# A Siphophage Like No Other: Deciphering the D3 Tail Tip

Tatiana Lenskaia[1,*] and Alan R. Davidson[1,2]

[1] Department of Molecular Genetics, University of Toronto, Toronto, Canada
[2] Department of Biochemistry, University of Toronto, Toronto, Canada
* Corresponding author: t.lenskaia@utoronto.ca

## Abstract

In the microbial world, phages often guard their secrets closely - but none more so than D3, a siphophage which tail tip architecture has long eluded explanation. This study embarks on a genomic journey to uncover the structure and function of the D3 tail tip, drawing on computational sleuthing techniques including comparative genomics, Hidden Markov Models (HMMs), structural and sequence homology, and gene synteny mapping. Through layered inference and pattern recognition, a picture begins to emerge: D3 harbors a distinct set of tail tip proteins – Distal Tail (DTN), Tail Hub (THN), and Central Fiber (CF) – configured in many ways unlike any previously solved structure. Intriguing clues, such as a uniquely extended DTN with a hinge domain and variably sized CFs in related phages, suggest alternative mechanisms for host specificity.

## Keywords

Tail tip, comparative genomics, phage mosaicism, host specificity

## 1. Introduction

It began, as many good mysteries do, with something that did not quite fit. Amid the expanding catalog of siphophages – those elegantly built bacterial viruses armed with long, flexible tails - one phage stood apart. Phage D3, infecting a Gram-negative host, *Pseudomonas aeruginosa*, carried the genomic markings of a typical siphophage but resisted identification where it mattered most: at its tail tip, the molecular spearhead of infection. While model systems like phage lambda and T5 have long served as blueprints for tail architecture in siphophages infecting Gram-negative hosts, D3 hinted at a different story – an evolutionary divergence written not in grand rearrangements but in genuine structural innovations. Its genome spoke in cryptic modularity: genes that resembled known tail tip components, but were fused, shuffled, or somewhat reimagined. There was no roadmap to follow, no clear evolutionary trail. This study is the chronicle

of that investigation - a forensic reconstruction of how D3 assembles its uniquely modular tail tip, and how that structure reshapes our understanding of siphophage diversity, evolution, and host adaptation.

Siphophages typically employ a modular tail architecture for host recognition, attachment, and DNA ejection. The tail tip is a key determinant of host range, composed of diverse tail tip proteins such as the distal tail protein (DTN), tail hub protein (THN), and central fiber (CF). Previous structural studies that analyzed tail tip organization before and after DNA ejection (e.g., phage Lambda and T5) have revealed conserved core architectural solutions, yet substantial variability exists in the actual modules that fulfill the necessary functions.

The canonical model of siphophage tail tip architecture is epitomized by phage lambda, whose infection mechanism has been studied in molecular detail for decades. In lambda, the distal end of the tail is capped by the gpJ protein, a trimeric structure that functions as a terminal tail tip component with the receptor-binding domain responsible for engaging the *Escherichia coli* outer membrane protein LamB (Ge & Wang, 2024). Structural studies have revealed that gpJ adopts a long, coiled-coil architecture extending from the tail tube, terminating in a globular C-terminal domain that mediates host binding. Notably, phage lambda lacks a large baseplate or extended fibers, distinguishing it from other siphophages and underscoring the minimalism of its tail tip machinery (Wang et al., 2024). This organization has served as a benchmark for understanding siphophage infection systems, yet it represents only one solution among many. As comparative genomic data have accumulated, it has become increasingly clear that this classical model may obscure a broader diversity of tail tip strategies in siphophages, many of which remain structurally uncharacterized.

Phage T5 offers a contrasting blueprint for long tail tip organization among siphophages, showcasing an elaborate and modular infection apparatus (Linares et al., 2023). Central to this assembly is the distal tail protein pb9, which acts as a tail needle, capping the tail tube and contributing to DNA ejection. Attached to this is the tail hub protein pb3 and a set of lateral tail fibers, notably pb4 and pb5, that extend from the tip complex to mediate host interaction. Pb5, the primary receptor-binding protein, interacts with FhuA, a TonB-dependent outer membrane transporter in *E. coli*, initiating DNA injection in a two-step process that first delivers early genes for host takeover. Cryo-electron microscopy has revealed that the T5 tail tip includes an extended baseplate-like structure. In this light, phages such as D3 - distinct yet unsolved until recently - may represent entirely new paradigms that depart from both the lambd and T5 models.

Phage R4C introduces yet another distinctive variant in siphophage tail tip architecture (Huang et al., 2023). Phage R4C, a siphophage infecting Roseobacter species, enriches the structural repertoire of siphophage tail tips by presenting a multi-layered and architecturally sophisticated tail tip complex. This composite assembly reflects a finely tuned system for host recognition, with tail fibers likely mediating interactions with complex or variable surface structures typical of marine bacteria. Genomic analysis supports this complexity: the tail tip genes reside in a distinct syntenic block, consistent with a modular assembly strategy observed in several long-tailed phages, including D3.

The classical models of siphophage infection – exemplified by the streamlined tail tip of lambda and the more elaborate, baseplate-equipped architecture of T5 – have shaped our understanding of how long-tailed phages engage their hosts. More recently, phages like R4C have

revealed a new level of modularity and complexity, with extended receptor-binding elements and structural innovations adapted to specific ecological niches. Yet even within this growing landscape, phage D3 stands apart. Its gene organization, protein domain structure, and host range patterns suggest a tail tip strategy unlike any previously characterized. With the newly resolved D3 structure now in hand, we uncover a distinct architectural and functional paradigm – one that expands the known repertoire of siphophage tail tips and redefines what these elegant machines are capable of.

## 2. Results

### 2.1. Overview of Unique Features for the D3 Tail Tip

In contrast to the previously solved models described above, the D3 phage tail tip exhibits both structural innovation and evolutionary divergence. For example, its DTN is significantly larger than that of phage lambda and contains an additional domain acting as a flexible hinge. This hinge domain enables docking of a highly variable receptor-binding protein (RBP), allowing for adaptation to a variety of Gram-negative bacterial hosts. Here, we unveil the remarkable structure of the D3 phage tail tip, a molecular apparatus that redefines the known boundaries of siphophage architecture.

Through comparative genomics and evolutionary analysis, we reveal a tail tip that is neither derivative nor transitional – but boldly distinct from the solved siphophage tail tip structure of lambda, T5, and R4C (**Fig.1**). D3 encodes a tail tip module that includes uniquely arranged and structurally distinct families of DTN, THN, and CF proteins. These proteins differ in both domain architecture and interaction modes compared to their counterparts in other siphophages with the solved structures.

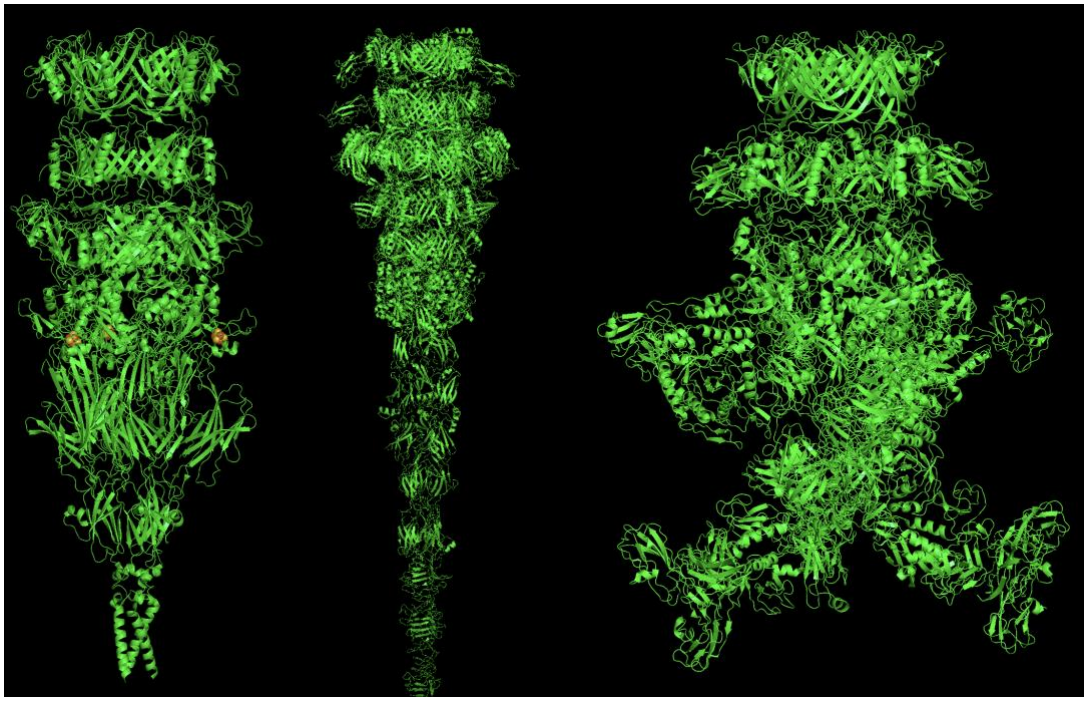

**Figure 1.** Comparison of the solved structure of siphophage lambda (PDB: 8K35, left), T5 (PDB: 7ZQB, center), and R4C (PDB: 8GTC, right).

### 2.2. Extended DTN with Hinge Domain

One of the most striking features of the D3 phage tail tip is its elongated DTN protein, which, unlike the compact form found in phage Lambda, is expanded by a hinge-like C-terminal domain. This structural innovation forms the anchor point of a novel modular interface that enables the dynamic attachment of receptor-binding proteins (RBPs), facilitating host adaptation without requiring wholesale genetic replacement. This flexibility suggests a previously unrecognized mechanism of host range expansion driven by structural plasticity. The D3 tail tip thus offers a compelling glimpse into the evolutionary creativity of siphophage architecture.

For DTN, there were 83 DTN proteins detected by the DTN-D3 HMM hits with lengths range from 72 to 372 aa, with length about 150 aa on average. The DTN proteins were separated into 8 clusters by MMseq2 (0 0). Cluster 1 contains 4 closely related DTN proteins of length 150-170 aa from Salmonella phages clustered with Salmonella phage SE2. These phages have a relatively short CF ranging in length from 600 to 850 aa and a spike protein (SF) of length 684 aa adjacent to the CF ORF. The spike proteins aligned almost perfectly with a few amino acid changes defining their host specificity. The structural alignment between hexamers of two distant DTNs from Cluster 1, Salmonella phage SE2 (147 aa) in cyan and Salmonella phage vB_SenS-Ent1 (166 aa) in green are shown on **Fig. 2.** The predicted interactions between 1 DTN and 3 SFs of Salmonella phage vB_SenS-Ent1 are shown on **Fig. 3.**

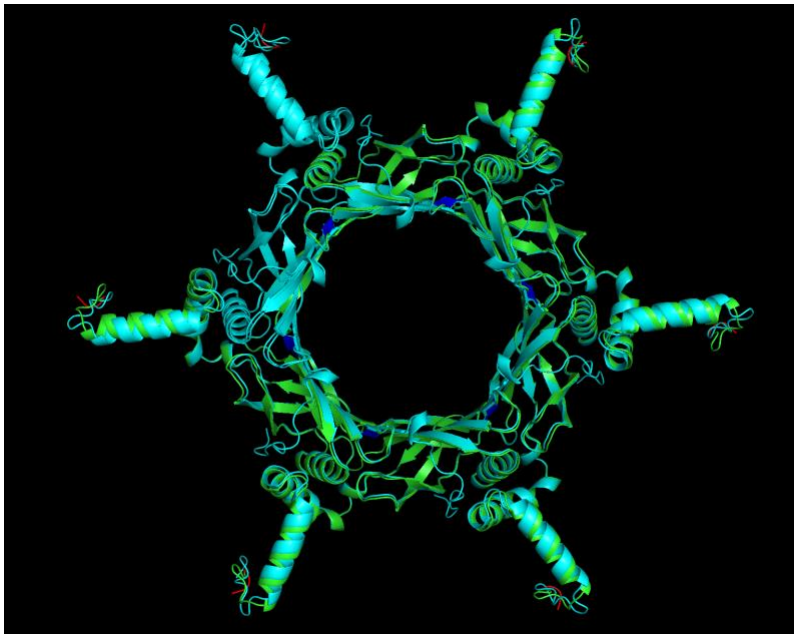

**Fig. 2.** The structural alignment between hexamers of two different DTNs from Cluster 1, Salmonella phage SE2 (147 aa) in green and Salmonella phage vB_SenS-Ent1 (166 aa) in cyan. The N-terminus of DTN SE2 is highlighted in blue, and the C-terminus – in red. The DTN monomer of vB_SenS-Ent1 has an extra 19 aa in the unstructured region at the N-terminus that SE2 does not have.

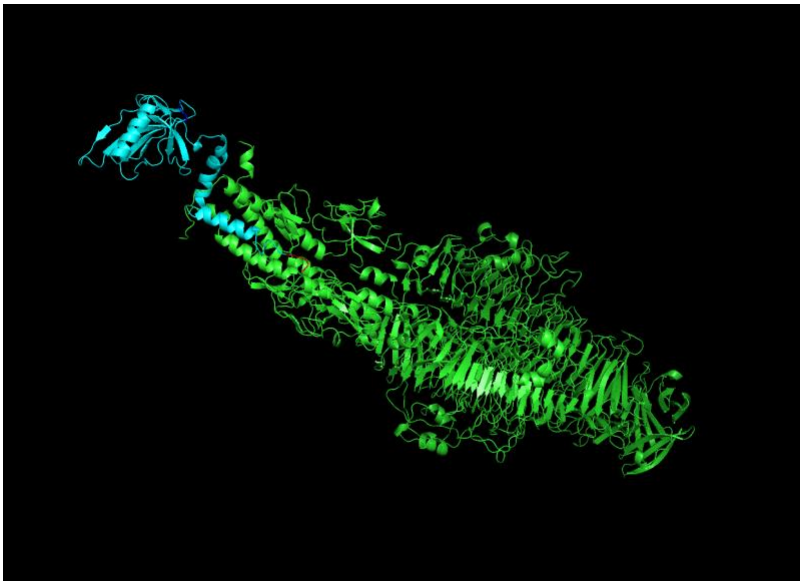

**Fig. 3.** The predicted interactions between 1 DTN (cyan) and 3 SFs (green) of Salmonella phage vB_SenS-Ent1 by AF3. The N-terminus of DTN is highlighted in blue, and the C-terminus – in red.

Cluster 2 contained 17 DTNs with all except one infecting *Acinetobacter baumannii*. The remaining one was a prophage from *Achromobacter xylosoxidans*. The vast majority of DTNs in this cluster have length of 132 aa. The corresponding CFs were relatively long (~1,150 aa) and did not have a suitable ORF to accommodate a spike protein adjacent to CF. Two prophages, one of *Acinetobacter baumannii SDF* and another one from *Achromobacter xylosoxidans A8,* have much longer DTNs (~160 aa) and relatively short CFs, 943 aa and 884 aa respectively. Both have a suitable ORF for a spike protein next to CF. The predicted interactions between 1 DTN and 3 SFs for these prophages are shown on **Fig.4**.

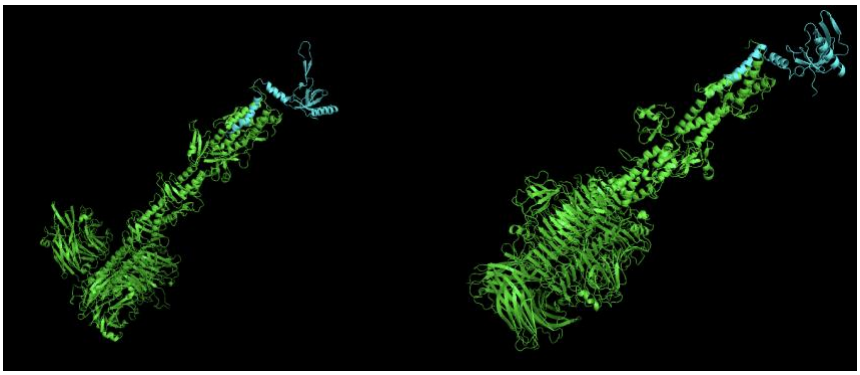

**Fig.4.** The predicted interactions between 1 DTN (cyan) and 3 SFs (green) for prophage SDF (left) and prophage A8 (right) prophages.

Cluster 3 contained 5 DTNs infecting a variety of Gram-negative hosts. The DTNs in this cluster have length of 125-145 aa. Two of five infecting agents in Cluster 3 with longer DTNs had relatively shorter CFs (~800-900 aa) and putative spike proteins next to them. The predicted interactions between 1 DTN and 3 SFs for these prophages, SC2 and Py2 are shown on **Fig.5**. The remaining three prophages have longer CFs of ~1200 aa and no apparent spike proteins.

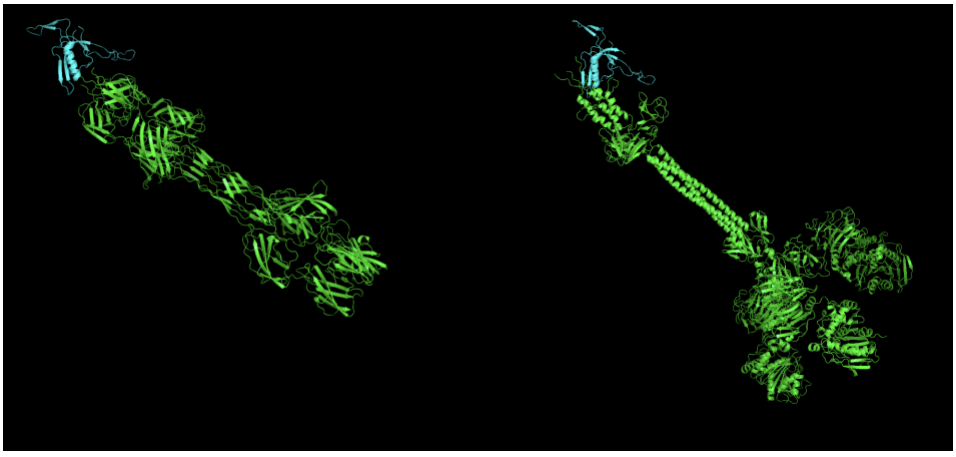

**Fig.5.** Predicted interactions between 1 DTN (cyan) and 3 SFs (green) for two prophages from Cluster 3, *Xanthobacter autotrophicus Py2* (left) and *Methylocystis sp. SC2* (right).

In Cluster 4, there were two prophages of *Methylorubrum extorquens CM4.* One of them had very short DTN protein of 85 aa, and the other prophage had a DTN protein of 124 aa. The sequence and structure alignments of these two DTN is shown on **Fig.6.** Both prophages have relatively large CFs and no large enough ORFs next to CF to accommodate separate spike proteins.

**A.**

**B.**

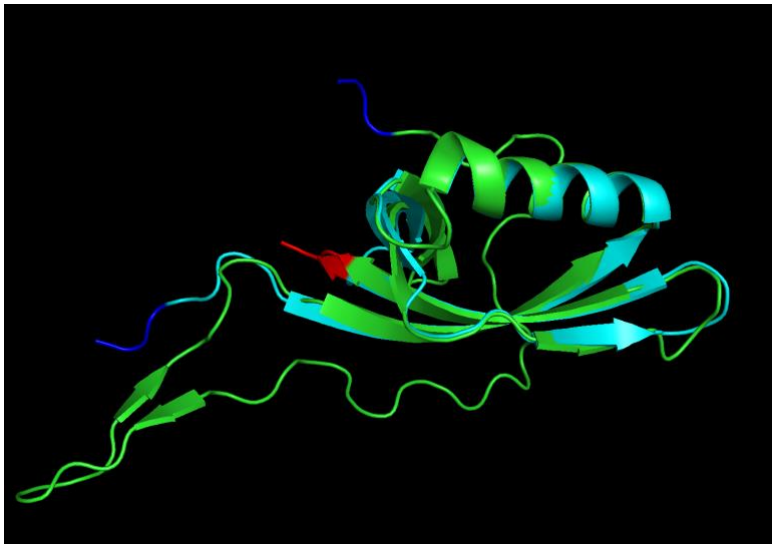

**Fig.6**. The alignments of DTN at the sequence (**A**) and structure (**B**) levels for two prophages of *Methylorubrum extorquens CM4* from Cluster 4. DTN 85 aa is in cyan and DTN 124 aa – in green; the C-terminal end of proteins is colored in blue and N-terminal end – in red.

Cluster 5 consisted of 5 DTN proteins (~130 aa) from prophages infecting diverse Gram-negative hosts. The CF proteins were about 1,200 aa and longer. Cluster 6 consisted of two DTNs of 154 aa in length from closely related prophages of *Acinetobacter baumannii*. These prophages have relatively short CFs of 944 aa each and spikes proteins of 648 aa adjacent to CF.

Cluster 7 was the largest cluster of all with 33 DTN proteins from phages and prophages infecting a range of Gram-negative hosts. The length of DTN was about 150 aa and longer. There was one prophage of *Rhodobacter capsulatus SB 1003* that had DTN of twice that length. This large DTN had an additional domain (**Fig.7A**) of about 170 aa in length in the middle of DTN extending outward from the DTN ring approximately between residue 70 and residue 240 (**Fig.7B**). Most of the infecting agents had Pseudomonas (12 of 33) or Klebsiella (9 of 33) as their host genus. The length of CF proteins varied from about 800 to 1,000 aa. All of them had spike proteins next to CF except for one Klebsiella prophage.

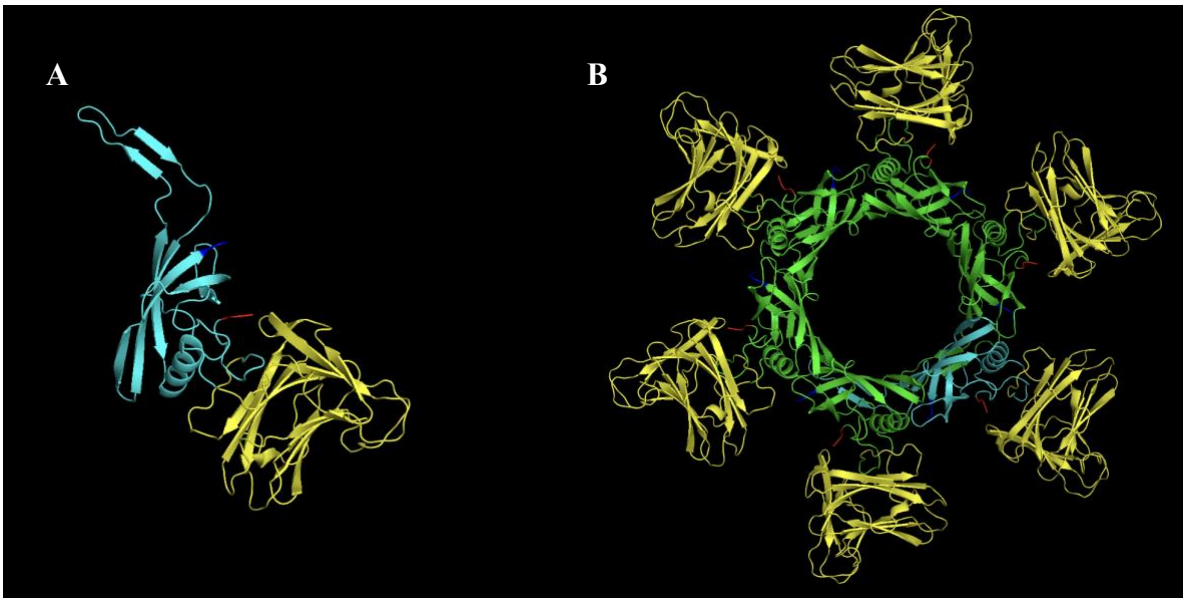


**Fig.7. (A)** The DTN of 299 aa with an extra domain in the middle of the protein (yellow). (**B**) The DTN hexamer with a DTN monomer in cyan; extra domains are in yellow. The C-term is blue, and the N-term is red.

In Cluster 8, there were 15 DTNs from a variety of infecting agents including 5 Xanthomonas phages and 3 Bordetella prophages. Also, one Bordetella prophage had DTN of 72 aa and no annotated CF, and one Burkholderia phage had an extremely long DTN of 372 aa representing two DTNs fused together. In this Burkholderia phage, a DTN ring was formed by a trimer with pseudo-sixfold symmetry because of the fusion (**Fig.8**). Only 2 of 15 had apparent spike proteins near CF.

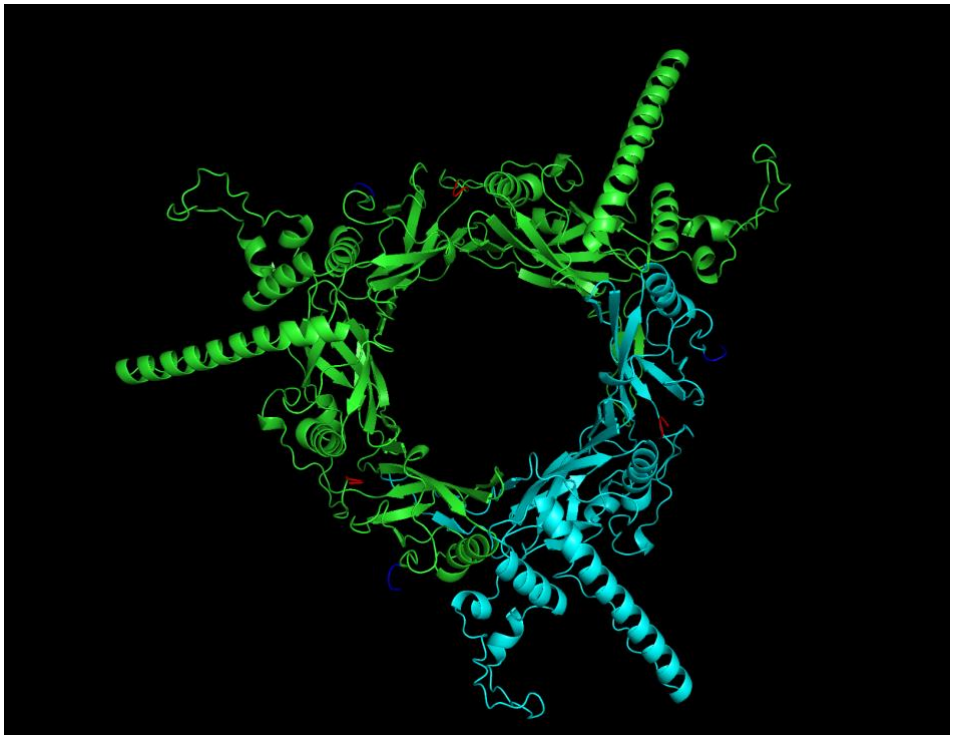

**Fig.8.** A trimer formed by the fused DTNs in Burkholderia phage (green). A DTN monomer is highlighted in cyan. The C-term is in blue, and the N-term is in red.

### 2.3.Variability in THN and the CF length and modularity

The THN and CF proteins in D3 also exhibit unique structural adaptations. In D3 phages, the CF is significantly enlarged, potentially compensating for a shorter or non-extended DTN. These variations suggest alternative mechanisms of host specificity, where a larger CF may bear the functional load typically carried by an RBP-attached DTN.

### 2.4. Diversity of D3-like Tail Tip Architectures

Our comparative genomic and structural analysis indicates that D3-type tail tips exist across a wide array of Gram-negative bacterial phages. We describe several variants of tail tip architecture in D3 related phages:

(1) D3-classic: elongated DTN with hinge and diverse RBP, moderate-sized CF
(2) Compact-DTN: short DTN, oversized CF potentially with built-in host binding function
(3) Hybrid types: intermediate DTN with multifunctional CF or THN extensions

These variations reflect a spectrum of strategies for environmental adaptation and host range expansion.

## 3. Materials and methods

### 3.1. The D3 phage group in PAT

In our in-house PAT database (Büttner et al., 2016), we identified a set of 18 phages and 65 prophages related to D3 with shared architectural features, 83 entries in total (**Fig.9**). The host genus distribution among these sequences is summarized in **Table 1.** These phages exhibit high conservation in their tail gene synteny, particularly across the distal tail (DTN), tail hub (THN), and central fiber (CF) regions, yet display striking diversity in their RBP modules. Most notably, a conserved hinge-like domain is consistently present within the DTN homologs across this group, suggesting a common structural strategy for modular RBP attachment. Phylogenetic clustering based on whole-genome alignments and core tail gene sequences places these phages into a distinct subclade, separated from canonical models like lambda and T5. The genomic variability localized to the RBP-coding region further supports a model of targeted recombination at the hinge site, facilitating host-specific adaptation. This collection not only reinforces the generality of the D3 architecture but also provides a rich resource for comparative and functional studies aimed at understanding the molecular basis of phage host range diversification.

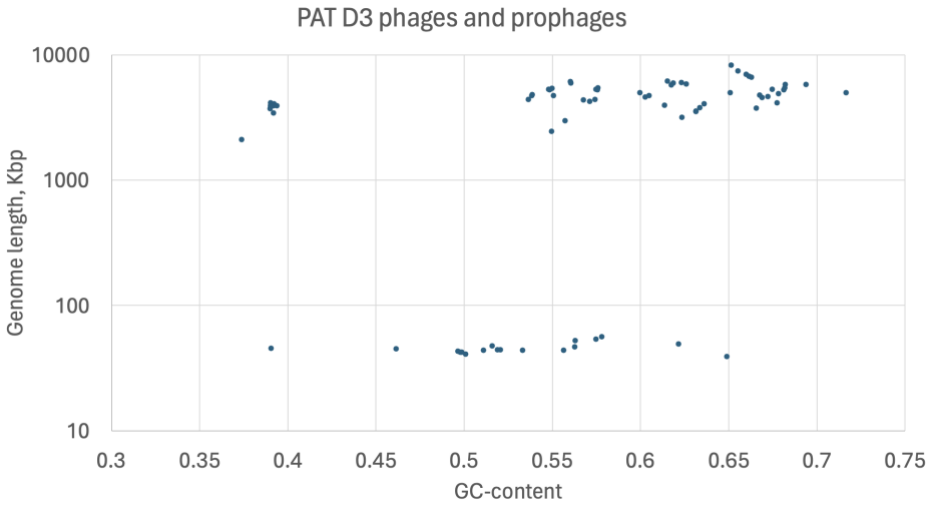


**Fig. 9.** Distribution of GC-content and sequence length (for a prophage a host genome sequence length is shown)

**Table 2.** Host genus distribution among 83 PAT phages and prophages related to D3.

| Host genus | #entries | Percent |
|---|---|---|
| Acinetobacter | 18 | 22% |
| Pseudomonas | 12 | 14% |
| Klebsiella | 9 | 11% |
| Enterobacter | 5 | 6% |
| Salmonella | 5 | 6% |
| Xanthomonas | 5 | 6% |
| Other | 29 | 35% |
| Total | 83 | |

### 3.2. Comparative analysis

To characterize the D3-like phages and elucidate their tail tip architecture, we employed a multi-tiered comparative genomics approach integrating both sequence- and structure-based analyses. Hidden Markov Models (HMMs) were used to detect distant homologs of tail tip proteins across diverse phage genomes, enabling the identification of conserved domains even in highly divergent sequences. Structural predictions from AlphaFold3 were compared using structure alignment tools like Dali to validate functional similarities among predicted DTN, THN, and CF proteins. Sequence similarity analysis was conducted to explore relationships between variants and assess clustering patterns. Additionally, we analyzed genome context and gene synteny, focusing on the positional conservation of tail tip genes to infer evolutionary relationships and potential recombination hotspots. This integrated framework allowed us to trace the modular organization of tail components and to distinguish conserved scaffolds from hypervariable RBP regions, revealing patterns of structural conservation paired with ecological diversification.

## 4. Discussion

### 4.1. Functional Implications

The structural modularity uncovered in the D3 tail tip provides compelling insights into the functional strategies that underlie phage host recognition. Most notably, the presence of a hinge-like domain within the enlarged distal tail needle (DTN) protein suggests a previously uncharacterized mode of structural flexibility that enables dynamic RBP integration. This hinge domain acts not only as a physical linker between the tail shaft and the receptor-binding protein (RBP), but also as a modular docking site, decoupling host specificity from the core architecture of the tail tip. Such an arrangement allows the phage to exchange or reconfigure RBPs without

perturbing essential interactions among the structural proteins responsible for DNA delivery. This configuration provides a strategic advantage: by relegating host specificity to a modular appendage rather than a deeply embedded domain, D3 minimizes the risk of functional disruption while maximizing adaptability.

From a mechanistic perspective, this modularity may streamline the process of host switching. Rather than requiring wholesale changes to the tail fiber complex or remodeling of core tail tip components – strategies that often come with significant structural trade-offs – D3 can potentially achieve host range expansion through a single recombination event at the hinge site. The hinge effectively serves as a pivot point, both literally and genetically, allowing insertion of diverse RBP modules tailored to specific Gram-negative bacterial receptors. This plug-and-play architecture sharply contrasts with more integrated systems like those of phages lambda and T5, where the receptor-binding domains are embedded within multifunctional proteins that participate in both structural stability and infection mechanics. In such systems, altering host specificity typically requires balancing conflicting structural constraints – a problem D3 avoids by isolating RBP function on a non-interacting, structurally permissive hinge module.

This configuration also provides a tangible molecular route for engineering phages with novel host specificities. In phage therapy and synthetic biology, host range specificity is a key design parameter. D3's hinge-based system offers an elegant genetic target: a recombination-prone site flanked by structurally self-contained domains, ideal for modular swapping. Unlike systems that require careful structural refolding or stabilization of large multi-domain proteins, the D3 hinge architecture allows for the simple replacement of an RBP module with minimal disruption to tail tip function. Switching host specificity in the D3 phage is akin to a cake decorator swapping the piping tip on an icing syringe: the core tool remains the same, but with a simple twist and replacement of the detachable tip, the decorator can craft entirely new patterns to suit each cake. Similarly, the D3 hinge domain acts like the coupler on a piping bag - designed to securely hold interchangeable components (RBPs) without altering the underlying mechanism. This modularity allows the phage to rapidly adapt to different bacterial hosts, not by rebuilding its tail apparatus from scratch, but by effortlessly changing the part that shapes its interaction with the host.

### 4.2. Evolutionary Perspectives

The D3 tail tip reveals an evolutionary trajectory that emphasizes flexibility over conservation, modularity over complexity. While siphophages such as lambda and T5 follow more rigidly defined architectural blueprints, D3 represents an alternative evolutionary solution in which tail tip diversity arises through recombination at discrete structural loci. The hinge domain in the DTN appears to function as an evolutionary nexus – a genetically mobile interface that supports rapid acquisition of new host-binding modules. This plasticity provides a selective advantage in fluctuating microbial environments, where host availability can shift rapidly and unpredictably.

The further analysis of D3-related phages suggest that such hinge-mediated modularity may be more widespread than previously appreciated. Tail tip gene neighborhoods in these phages display high local synteny despite divergent RBP content. This pattern supports a model in which the hinge domain not only tolerates but promotes genetic turnover, functioning as a scaffold for

evolutionary experimentation. In contrast to recombination events that disrupt core structural proteins or lead to partial tail tip dysfunction, hinge-mediated modularity preserves the mechanical integrity of the infection machinery while allowing fine-tuned adaptation to new hosts.

The implications extend beyond D3 itself: the presence of such modular systems suggests that phage tail tips are evolutionary mosaics, shaped by localized recombination rather than gradual modification. This redefines how we interpret the evolutionary landscape of siphophages - not as a spectrum from simple to complex, but as a network of modular solutions optimized for ecological fluidity. D3, in this context, is not an outlier but a representative of a broader evolutionary strategy: one that favors architectural flexibility as a mechanism for ecological resilience.

### 4.3. Other implications and future directions

This study provides a detailed structural and genomic dissection of the D3 siphophage tail tip, highlighting a previously unrecognized mode of modularity that challenges classical models of phage architecture. The identification of the hinge domain as a modular interface opens new avenues for understanding – and potentially engineering – phage host range. However, limitations remain. The functional dynamics of RBP exchange at the hinge are inferred but not directly demonstrated, and the infection process in live host systems has yet to be fully characterized. Additionally, while the hinge appears to promote modularity, the precise molecular mechanisms driving recombination and selection at this site warrant further investigation. Future work should focus on in vivo validation of hinge-mediated host switching, structural analysis of diverse hinge variants, and exploration of similar architectures in other phage families. These directions will be critical for harnessing the evolutionary potential of modular phages in both natural and applied contexts.

## 5. Conclusion

The structural and genomic analysis of the D3 phage tail tip reveals a profound shift in our understanding of siphophage diversity and adaptability. At the heart of this discovery lies a hinge domain – a structurally permissive, genetically accessible site that enables modular swapping of receptor-binding proteins without compromising the mechanical integrity of the tail apparatus. This architecture not only uncouples host specificity from core structural function, but also introduces a practical and evolvable mechanism for rapid adaptation. Unlike previously studied siphophages, where host recognition is intricately woven into multifunctional structural proteins, D3 presents a blueprint of architectural economy and evolutionary foresight.

The study's strength lies in its synthesis of comparative genomics and synteny analysis, producing a cohesive model that bridges molecular form with evolutionary function. While the findings provide a strong conceptual framework, future studies incorporating direct functional assays will be essential to fully validate hinge-mediated host switching. The implications of this

study are far-reaching: D3 challenges long-held assumptions about phage tail architecture and suggests a model of evolution driven by modular recombination rather than incremental change. Looking forward, this work opens exciting opportunities for rational phage engineering, particularly in the realm of synthetic biology and therapeutic phage design, where precision targeting of bacterial hosts is paramount. As more hinge-like systems are uncovered and experimentally characterized, D3 may stand not as a rare exception, but as a foundational example of evolutionary ingenuity.

## References


Büttner, C. R., Wu, Y., Maxwell, K. L., & Davidson, A. R. (2016). Baseplate assembly of phage Mu: Defining the conserved core components of contractile-tailed phages and related bacterial systems. *Proceedings of the National Academy of Sciences, 113*(36), 10174–10179.

Ge, X., & Wang, J. (2024). Structural mechanism of bacteriophage lambda tail's interaction with the bacterial receptor. *Nature Communications, 15*(1), 4185. 10.1038/s41467-024-48686-3

Huang, Y., Sun, H., Wei, S., Cai, L., Liu, L., Jiang, Y., Xin, J., Chen, Z., Que, Y., Kong, Z., Li, T., Yu, H., Zhang, J., Gu, Y., Zheng, Q., Li, S., Zhang, R., & Xia, N. (2023). Structure and proposed DNA delivery mechanism of a marine roseophage. *Nature Communications, 14*(1), 3609. 10.1038/s41467-023-39220-y

Linares, R., Arnaud, C., Effantin, G., Darnault, C., Epalle, N. H., Boeri Erba, E., Schoehn, G., & Breyton, C. (2023). Structural basis of bacteriophage T5 infection trigger and E. coli cell wall perforation. *Science Advances, 9*(12), eade9674. 10.1126/sciadv.ade9674

Wang, C., Duan, J., Gu, Z., Ge, X., Zeng, J., & Wang, J. (2024). Architecture of the bacteriophage lambda tail. *Structure, 32*(1), 35–46.e3. 10.1016/j.str.2023.10.006