Engineering modular bacteriophage genomes for targeted bacterial elimination

Introduction

Bacteriophages (phages) are viruses that infect and kill bacteria. They have shown great promise as therapeutic agents for treating bacterial infections, especially in the face of rising antibiotic resistance[1]. However, a major limitation of phage therapy is the narrow and unpredictable host range of most phages, which restricts their therapeutic applications and necessitates laborious screening processes to identify the appropriate phage for a given bacterial strain[2]. Phage engineering enables the customization of phages for enhanced bacterial targeting and killing, while minimizing the amount of experimental screening required.

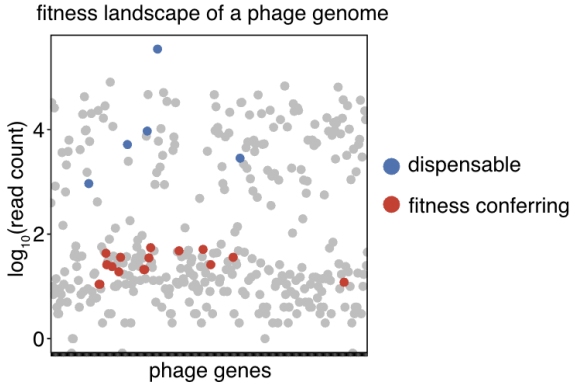
Customizing phage entry into specific bacterial hosts generally requires knowledge of what bacterial surface receptor to target, as well as the cognate receptor-binding phage protein. Whether a bacterial strain is susceptible or not to infection by a particular phage is usually dictated by its mechanism of entry and surface receptor compatibility. Receptors are often large protein complexes present at the surface of the bacterial cell such as the pilus or flagella. Receptors are conventionally identified by sequencing bacteria that have rapidly evolved resistance to a particular phage infection, but this does not necessarily result in a comprehensive list of receptors. Receptor-binding proteins (RBPs) on the phage can be identified similarly by sequencing phage that have evolved the ability to infect closely-related bacterial strains.

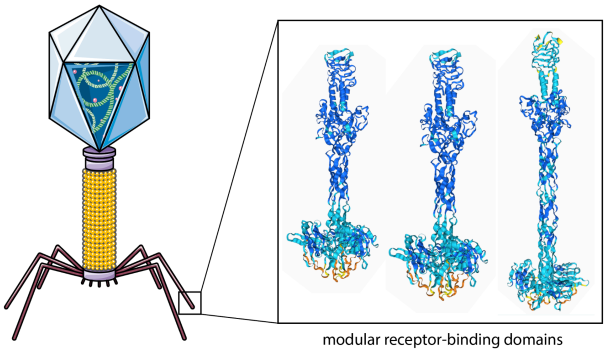
Previous phage engineering efforts, including deep-mutational scanning of phage receptor-binding domains[3] and concerted swaps of tail proteins between homologous phages[4], have attempted to establish non-natural phage-host pairings by exploring sequence space. While some of these methods have yielded desirable alterations to the phage host range, they cannot easily generalize to other phage or bacterial species beyond a certain threshold.

A major limitation in engineering phage genomes for targeting distant bacterial hosts is knowing whether a particular genomic edit will be tolerated or deleterious for its life cycle. The phage genome contains a subset of genes that are required for fitness in a majority of contexts (genome replication, virion production and assembly) and a subset of dispensable genes required in specific contexts (immune evasion factors and metabolism). Identifying specific genomic regions that can be replaced or removed is necessary for modular design of phages, yet this has remained challenging to achieve experimentally because of the genomic diversity of phages and limitation of current technology.

Results

To this end, we have developed a new high throughput genetic tool that identifies dispensable genomic regions in phages to facilitate their editing. We have applied this tool to identify dispensable genes in a non-model phage under active investigation for phage therapy. We use this to rapidly identify genomic regions amenable for transgene insertion, as well as specific phage genes that are essential for infection of a model pathogenic bacterium. This enables the design of a minimal phage scaffold that can be modified in a modular manner.





By pairing this minimal scaffold with the redesign of phage tail fibers to incorporate synthetic nanobody domains that can theoretically bind any given bacterial cell surface receptor, we further increase the generalizability of this platform. Preliminary computational results show that we can introduce nanobody domains to phage tail fibers to a model phage without disruption to the structural context.

Methods

To identify bacterial cell surface receptors, we generated mutated libraries of a bacterial isolate by transposon mutagenesis. Upon infecting a pool of bacterial mutants with phage, bacterial cells harboring mutations in receptors are shown to become enriched as they can resist infection and are subsequently identified by sequencing. This method can be applied to identify bacterial receptors from a wide array of phage-host pairs to experimentally generate a matrix detailing compatible phage receptor combinations.

To identify the putative phage proteins responsible for binding to the bacterial surface a priori, we generated embeddings of each phage gene using a protein language model that pairs ESM with contrastive learning[5], and queried for proteins with similarity to known phage RBPs.

To redesign the phage tail fibers while retaining stability in their native context, we used a combination of AlphaFold3 and ProteinMPNN, as outlined previously[6].

Discussion

In summary, we introduce a new experimental method that is augmented by computation to generate phage genomic scaffolds that can be used as starting points for phage engineering. Data generated by this method is particularly relevant to engineer phages that can both enter and replicate within specific host bacteria. Phage replication is thought to be a key component of the therapeutic benefit of bacteriophages in physiological contexts[7]. Additionally, creating multiple minimal phage scaffolds that are orthogonal to one another can enable the formation of phage cocktails that minimize the probability of resistance[8]. A key limitation of this genetic method is that if a given protein being queried is essential, then its change in fitness as a result of a perturbation (such as bacterial infection) cannot be quantified.

Experimental tools to increase the granularity of data elucidating phage-bacterial interactions are sorely needed, as most existing datasets simply capture phage or bacterial diversity independently. Data that is enabled by this tool may also serve as the foundation of more powerful models that can accurately predict the interactions between phage and bacteria that have never been experimentally verified.

Additionally, establishing modular genomic scaffolds for engineering phages for therapeutic purposes will be important for their experimental tractability. Minimizing the number of loci that need to be genetically modified enables more facile phage assembly, and limits variability between phages which is also amenable from a regulatory standpoint. Furthermore, the ability to engineer phages with specific binding specificities could also open up new possibilities for using phages such as diagnostic tools for detecting bacterial infections.

References

[1] Murray, C. J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655 (2022).
[2] Roucourt, B. & Lavigne, R. The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environmental Microbiology* **11**, 2789–2805 (2009).
[3] Huss, P., Meger, A., Leander, M., Nishikawa, K. & Raman, S. Mapping the functional landscape of the receptor binding domain of T7 bacteriophage by deep mutational scanning. *Elife* **10**, e63775 (2021).
[4] Yehl, K. *et al.* Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell* **179**, 459-469.e9 (2019).
[5] Hong, L., Hu, Z., Sun, S. *et al.* Fast, sensitive detection of protein homologs using deep dense retrieval. *Nat Biotechnol* (2024). https://doi.org/10.1038/s41587-024-02353-6
[6] Wang, J., Watson, J. L. & Lisanza, S. L. Protein Design Using Structure-Prediction Networks: AlphaFold and RoseTTAFold as Protein Structure Foundation Models. *Cold Spring Harb Perspect Biol* **16**, a041472 (2024).
[7] Bosco, K., Lynch, S., Sandaradura, I. & Khatami, A. Therapeutic Phage Monitoring: A Review. *Clinical Infectious Diseases* **77**, S384–S394 (2023).
[8] Kim, M. K. *et al.* A blueprint for broadly effective bacteriophage-antibiotic cocktails against bacterial infections. *Nat Commun* **15**, 9987 (2024).