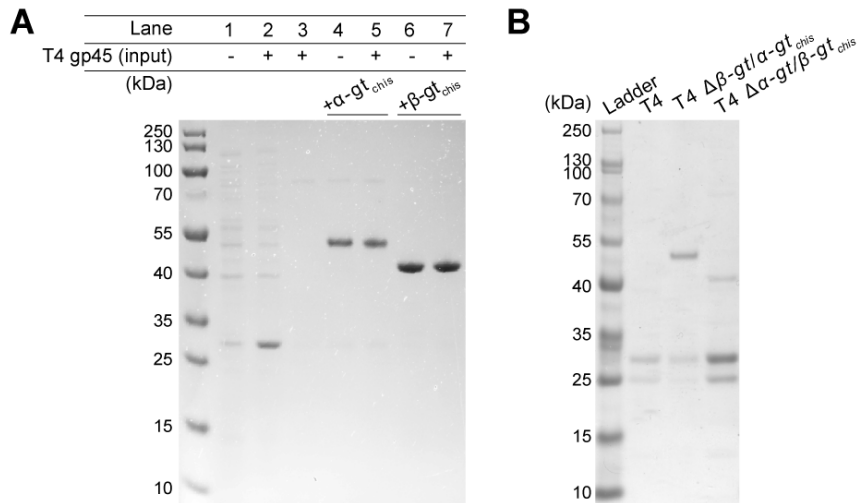
**Figure 1: Distribution of glucosyltransferases among phage genomes.**

**A.** Distribution of  $\alpha$ -GT,  $\beta$ -GT and  $\beta$ -glucosyl-HMC- $\alpha$ -GT homologs in phage genomes from the RefSeq database. **B.** Bar-plot showing the total number of  $\alpha$ -GT,  $\beta$ -GT and  $\beta$ -glucosyl-HMC- $\alpha$ -GT homologs in viruses (taxid:10239), in the non-redundant (nr) NCBI GenBank database. The number of GT homologs encoded either individually or in combination with other GT(s) is indicated below the x-axis. **C.** Heatmap showing pairwise sequence similarities between the six *Escherichia* phages that encode both  $\alpha$ -GT and  $\beta$ -GT. The top-right half of the heatmap, above the diagonal, displays the percentage of similarity between each genome pair. In the bottom-left half, three values are provided for each pair: the top and bottom values represent the aligned fraction for the phage genome pair corresponding to that row and column respectively. The middle value is the ratio of their genome lengths.



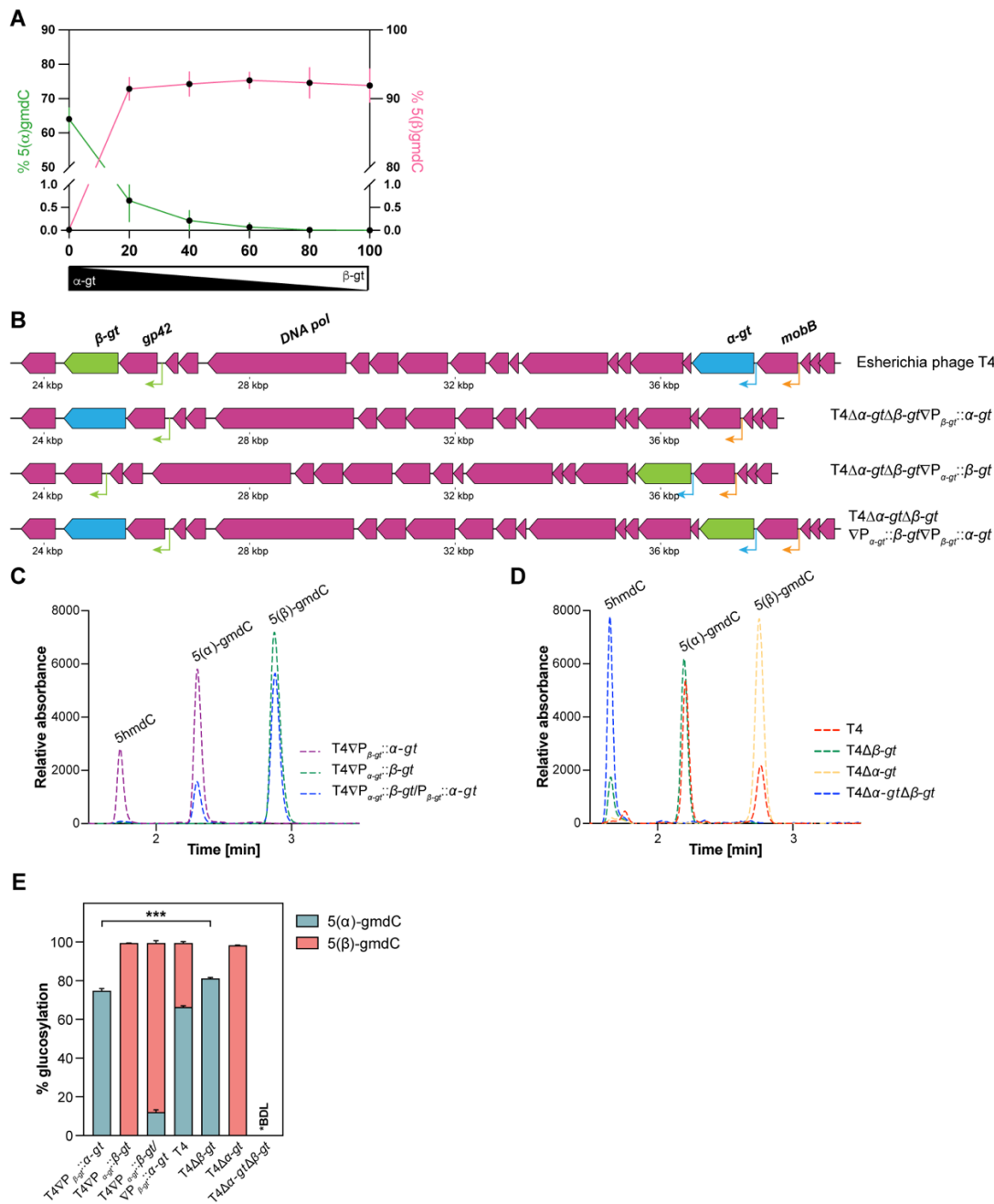
**Figure 2: Relationships between  $\alpha$ -GT,  $\beta$ -GT and  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT.**

A dendrogram and heatmap based on pairwise Z-scores, obtained using DALI, comparing T4  $\alpha$ -GT, T4  $\beta$ -GT and modelled structure of  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT along with representative cellular GTs from diverse families of the GT-A and GT-B superfamilies. Experimentally determined structures and structural models are indicated with the corresponding PDB and UniProt accession numbers, respectively.  $\alpha$ -GT and  $\beta$ -GT of phage T4 and  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT of phage T6 are underlined. Different GT families are highlighted with different background colors on the dendrogram. The color scale indicates the corresponding Z scores. On the right, the crystal structures of T4  $\alpha$ -GT (PDB: 1XV5), T4  $\beta$ -GT (PDB: 1JG7) and the modelled structure of  $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT (AF3) are compared to the structures of closest cellular representatives from the GT-B and GT-A fold superfamilies, GT113 (PDB: 9HTX) and GT2 (PDB: 6HI1J), respectively. Structural models are colored according to the secondary structure elements:  $\alpha$ -helices, red;  $\beta$ -strands, blue; random coil, grey.



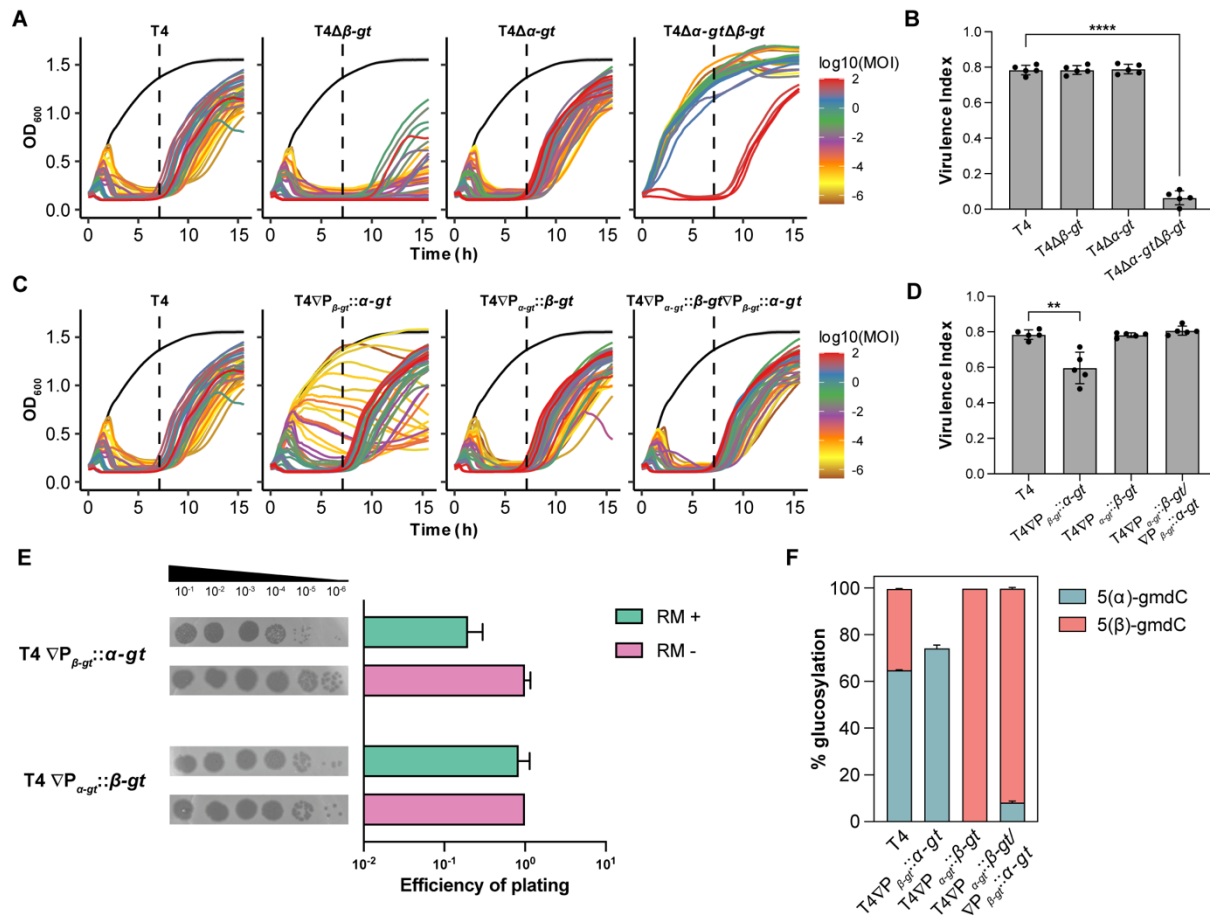
**Figure 3: Protein-protein interaction analysis of  $\alpha$ -GT and  $\beta$ -GT.**

**A.** *In vitro* pull-down of histidine tagged  $\alpha$ -GT or  $\beta$ -GT, mixed with un-tagged T4gp45. Lane 1 and lane 2 correspond to the total cell extract of *E. coli* BL21(DE3) carrying plasmid-borne un-tagged T4 gp45, before and after induction with IPTG. Lane 3 is a control IMAC purification of un-tagged gp45. Lanes 4 and 5 are IMAC purification of C-terminal histidine tagged  $\alpha$ -GT without or with fraction containing tag-free gp45 (shown in lane 2). Lanes 6 and 7 are IMAC purification of C-terminal histidine tagged  $\beta$ -GT without or with fraction containing tag-free gp45 (shown in lane 2). **B.** *In vivo* pull-down of  $\alpha$ -GT(chis) (lane 2) and  $\beta$ -GT(chis) (lane 3), based on phage-encoded endogenous expression, after infection of *E. coli* MG1655 with the respective phage T4 mutants. Lane 1 is IMAC purification control from *E. coli* MG1655 infected with the wild-type phage.



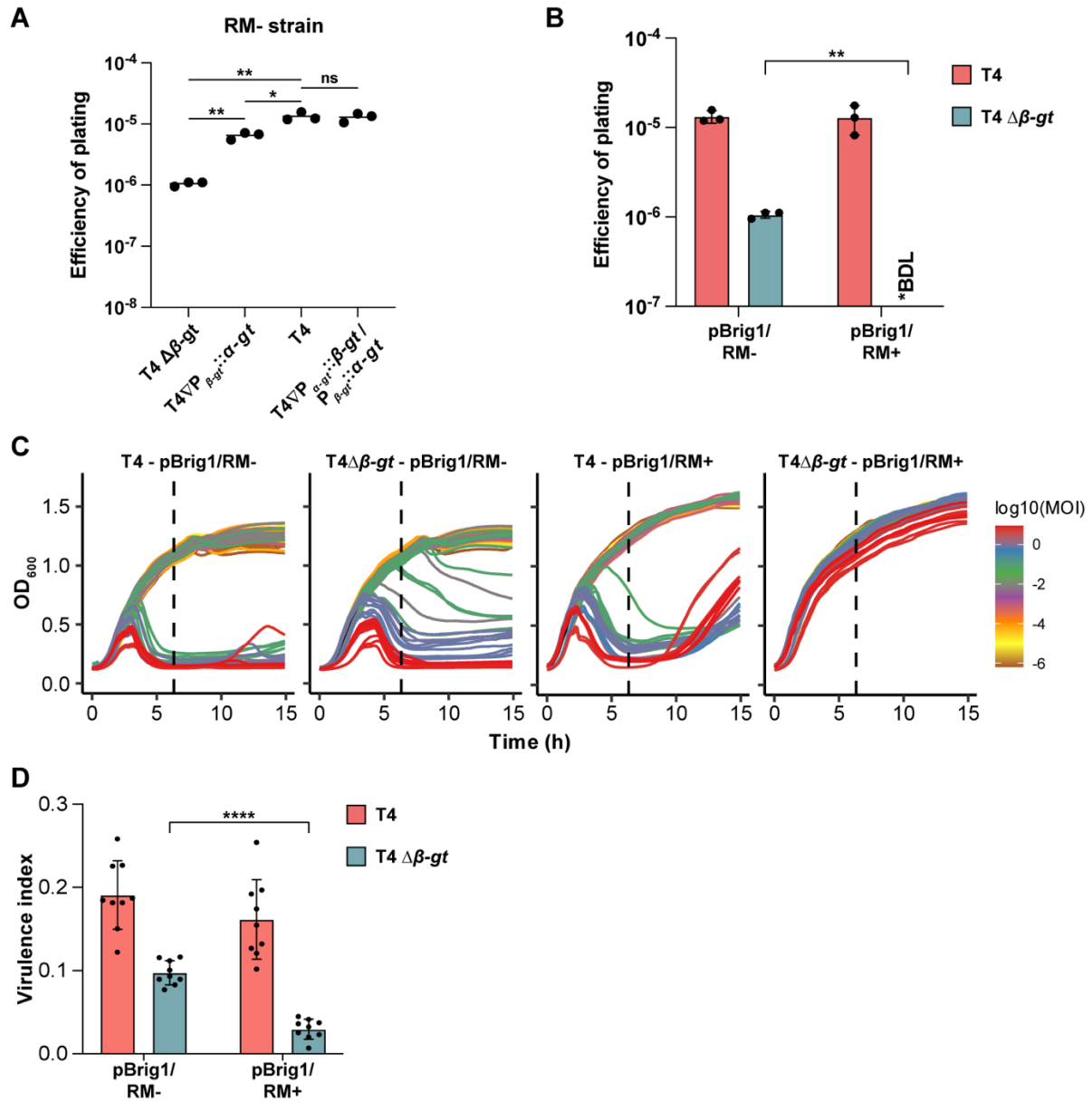
**Figure 4: Molecular mechanism behind the competitive glucosylation by  $\alpha$ -GT and  $\beta$ -GT.**

**A.** Competition assay to estimate the relative rate of glucosylation by  $\alpha$ -GT and  $\beta$ -GT on a fragment of DNA containing 5-hmdC. The average percentage of 5( $\alpha$ )-gmdC and 5( $\beta$ )-gmdC nucleotides were estimated by mass-spectrometry. The values shown are average of 6 independent experiments, represented as mean  $\pm$  SD. **B.** Illustration of phage mutants with genomic location of  $\alpha$ -gt and  $\beta$ -gt swapped between their native locations in the wild-type T4 phage. **C.** LC-MS/MS profile of the mutant phages, T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}$ , T4  $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$  and T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}/P_{\beta\text{-gt}}::\alpha\text{-gt}$  propagated in *E. coli* DH10B. The curves show the abundance of 5-hmdC, 5( $\alpha$ )-gmdC and 5( $\beta$ )-gmdC nucleotides. **D.** Mass-spectrometric LC-MS/MS profile of the mutant phages, T4 $\Delta\alpha\text{-gt}$ , T4 $\Delta\beta\text{-gt}$  and T4 $\Delta\alpha\text{-gt}\Delta\beta\text{-gt}$ , in comparison to the wild-type T4 phage. All phages were propagated in *E. coli* DH10B prior to estimation of their nucleotide composition. The curves show the abundance of 5-hmdC, 5( $\alpha$ )-gmdC and 5( $\beta$ )-gmdC nucleotides. **E.** Average percentage of 5( $\alpha$ )-gmdC and 5( $\beta$ )-gmdC in phages T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}$ , T4  $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$ , T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}/P_{\beta\text{-gt}}::\alpha\text{-gt}$ , T4, T4  $\Delta\alpha\text{-gt}$ , T4  $\Delta\beta\text{-gt}$  and T4  $\Delta\alpha\text{-gt}\Delta\beta\text{-gt}$ , as estimated by mass spectrometric analysis. The values shown are from three biological replicates, represented as mean  $\pm$  SD. A two-tailed unpaired t-test of the percentage values of 5( $\alpha$ )-gmdC was used to calculate the *P* values; \*\*\**P* < 0.001.



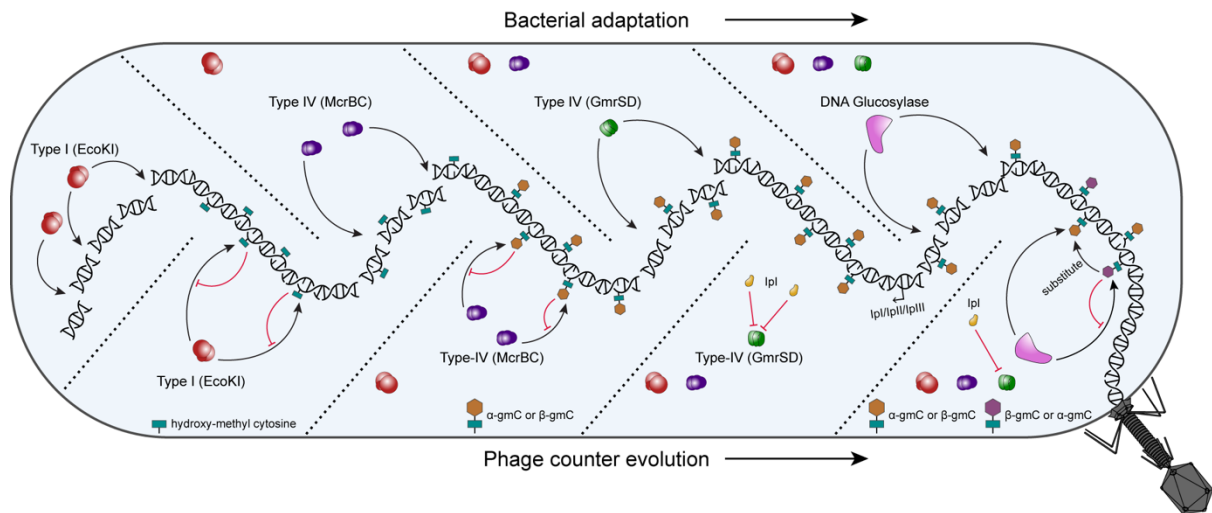
**Figure 5: Virulence estimation of phages T4 wild-type, T4  $\Delta\alpha$ -gt, T4  $\Delta\beta$ -gt and T4  $\Delta\alpha$ -gt $\Delta\beta$ -gt in *E. coli* MG1655.**

**A.** Growth curves of *E. coli* MG1655 infected with phages T4, T4 $\Delta\alpha$ -gt, T4 $\Delta\beta$ -gt and T4 $\Delta\alpha$ -gt $\Delta\beta$ -gt, over a range of MOIs between  $10^2$  to  $10^{-6}$ . Uninfected culture is represented in black. **B.** Virulence index of the respective phages, as calculated from the growth curves shown in (A), up to 7 hours post-infection. Data shown are mean of four biological replicates, represented as mean value  $\pm$  SD. A two-tailed unpaired t-test of the virulence index values was used to calculate the *P* values; \*\*\*\**P* < 0.0001. **C.** Growth curves of *E. coli* MG1655 infected with phages T4  $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$ , T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}$  and T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}/P_{\beta\text{-gt}}::\alpha\text{-gt}$  and the wild-type T4, over a range of MOIs between  $10^2$  to  $10^{-6}$ . Uninfected culture is represented in black. **D.** Virulence index of the respective phages, as calculated from the growth curves shown in (c), up to 7 hours post-infection. Data shown are mean of four biological replicates, represented as mean value  $\pm$  SD. A two-tailed unpaired t-test of the virulence index values was used to calculate the *P* values; \*\**P* < 0.01. **E.** Spot assay of serially diluted phages, T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}$  and T4  $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$  on *E. coli* DH10B (RM-) or *E. coli* MG1655 (RM+) background. Efficiency of plating (EOP) of phage infection, as estimated from the spot assay, are from four biological replicates, represented as mean  $\pm$  SD. **F.** Average percentage of 5( $\alpha$ )-gmdC and 5( $\beta$ )-gmdC in phages T4, T4 $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}$ , T4 $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$  and T4 $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}/P_{\beta\text{-gt}}::\alpha\text{-gt}$ , as estimated by mass spectrometric analysis. All phages were propagated in *E. coli* MG1655 prior to estimation of their nucleotide composition. The values shown are from two biological replicates, represented as mean  $\pm$  SD.



**Figure 6: Role of  $\beta$ -GT in protection against Brig1.**

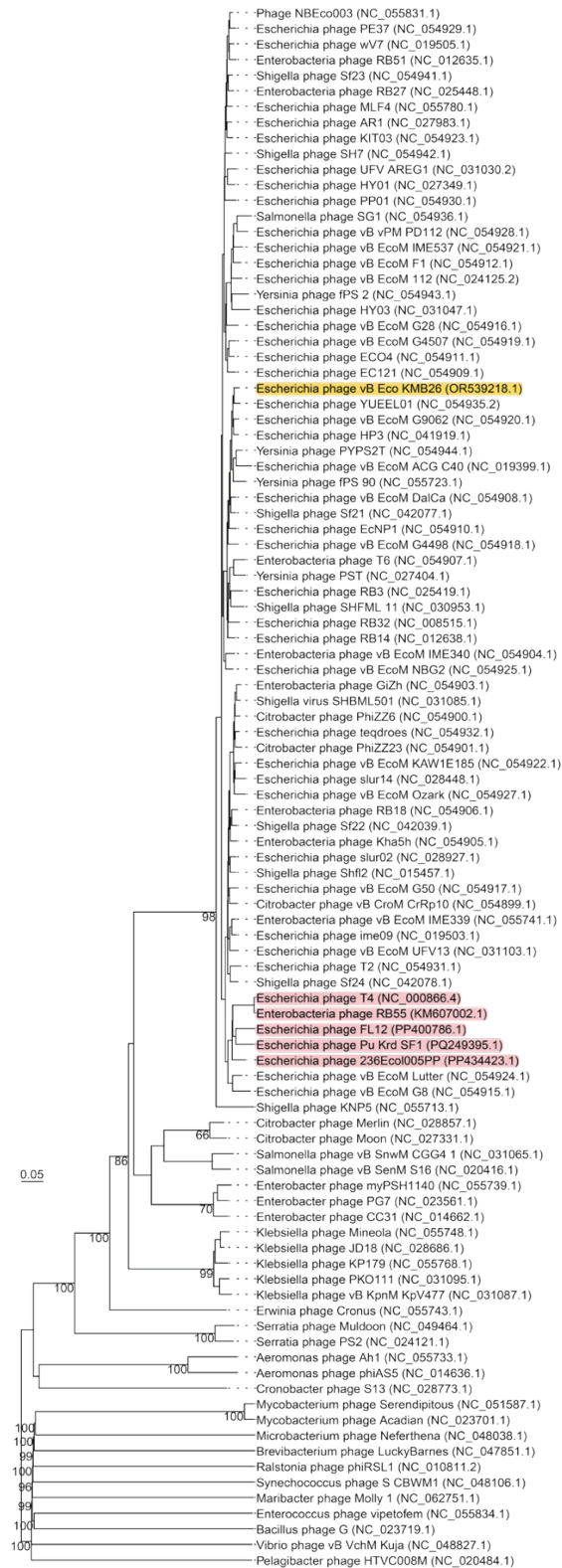
**A.** The frequency of Brig1 escapers, estimated as the efficiency of plating (EOP) of phages T4 $\Delta\beta$ -gt, T4VP  $\alpha$ -gt:: $\beta$ -gt, T4 and T4VP  $\alpha$ -gt:: $\beta$ -gt/ $P_{\beta$ -gt:: $\alpha$ -gt on *E. coli* DH10B carrying a *brig1* expressing plasmid. Data shown are mean of three biological replicates, the horizontal line represents the average value. A two-tailed unpaired t-test of the EOP values was used to calculate the *P* values; \*\**P* < 0.01, \**P* < 0.05, ns – not significant. **B.** Frequency of brig1 escapers (efficiency of plating of phages) T4 and T4 $\Delta\beta$ -gt in *E. coli* strains DH10B (RM-) and MG1655 (RM+), both carrying a *brig1* expressing plasmid. The results are represented in terms of EOP. \*BDL = Below Detection Limit. Data shown are mean of three biological replicates. A two-tailed unpaired t-test of the EOP values was used to calculate the *P* values; \*\**P* < 0.01. **C.** Growth curves of *E. coli* EC100 (RM-) and *E. coli* MG1655 (RM+) infected with phages T4 or T4 $\Delta\beta$ -gt over a range of MOIs between  $10^2$  to  $10^{-6}$ . Uninfected culture is represented in black. **D.** Virulence index of the respective phages, as calculated from the growth curves shown in (C), up to 7 hours post-infection. Data shown are mean of nine biological replicates, represented as mean value  $\pm$  SD. A two-tailed unpaired t-test of the virulence index values was used to calculate the *P* values; \*\*\*\**P* < 0.0001.



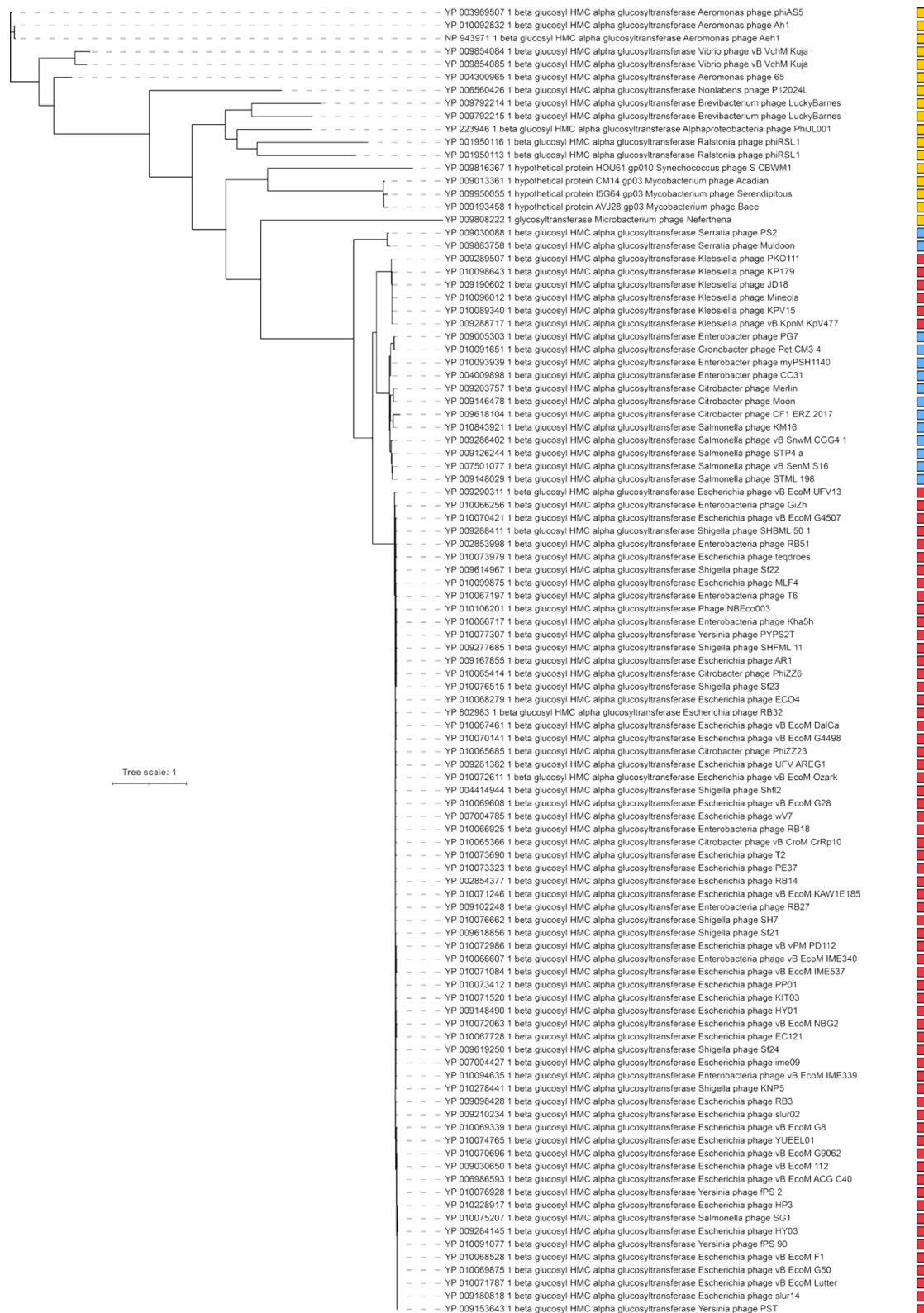
**Figure 7: Model of T4 counter evolution to overcome host defense systems.**

From left to right, upon evolution of the type I RM systems, phages likely adapted a host enzyme to hydroxy methylate the cytosine base in their genome. In response, host evolved the Type IV McrBC RM system, which was in turn countered by hypermodification of the 5-hmC in the phage genome. The hypermodification was in turn countered by Type IV GmrSD RM systems, which are suppressed by the phage protein Ipl. Following this, evolution of DNA glycosylases such as Brig1, specific to  $\alpha$ -glucosylation, are countered by acquisition of  $\beta$ -GT enabling viability of Brig1 escapers.

## SUPPLEMENTAL FIGURES

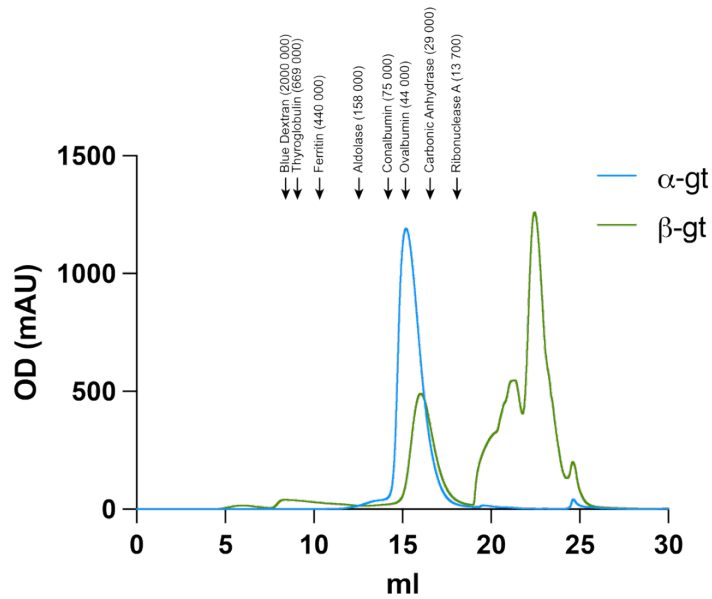


**Figure S1: Phylogenetic tree based on the whole-genome comparison of phages encoding GTs.** The phages highlighted in red encode both primary GTs and the phage highlighted in yellow encodes all three GTs. The Genome-BLAST Distance Phylogeny pseudo-bootstrap support values from 100 replicates are shown above the branches. The tree was generated using the VICTOR web service (<https://ggdc.dsmz.de/victor.php>)<sup>49</sup>, with branch lengths scaled according to the nucleotide distance formula (D0).



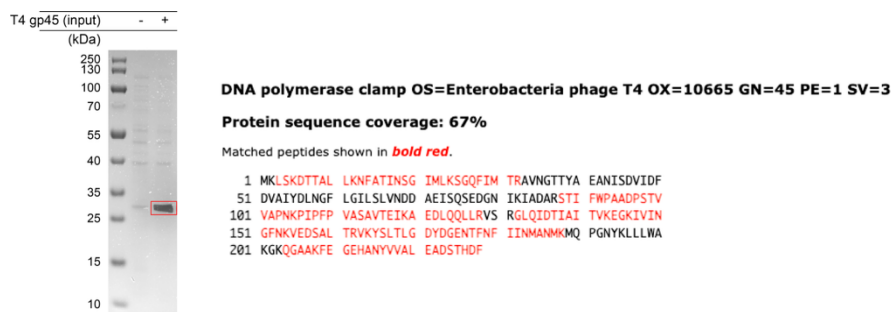
**Figure S2: Maximum likelihood phylogeny of the secondary glucosyltransferase ( $\beta$ -glucosyl-HMC- $\alpha/\beta$ -GT).**

The box adjacent to each leaf label indicates co-occurrence with either  $\alpha$ -GT (red) or  $\beta$ -GT (blue), or lack of association with a primary GT (yellow). The tree was generated using the NGphylogeny.fr<sup>52</sup> web service.



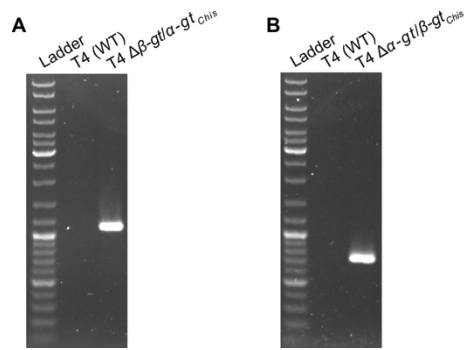
**Figure S3: Size-Exclusion Chromatography of C-terminally histidine tagged  $\alpha$ -GT and  $\beta$ -GT.**

Elution profiles of histidine tagged  $\alpha$ -GT (blue) and  $\beta$ -GT (green) obtained upon separation of the proteins on a Superdex 200 Increase 10/300 GL column. The absorbance was monitored at 280 nm. The peak elution position of each standard protein is indicated as an arrow, along with their theoretical molecular weights (Da) in brackets. The molecular weights of  $\alpha$ -GT and  $\beta$ -GT, based on their peak elution volumes, were estimated to be 48.36 kDa and 33.61 kDa, respectively.



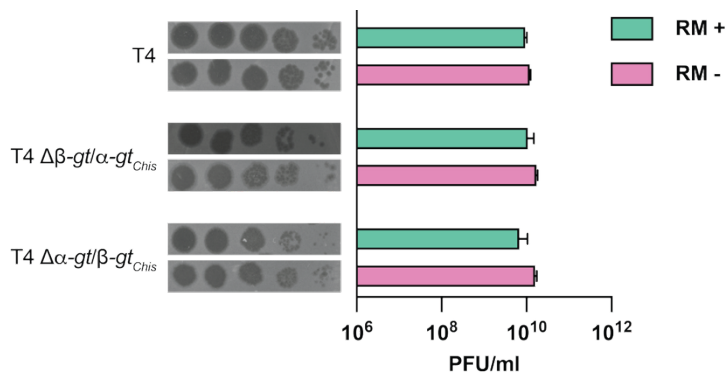
**Figure S4: Validation of un-tagged T4 gp45 expression in *E. coli* BL21(DE3).**

Lane 1 and lane 2 correspond to the total cell extract of *E. coli* BL21(DE3) with un-tagged T4 gp45, before and after induction with IPTG. The induced protein (intense band) corresponding to the size of the un-tagged sliding clamp (T4 gp45, 24858.32 Da) monomer, was excised from the gel (shown inside the red box) and the protein's identity confirmed by LC-MS/MS (right panel).



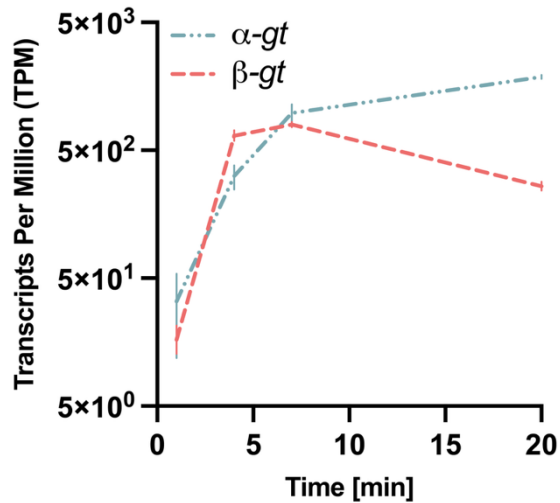
**Figure S5: Sequence confirmation of the histidine-tag at the C-terminus of the GTs in the phage mutants T4  $\Delta\beta$ -gt/ $\alpha$ -gt<sub>Chis</sub> and T4  $\Delta\alpha$ -gt/ $\beta$ -gt<sub>Chis</sub>.**

**A.** PCR confirmation of the histidine tag coding sequence downstream of  $\alpha$ -gt in the phage T4  $\Delta\beta$ -gt/ $\alpha$ -gt<sub>Chis</sub> using primers T4\_ $\alpha$ -gt\_chk\_in\_rev and Histag-C\_chk\_rev (listed in Table S4). T4 wild-type was used as control. **B.** PCR confirmation of the histidine tag coding sequence downstream of  $\beta$ -gt in the phage T4  $\Delta\alpha$ -gt/ $\beta$ -gt<sub>Chis</sub> using primers T4\_ $\beta$ -gt\_chk\_in\_rev and Histag-C\_chk\_rev (listed in Table S4). Wild-type T4 was used as control.



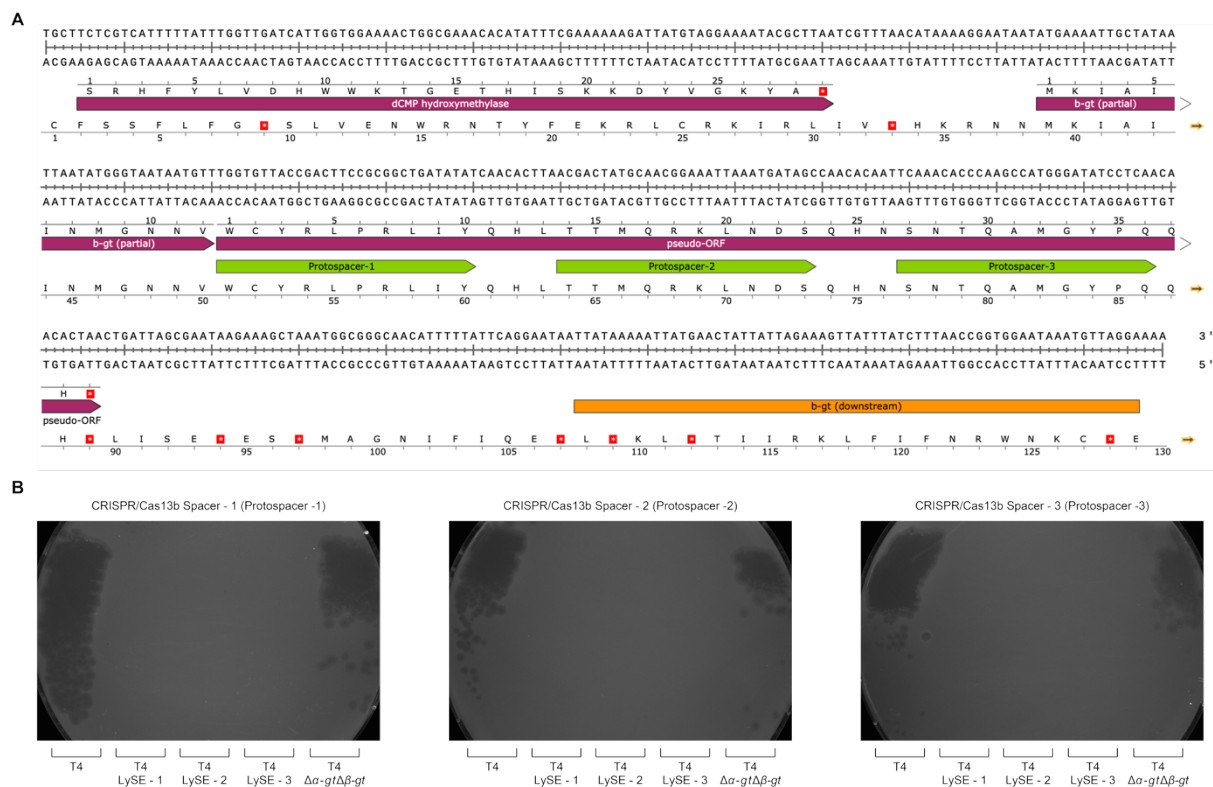
**Figure S6: Estimation of the infectivity of phages T4  $\Delta\beta$ -gt/ $\alpha$ -gt<sub>Chis</sub> and T4  $\Delta\alpha$ -gt/ $\beta$ -gt<sub>Chis</sub> in the presence of RM systems.**

Phage spot assays demonstrating the infectivity of the phage mutants T4  $\Delta\beta$ -gt/ $\alpha$ -gt<sub>Chis</sub> and T4  $\Delta\alpha$ -gt/ $\beta$ -gt<sub>Chis</sub> in comparison to the wild-type phage. The spot assays were performed on *E. coli* MG1655 (RM+) and *E. coli* DH10B (RM-). The phage titers (PFU/ml), estimated from the spot assay, are from three biological replicates, represented as mean  $\pm$  SD.



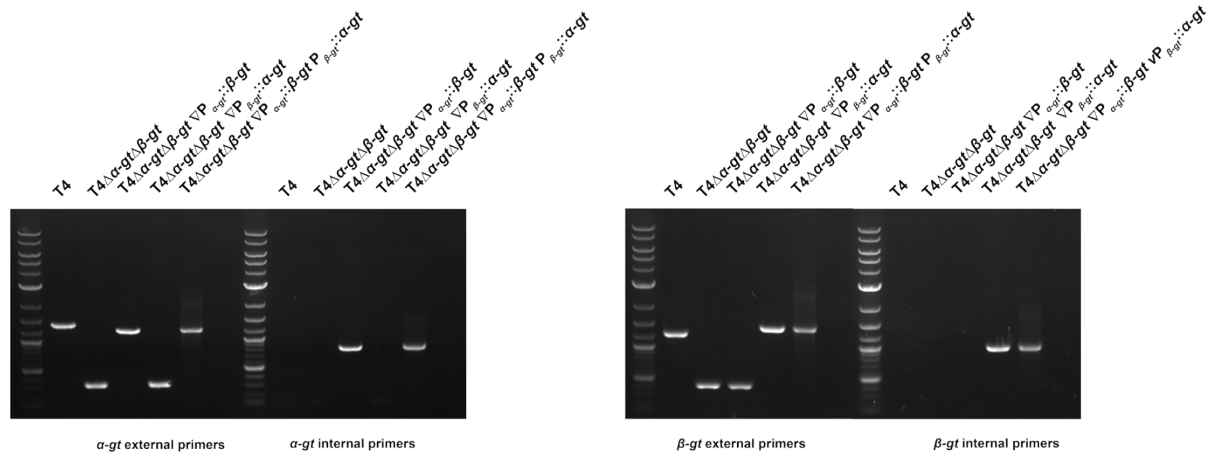
**Figure S7: Transcript levels of  $\alpha$ -gt and  $\beta$ -gt during the T4 infection cycle.**

Average of the three replicates, each normalized with the total reads corresponding to phage transcripts, was plotted against time post infection. Raw data was obtained from Wolfram-Schauerte and Pozhydaieva et al., 2022<sup>42</sup>.



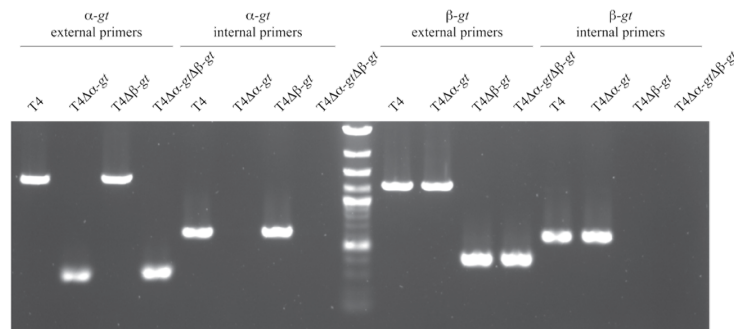
**Figure S8: Genome editing strategy for swapping genomic locations of  $\alpha$ -gt and  $\beta$ -gt.**

**A.** The coding region including the pseudo-ORF, in-frame, with the partial  $\beta$ -gt sequence is shown. The protospacers for Cas13b targeting are shown as arrows (in green). The upstream (*gp42* and partial  $\beta$ -gt) and downstream ( $\beta$ -gt downstream) sequences, necessary for recombination into the wild-type T4, are shown in purple and orange respectively. **B.** Susceptibility of the phage T4 LySE (3 clonal copies) to Cas13b is shown. Left panel: Cas13b spacer complimentary to protospacer-1, middle panel: spacer complimentary to protospacer-2 and right panel: spacer complimentary to protospacer-3. T4 wild-type and T4 $\Delta\alpha$ -gt $\Delta\beta$ -gt, both lacking the pseudo-ORF, are shown as controls.



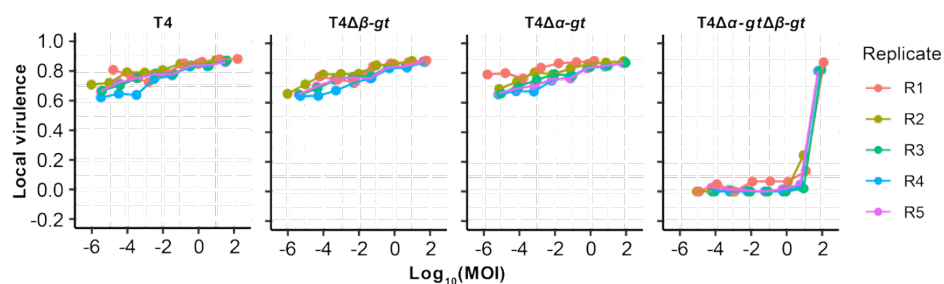
**Figure S9: PCR confirmation of the  $\alpha$ -gt and  $\beta$ -gt genomic location swapping in different phage mutants.**

The positional swapping of  $\alpha$ -gt and  $\beta$ -gt, along with their presence and absence, in the phage mutants T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}$ , T4  $\nabla P_{\beta\text{-gt}}::\alpha\text{-gt}$  and T4  $\nabla P_{\alpha\text{-gt}}::\beta\text{-gt}$   $P_{\beta\text{-gt}}::\alpha\text{-gt}$  was confirmed with internal primers and external primers for  $\alpha$ -gt and  $\beta$ -gt (listed in [Table S4](#)). Phage T4 was used as PCR control.



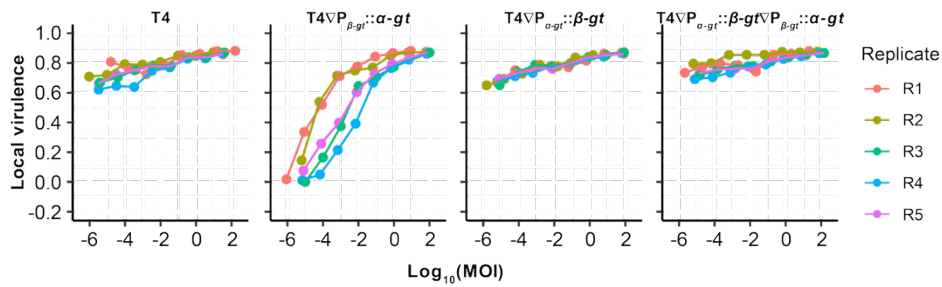
**Figure S10: PCR confirmation of the desired deletion in T4 $\Delta\alpha$ -gt $\Delta\beta$ -gt.**

The deletion of  $\alpha$ -gt and  $\beta$ -gt in the mutant phage T4 $\Delta\alpha$ -gt $\Delta\beta$ -gt is confirmed with internal primers and external primers (listed in [Table S4](#)) for  $\alpha$ -gt and  $\beta$ -gt. Mutant phages T4 $\Delta\alpha$ -gt and T4 $\Delta\beta$ -gt, constructed earlier<sup>43</sup>, along with the phage T4 were used as PCR controls.



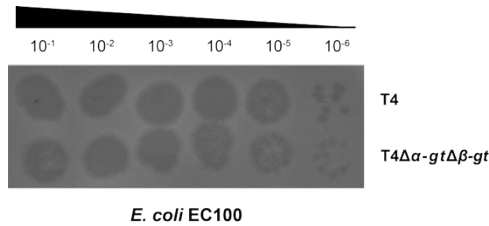
**Figure S11: Local virulence corresponding to phages T4, T4 $\Delta\alpha$ -gt, T4 $\Delta\beta$ -gt and T4 $\Delta\alpha$ -gt $\Delta\beta$ -gt upon infection in *E. coli* MG1655.**

The local virulence for each MOI was calculated from the corresponding growth curve in [Figure 5A](#).



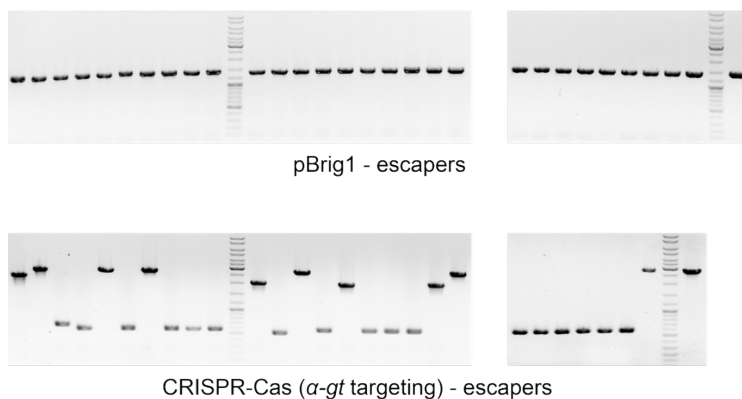
**Figure S12: Local virulence of phages T4  $\nabla P_{\alpha-gt}::\beta-gt$ , T4  $\nabla P_{\beta-gt}::\alpha-gt$  and T4  $\nabla P_{\alpha-gt}::\beta-gt/P_{\beta-gt}::\alpha-gt$  in comparison to wild-type T4, upon infection in *E. coli* MG1655.**

The local virulence for each MOI was calculated from the corresponding growth curve in [Figure 5C](#).

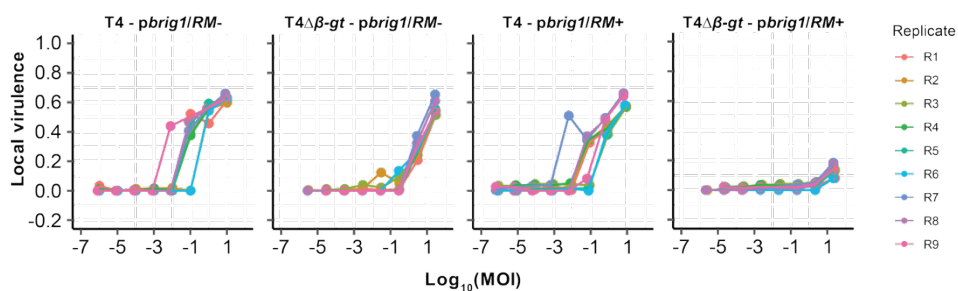


**Figure S13: *E. coli* EC100 strain lacks type IV RM systems that target 5-hmC.**

Spot assay of serially diluted wild-type T4 and T4  $\Delta\alpha-gt\Delta\beta-gt$  on a lawn of *E. coli* EC100.



**Figure S14: PCR analysis of the  $\alpha-gt$  locus in phage T4 escapers generated post targeting by Brig1 (upper panel) or the CRISPR-Cas13b system (with a spacer complementary to the transcript of  $\alpha-gt$ ).**



**Figure S15: Local virulence corresponding to phage T4 and T4 $\Delta\beta-gt$  upon infection in *E. coli* EC100 (RM-) and *E. coli* MG1655 (RM+).**

The local virulence for each MOI was calculated from the corresponding growth curve in [Figure 6C](#).

## SUPPLEMENTAL TABLE LEGENDS

**Table S1.** Identification and analysis of  $\alpha$ -GT,  $\beta$ -GT and  $\beta$ -glucosyl-HMC- $\alpha$ -GT homologs in phage genomes from the RefSeq and non-redundant (nr) NCBI GenBank database.

**Table S2.** Proteomic analysis of histidine-tagged  $\alpha$ -GT and  $\beta$ -GT pulldown samples.

**Table S3.** List of mutations in the  *$\alpha$ -gt* locus from T4 escapers of Brig1 immunity.

**Table S4.** List of bacterial and phage strains, plasmids, and oligonucleotides used in this study.