RESEARCH ARTICLE SUMMARY

SYNTHETIC BIOLOGY

Generative design of bacteriophages with genome language models

Samuel H. King, Claudia L. Driscoll, David B. Li, Daniel Guo, Aditi T. Merchant, Garyk Brixi, Max E. Wilkinson, Brian L. Hie*

Full article and list of author affiliations: https://doi.org/10.1126/science.aec2657

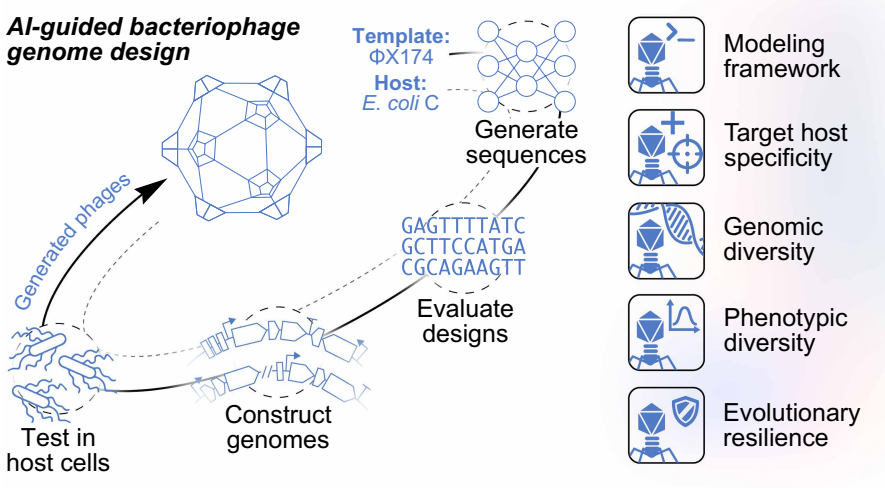
INTRODUCTION: Evolution continuously forges new biological innovations written in genomes. Navigating this vast design space could access functions that would transform biotechnology, but even the simplest genomes are highly complex and can be rendered nonviable by a single mutation. Accordingly, most progress in biological design has been made at the scale of individual genes and gene circuits, whereas design at the scale of whole genomes has remained largely beyond reach.

RATIONALE: Genome language models are artificial intelligence (AI) algorithms that have shown promise in designing biological systems. Much like how other language models are trained on large corpora of text, genome language models are trained on large corpora of DNA comprising millions of genomes from all domains of life. This enables these models to learn the evolutionary constraints that shape DNA sequences in nature. However, the ability of genome language models to generate entire functional genomes has not been tested. Bacteriophages, viruses that infect bacteria, are specifically well suited for this task, as they are relatively small, experimentally tractable, and have broad applications in molecular biology, microbial engineering, and therapeutics.

RESULTS: In this work, we leveraged genome language models, Evo 1 and Evo 2, to generate complete phage genomes with realistic genetic architectures and specificity for a bacterial host, *Escherichia coli* C. Using the natural phage ΦX174 as a design template, we established a framework for generating and evaluating thousands of AI-generated genomes, nearly 300 of which we chemically synthesized and tested in laboratory conditions, yielding 16 viable phages. The viable generated phages showed strong host specificity and diverse fitness profiles, including competitive infection kinetics. The generated phages were different from any known natural phages, exhibiting de novo mutations, divergent genes and regulatory elements, and variable genome lengths. One of the phages utilized a DNA packaging protein from an evolutionarily distant phage in its capsid structure. We also tested whether the generated phages could overcome bacterial resistance, a central challenge in developing phage-based antimicrobial therapies, and found that a mixture of designed phages rapidly overcame ΦX174-resistant *E. coli* strains, whereas a comparable mixture of naturally sourced ΦX174-like phages could not.

CONCLUSION: Our results demonstrate that generative models capture evolutionary constraints in DNA sequences with enough fidelity to produce complete bacteriophage genomes divergent from those observed in nature and with prespecified traits. Our approach expands what synthetic genomics can achieve alongside methods such as directed evolution and rational engineering, lays out a path for generating adaptive and resilient phage therapies against rapidly evolving pathogens, and establishes a foundation for the generative design of larger, more complex genomes. Genome design can augment the broader toolkit of genome sequencing, synthesis, and editing, enabling the composition of biological systems at the genome scale. □

*Corresponding author. Email: brianhie@stanford.edu Cite this article as S. H. King *et al.*, *Science* **393**, eaec2657 (2026). DOI: 10.1126/science.aec2657



A framework for AI-guided bacteriophage genome design. Language models trained on millions of genomes enable the generation of complete phages. Sequences are generated and computationally evaluated with design criteria inspired from the phage ΦX174 and its host *E. coli* C. The most promising designs are chemically synthesized and tested in host cells, yielding viable generated phage genomes divergent from known natural phages, while retaining host specificity.

SYNTHETIC BIOLOGY

Generative design of bacteriophages with genome language models

Samuel H. King[1,2], Claudia L. Driscoll[2,3], David B. Li[1,2], Daniel Guo[2,4], Aditi T. Merchant[1,2], Garyk Brixi[2,5], Max E. Wilkinson[6]†, Brian L. Hie[2,3,7]*

Many important biological functions arise not from single genes but from complex interactions encoded by entire genomes. We report the first generative design of complete bacteriophage genomes using genome language models. We generated viable bacteriophages with target host tropism, using the phage ΦX174 as our design template. Experimental testing yielded 16 phages with diverse fitness profiles in laboratory conditions. Cryo–electron microscopy confirmed that a generated phage utilizes an evolutionarily distant DNA packaging protein in its capsid. A cocktail of generated phages rapidly overcomes ΦX174-resistant *Escherichia coli* strains, demonstrating a path toward artificial intelligence–generated phage therapies against rapidly evolving bacterial pathogens. This work provides a blueprint for the design of diverse synthetic bacteriophages and useful biological systems at the genome scale.

Advances in DNA sequencing and synthesis have vastly improved our ability to read and write DNA at the scale of whole genomes (*1*–*5*). Although these technologies could also facilitate whole-genome design, potentially enabling new biotechnologies or therapeutic modalities (*6*–*9*), intelligently composing new genomes still represents a monumental challenge. Genomes contain emergent complexity, wherein interactions involving multiple genes, regulatory elements, recognition sequences, and other features are tightly orchestrated to enable replication and other higher-order functions (Fig. 1A) (*10*–*12*). Genomes are also highly sensitive to their sequence composition, as even a single mutation can render an entire genome nonviable (*13*–*16*). Despite efforts to design new proteins, reprogram genetic codes, and engineer increasingly complex biological circuits (*17*–*23*), there remains no generalizable framework for designing entire genomes.

Recent advances in artificial intelligence (AI) leverage data and compute at scale to enable complex generative tasks (*24*, *25*). In biology, language models trained on vast genomic sequence datasets learn complex rules that enable the generative design of new DNA sequences with desired functions (*26*–*28*). For example, we previously reported the successful design of functional systems, such as CRISPR-Cas complexes, transposable elements, and toxin-antitoxin interactions, with genomic language models Evo 1 and Evo 2 (*27*, *28*). However, these systems contain only a small number of genetic components, whereas the design of complete genomes would require new methodological advances and much deeper integration of machine learning, computational biology, and experimental biology.

In this work, we report the first generative design of complete genomes. We focused on the design of bacteriophages, which have substantial utility as biotechnologies and have therapeutic relevance as treatments for bacterial infections (*6*, *29*–*33*). As a design template, we selected ΦX174, a lytic *Microviridae* phage with a ~5.4–kilobase (kb) genome containing 11 genes, at least seven regulatory elements, and two recognition sequences (*34*–*37*). Although ΦX174 has a much more complex genetic architecture than any previously AI-generated biological system, its relatively small genomic length and rich history of experimental work (*38*) also make it a tractable and safe model for establishing whole-genome design. Notably, ΦX174 was the first complete DNA genome sequenced and synthesized (*36*, *39*) and has continually served as a pivotal model within molecular biology (*13*, *40*, *41*).

Enabling generative design of complete bacteriophages first required substantial computational development of both generative models and bioinformatic prediction tools. Our full generative pipeline consisted of unsupervised pretraining with Evo 1 and Evo 2, task-specific fine-tuning on *Microviridae* genomes, prompt engineering with ΦX174-specific sequences, and inference-time guidance by means of predictive models of genomic architecture and host tropism. Whole-genome bacteriophage design also required extensive development of experimental methods, including a scalable protocol that enabled us to screen nearly 300 genome designs to yield 16 functional phages, with tropism successfully limited to the target host.

The generated phages have diverse sequences and structures. Some of these phages also have competitive growth rates and lysis kinetics compared with those of natural *Microviridae* phages. In a setting analogous to pathogen resistance to phage therapies, a cocktail of the generated phages rapidly overcame resistance in two different ΦX174-resistant *Escherichia coli* strains, whereas a natural ΦX174-like phage cocktail could not overcome resistance. These results establish generative genomics as a powerful strategy for accessing evolutionary sequence space and for potentially creating effective and evolutionarily resilient phage therapeutics.

More broadly, this study establishes a generalizable approach to generative bacteriophage genome design under steerable, user-specified constraints. We envision that techniques developed in this work offer a path toward designing more complex biological systems with desirable functions, potentially including larger bacteriophage genomes and prokaryotic genetic systems at the scale beyond tens of genes.

Evo generates realistic bacteriophage genomic sequences

Bacteriophage genome design is distinctively complex because a complete phage genome must reconcile features such as coding and noncoding elements, architectural arrangement, gene directionality, structural motifs, and host interactions to yield a temporally coordinated multistage life cycle within and outside of host cells (Fig. 1A). By contrast, genetic systems of similar length, such as prokaryotic operons, generally execute modular functions that typically do not require the same degree of global coordination. We hypothesized that, with advances in genome language models and with the appropriate training data, it would be possible to learn underlying evolutionary constraints that could allow for the generation of phage genome sequences not yet seen in nature (Fig. 1B).

We investigated whether Evo 1 and Evo 2, which were pretrained on large corpora of DNA sequences, including >2 million bacteriophage genomes (*26*, *28*), had baseline capability for generating phage-like sequences (Fig. 1C). We leveraged special taxonomic sequence labels that were included alongside genomic sequences during pretraining (*26*, *28*) to specifically prompt the models to generate phage-like sequences using three prompts corresponding to major viral realms: *Duplodnaviria* (double-stranded DNA viruses), *Monodnaviria* (single-stranded DNA viruses), and *Riboviria* (RNA viruses) (Fig. 1C; materials and methods).

We first evaluated whether the generated sequences resembled natural phage sequences using geNomad, a widely adopted viral classification tool (*42*). As a baseline, geNomad reliably classified 89 to 100% of

[1]Department of Bioengineering, Stanford University, Stanford, CA, USA. [2]Arc Institute, Palo Alto, CA, USA. [3]Department of Chemical Engineering, Stanford University, Stanford, CA, USA. [4]Department of Computer Science, Stanford University, Stanford, CA, USA. [5]Department of Genetics, Stanford University, Stanford, CA, USA. [6]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Stanford Data Science, Stanford University, Stanford, CA, USA.
***Corresponding author. Email: brianhie@stanford.edu** †Present address: Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

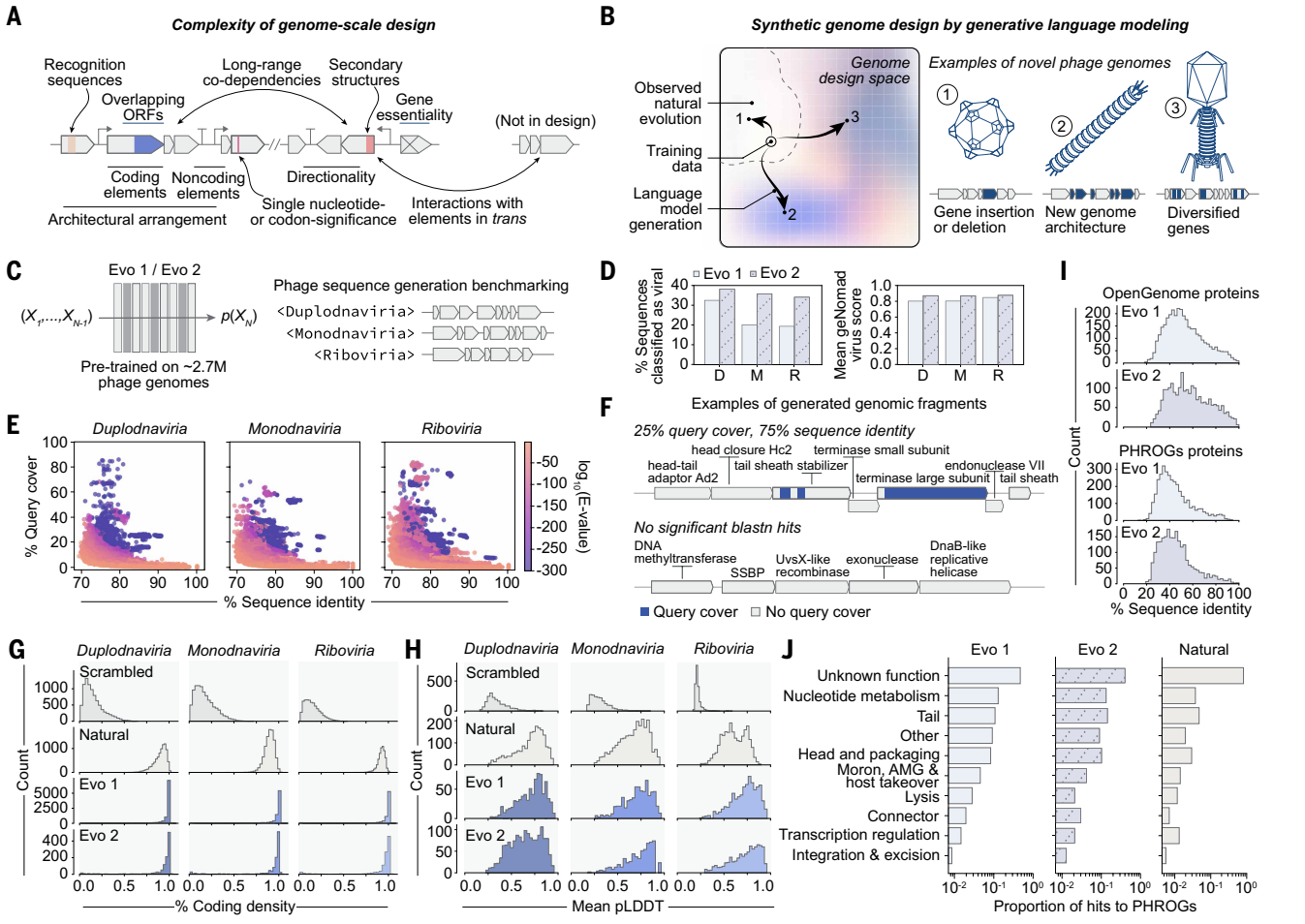


Fig. 1. Evo generates realistic bacteriophage genomic sequences. (**A**) Key considerations highlighting the complexity of genome design. (**B**) Generative genome language models have the potential to explore phage genome design space when trained on a subset of observed natural evolution. (**C**) We benchmarked Evo 1 and Evo 2 on a broad zero-shot prompting task for phage genomic sequence design. (**D**) Generated sequences classified as viral by geNomad (left) with high average virus classification scores across prompts (right). D, *Duplodnaviria*; M, *Monodnaviria*; R, *Riboviria*. (**E**) Generated sequences show low query cover and sequence identity against natural sequences in nucleotide BLAST searches. (**F**) Generated sequences contain predicted phage-like architectures, with regions that match natural sequences (blue) and regions without nucleotide BLAST hits (gray). (**G**) Predicted coding densities of generated sequences are reminiscent of but higher than natural sequences and unlike scrambled natural sequences. (**H**) ESMFold-predicted protein structures from generated sequences have mean predicted local distance difference test (pLDDT) scores similar to natural proteins, higher than predicted proteins from scrambled natural sequences. (**I**) Generated proteins align to proteins in the OpenGenome and PHROGs databases with generally low sequence identities, indicating sequence divergence. (**J**) Functional annotations of generated proteins closely match those of natural phages when queried against the PHROGs database.

natural phage sequences as viral depending on the realm (geNomad mean score > 0.96), whereas the same natural phage sequences but randomly scrambled were rarely classified as viral (1.9 to 11%, mean score > 0.73) (fig. S1A). By contrast, 19 to 33% and 34 to 38% of Evo 1 and Evo 2 generations, respectively, were classified as viral (Evo 1 mean score > 0.80, Evo 2 mean score > 0.87) depending on the prompt (Fig. 1D). These results support that both pretrained models can generate phage-like sequences, with Evo 2 showing stronger baseline performance.

Querying generated sequences against the nucleotide BLAST database (*43*) revealed that generated sequences are distinct from those in nature (Fig. 1, E and F), yet they retained a high predicted coding density and genetic features reminiscent of natural phage genomes (Fig. 1, F and G). Structure prediction (*44*) of the proteins generated by both Evo 1 and Evo 2 yielded predicted confidence scores consistent with those expected for structure prediction of natural proteins (Fig. 1H and fig. S1B). When compared with proteins in the Evo models' pretraining data, OpenGenome (*26*, *28*), and the PHROGs database, a comprehensive database of phage proteins (*45*), generated proteins generally showed low sequence identity with a realistic distribution of gene annotations (Fig. 1, I and J). Together, these data demonstrate that pretrained Evo models can design biologically realistic phage genomic sequences while introducing sequence diversity beyond observed natural evolution.

Generative design of bacteriophages with target host tropism

We next hypothesized that generative genomic design, with appropriate templates and constraints, could propose complete and viable phage genomes (Fig. 2). To achieve bacteriophage design, we devised a workflow (Fig. 2A) that consisted of (i) selecting a target host of biological or therapeutic interest; (ii) choosing a design template, a phage genome known to infect that host, and collecting evolutionarily related sequence data; (iii) training or fine-tuning a genome language model on that sequence data; (iv) establishing design constraints based on the design template; (v) computationally evaluating and filtering generated sequences using those constraints; and (vi) experimentally validating the designed genomes.

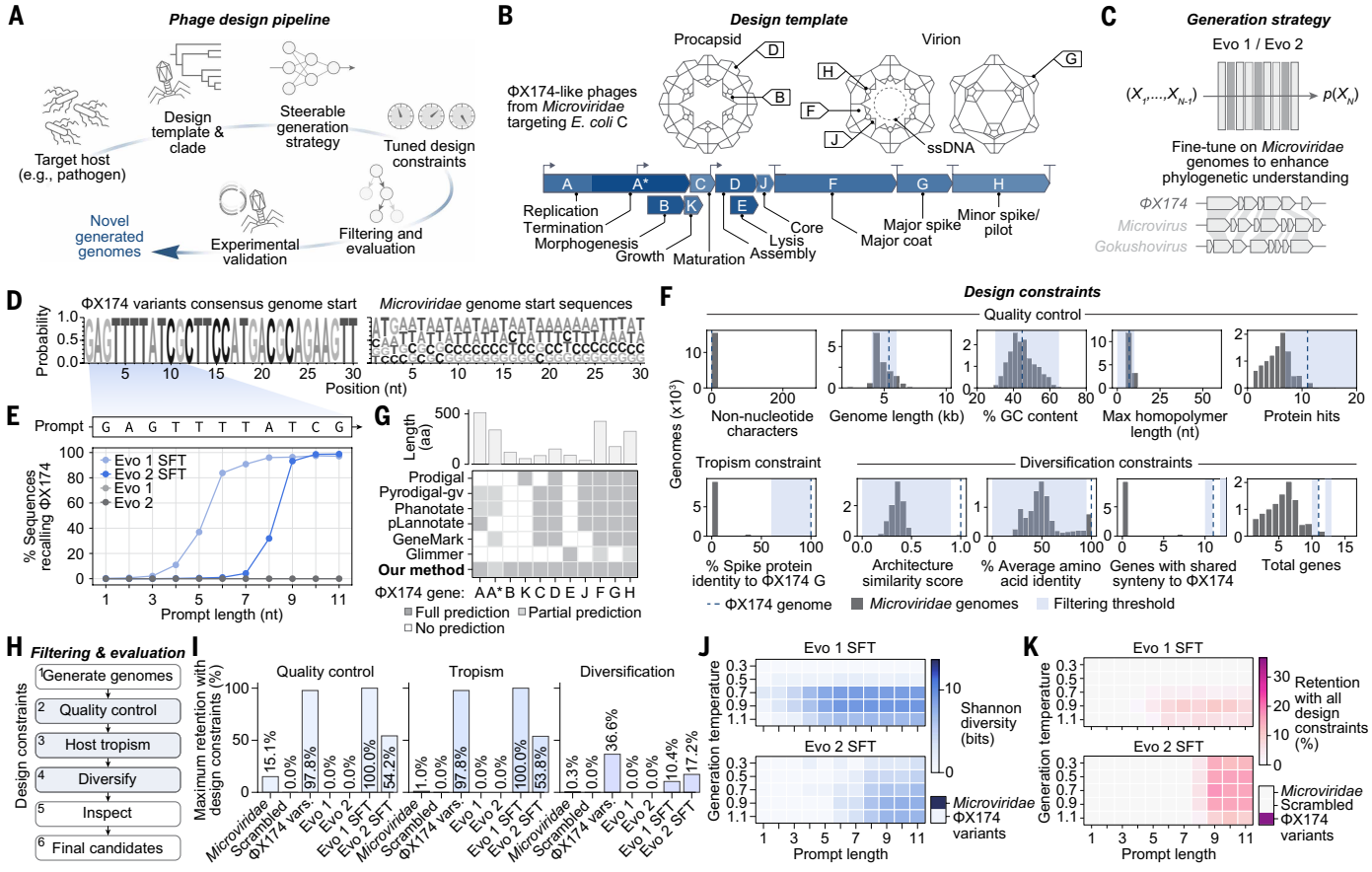


Fig. 2. Generative design of bacteriophages with target host tropism. (**A**) Phage genome design workflow. (**B**) Genetic architecture of our design template, ΦX174, a *Microviridae* phage that uses *E. coli* C as a host. (**C**) For our generation strategy, we specialized Evo 1 and Evo 2 on *Microviridae* genomes through SFT to enhance its ability to generate ΦX174-like sequences. (**D**) Sequence logos showing conserved nucleotides at the start of ΦX174 variant genomes compared with *Microviridae* genomes in our training data. (**E**) Increasing the number of ΦX174 nucleotides in the prompt quickly improved recall with the SFT models, whereas the base models failed to recall ΦX174 across all prompt lengths. (**F**) Design constraints selected for genome filtering, with thresholds (blue) chosen against natural *Microviridae* distributions (gray) and ΦX174 (dotted line). (**G**) Benchmarking of six gene prediction methods on the ΦX174 genome shows that overlapping genes are systematically missed, requiring us to create a method that predicts all genes in ΦX174. (**H**) Final filtering and evaluation steps for generated genomes. (**I**) Maximum sequence retention rate across Evo 1 and 2 base model– and SFT-generated sequences after applying design constraints, with natural *Microviridae*, scrambled *Microviridae*, and ΦX174 variant controls. The selected design constraints retained ΦX174-like sequences across quality control, tropism, and diversification filters. (**J** and **K**) Generated sequences had high Shannon diversity after tropism filtering (J) and maintained a high retention rate even after further diversification filtering (K). Diversity and retention rate both increased with generation temperature and prompt length. *Microviridae*, scrambled *Microviridae*, and ΦX174 variants are shown for comparison. $N = 1000$ sequences per parameter combination.

Following these steps, we selected *E. coli* C and its native phage ΦX174 as our design template (Fig. 2B). ΦX174 is a well-studied phage that belongs to *Microviridae*, a family of phages with small genomic sizes (typically ~4 to 6 kb) and an abundance of available sequencing data (fig. S2A) (*38*, *46*); *E. coli* C is a nonpathogenic, well-characterized host strain (*47*). These features make ΦX174 and *E. coli* C an ideal testbed for generative phage design.

Given the success of previous design tasks that required supervised fine-tuning (SFT) to generate coherent CRISPR-Cas systems and transposable elements (*28*), we fine-tuned Evo 1 7B 131K and Evo 2 7B 8K on a dataset of ~15,000 *Microviridae* sequences (Fig. 2C and figs. S2B and S3, A to F), resulting in higher fidelity language modeling of ΦX174-like genomes (fig. S3, G and H). We then sampled sequences from the *Microviridae* SFT models, leveraging the property that they were trained to initiate full-length genome generation from the first position of their input context. Whereas the genomic start sequences of all *Microviridae* sequences in the training data were highly diverse, all the ΦX174-like genomes started with the same consensus nucleotides (Fig. 2D). By prompting with a portion of this consensus sequence, we were able to generate ΦX174-like sequences with both Evo 1 and Evo 2 SFT models (Fig. 2E). By contrast, the base models failed to recall ΦX174 across all tested lengths of the consensus sequence prompt.

To assess the quality of generated DNA sequences, we developed a set of genome-level design constraints by computing various statistics on natural *Microviridae* sequences, including ΦX174 (Fig. 2F). We organized these constraints into three tiers: sequence quality, tropism specificity, and evolutionary diversity. As basic sequence quality control, we filtered out generated sequences containing nonnucleotide characters, enforced lengths between 4 to 6 kb and GC content within 30 to 65%, and excluded sequences with DNA homopolymers longer than 10 bases.

We also sought to add constraints on the protein-coding genes of our generated genomes. However, we found that none of six broadly used gene annotation tools were able to annotate all 11 genes on the wild-type ΦX174 sequence (Fig. 2G and fig. S4A), reflecting the challenge of predicting overlapping open reading frames (ORFs) (fig. S4, A and B) (*48*). To overcome this limitation, we built a bespoke coding

sequence (CDS) prediction method tailored to ΦX174-like sequences (fig. S4, C and D; materials and methods) that was able to fully annotate all genes in ΦX174, with the exception of gene A*, which was partially predicted (Fig. 2G). With our method, we applied an additional quality control constraint requiring at least seven predicted protein hits to natural phage proteins. Because host range is largely determined by the ability of viral spike proteins to bind host cell receptors (*47*, *49*), we applied a tropism constraint requiring that generated genomes encode spike proteins with moderate sequence identity (≥60%) to the ΦX174 spike protein (Fig. 2F).

We encouraged evolutionary novelty by individually applying any of three diversification filters (Fig. 2F). We preferred genomes with ≤95% average amino acid identity (AAI) to natural proteins, directly promoting diverged proteome sequences. We also developed a "genetic architecture" constraint to capture preservation of global gene arrangement relative to ΦX174, which scores genomes by fuzzy boundary matching of start and stop codons positions, allowing us to remove sequences that too closely resembled the exact coding position architecture of ΦX174 (Fig. 2F and fig. S5; materials and methods). We favored genomes with 10 or 12 genes in total or those sharing synteny with 10 or 12 of ΦX174's genes, allowing for variants with single gene losses or gains.

With our full set of design constraints, we narrowed the set of generated sequences through successive rounds of quality control, tropism filtering, and diversification filtering (Fig. 2H). These filters also separated natural ΦX174-like sequences from other *Microviridae* genomes and scrambled *Microviridae* sequences (Fig. 2I). After tropism filtering, we retained as much as 100% of the Evo 1 SFT generations and 53.8% of the Evo 2 SFT generations. After diversification-based filtering, we retained as much as 10.4% of Evo 1 generations and 17.2% of Evo 2 generations.

Lastly, building on our prior finding that sequence context strongly influences generative outputs (*27*), we systematically assessed how prompt length alongside autoregressive sampling temperature influenced the diversity and quality of generated genomes. We found that prompting with the first nine or more nucleotides from the ΦX174 consensus sequence (Fig. 2D) led to simple, memorized recall of ΦX174 with minimal diversity at low sampling temperatures (fig. S6). By contrast, prompting with only the first one or two nucleotides from the consensus sequence did not provide sufficiently strong conditioning to produce ΦX174-like generations. Notably, steering toward diverse, ΦX174-like sequences required carefully tuning the length of the consensus-sequence prompt to be around four to nine nucleotides and the sampling temperature to be in the 0.7 to 0.9 range (Fig. 2, J and K, and fig. S6). The ability to use such short prompts with the SFT models ensured that the majority of each generated sequence could be generated by the models.

Both Evo 1 and Evo 2 SFT models produced highly diverse sequence populations even after tropism filtering, with Shannon diversity, a measure of biodiversity (*50*), increasing alongside temperature and prompt length (Fig. 2J). This diversity did not come at the cost of design quality, with a high percentage of sequences being retained after applying our diversification filters (Fig. 2K). In total, these analyses show that by combining steerable generation with systematic design constraints, Evo models can propose phage genomes that are realistic, template-guided, and distinct from observed natural phages.

Creating viable generated bacteriophage genomes

Beyond computational design, we then sought to experimentally test whether the generated genomes encode viable phages (Fig. 3). Using our evaluation criteria, we curated a set of 302 diverse generated phage genome candidates (figs. S7 and S8). We designated these phages "Evo-Φ" followed by a numeric sequence identifier. These genomes spanned 4 to 6 kb in length, had AAIs to natural proteins as low as 63%, and retained spike protein sequence identities largely above 85% (Fig. 3A). Most candidates shared >40% nucleotide identity with ΦX174 and genomes in the *Microviridae* training data. CheckV, a tool for assessing the quality of viral sequences, classified >87% of the generated sequences as High Quality or Complete (*51*). Many of the generated genes also shared no coding sequence similarity with any gene in ΦX174, resulting in breaks in synteny (Fig. 3B). Most genomes encoded 11 genes in total, with 10 genes preserving synteny with ΦX174 (Fig. 3C).

We developed a growth assay for testing phage replication and lysis similar to previously established phage rebooting protocols (Fig. 3D) (*52*, *53*). A mature ΦX174 virion contains a circular single-stranded genome that becomes double-stranded during replication (*54*, *55*), a stage that enables phage rebooting through transformation of a circular double-stranded DNA (dsDNA) product (*52*, *53*) that can be obtained through in vitro assembly of dsDNA fragments (Fig. 3D). With this protocol, we observed that genomes of ΦX174 robustly formed plaques on a plate of *E. coli* C, whereas variants with mutated lysis genes did not (Fig. 3E and fig. S9A). In our growth inhibition assay, ΦX174 impeded growth of *E. coli* C in less than 2 hours (Fig. 3F). Upon successful growth inhibition, we created a stock of the phage-infected culture, verified its identity by long-read sequencing, then propagated and titrated the phage for downstream assays (fig. S9, B and C; materials and methods).

Upon validating our screening protocol with wild-type ΦX174, we successfully synthesized and assembled 285 out of 302 generated genomes; the remaining failed because of high-complexity DNA synthesis. Measuring bacterial growth inhibition as an indication of phage viability, we observed 16 generated phage transformations that inhibited growth of *E. coli* C (Fig. 3, G and H, and fig. S10A). When transformed in *E. coli* K-12, none of the 285 assemblies resulted in growth inhibition, supporting the robustness of the tropism constraint filter (fig. S10B). Sequence-verification of the 16 candidates revealed nine genomes with no acquired mutations, whereas the other seven acquired some single-nucleotide variants (SNVs) or deletions during synthesis or rebooting (fig. S11). Four out of the seven sets of acquired mutations made the phages more similar to their nearest natural genome, and some were essential for viability, suggesting selective pressure for some sequences observed in nature (fig. S12, A and B). Overall, 12 out of 227 (5.3%) tested sequences generated by Evo 1 and 4 out of 58 (6.9%) generated by Evo 2 were viable. Viability tended to increase with sequence similarity to natural genomes, with as high as 46.2% of generated phages viable in the 98 to 100% sequence similarity range (~100 mutations away) from their nearest natural genomes (fig. S12C).

Lastly, we tested the host range of the generated phages across eight *E. coli* strains, including *E. coli* C, *E. coli* B, *E. coli* W, and five variants of *E. coli* K-12 (Fig. 3I and figs. S13 and S14). ΦX174 and 15 out of 16 of the generated phages could also inhibit growth in *E. coli* W, a host not previously associated with ΦX174 (*47*, *56*), although they did so with much more variation in growth kinetics (fig. S14). We did not observe growth inhibition in the other six *E. coli* strains, suggesting that the generated phages maintain a high specificity for the intended host, *E. coli* C. More broadly, these results demonstrate that genome language models, combined with inference-time steering and filtering, can design viable phage genomes.

Generated bacteriophages reveal sequence and structural insights

Having validated the viability of our generated phages, we next examined the extent of their evolutionary divergence from natural phages (Fig. 4). Upon analyzing the mutational differences within the generated genomes relative to their nearest natural genomes and ΦX174, we observed hundreds of synonymous, nonsynonymous, and noncoding mutations (Fig. 4A and fig. S15, A to E) as well as diverse genetic innovations, including (i) a putative insertion of a new gene J in Evo-Φ63; (ii) extended noncoding regions in Evo-Φ63, Evo-Φ2147, Evo-Φ2483, and

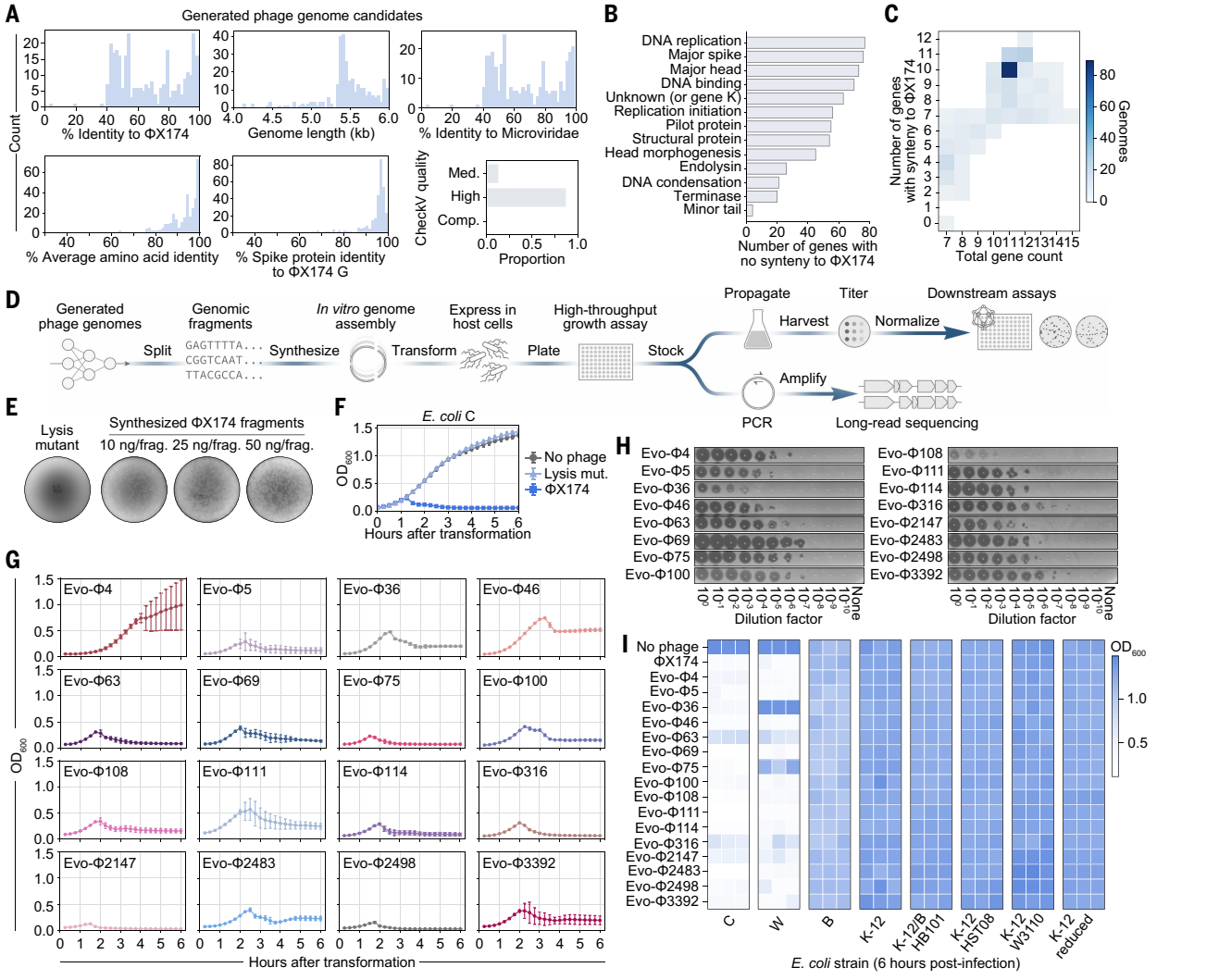


Fig. 3. Creating viable generated bacteriophage genomes. (**A**) Final generated phage candidates met our quality criteria while capturing abundant sequence diversity. kb, kilobase; Med., Medium; Comp. Complete. (**B**) Many generated sequences encode genes with low sequence identity to ΦX174, resulting in breaks in gene synteny. (**C**) A heatmap of total gene count versus number of syntenic genes shows that most sequences contain a single-gene break in synteny to ΦX174, balancing conservation with diversity. (**D**) Workflow for experimental validation of generated phage genomes. (**E**) *E. coli* C transformed with synthesized ΦX174 genome assemblies show that phage plaques formed robustly across assembly conditions. By contrast, the same genomes but with loss-of-function mutations in lysis genes did not form plaques. ng, nanogram; frag., fragment. (**F**) Growth curves of *E. coli* C transformed with no phage (gray), ΦX174 lysis mutant (mut., light blue), or wild-type ΦX174 (dark blue) genome assembly reveal strong growth inhibition by ΦX174. Data point, mean OD_{600} (optical density at 600 nm) value; error bar, standard deviation; $N = 3$ growth replicates. (**G**) Growth curves of generated genome assemblies transformed in *E. coli* C exhibiting growth inhibition. (**H**) Representative titrations of propagated phage candidates. (**I**) Growth inhibition measured by OD_{600} at 6 hours after infection of *E. coli* cultures with no phage, ΦX174, or generated phages shows that ΦX174 and generated phages inhibit growth in the target strain *E. coli* C and in *E. coli* W but not in six other strains tested, demonstrating the robustness of tropism filtering. Each column represents an infection replicate.

Evo-Φ2498; (iii) loss of gene K in Evo-Φ114; (iv) putative truncations of gene C in Evo-Φ75 and Evo-Φ100 and of gene B in Evo-Φ316; (v) putative elongations of gene E in Evo-Φ4 and Evo-Φ46; and (vi) swapping of gene J in Evo-Φ36 with a truncated gene J from the distant Escherichia phage G4, a swap previously found nonviable for wild-type ΦX174 (*57*–*59*).

The most mutated elements amongst the 16 genomes were promoter A, gene J, and terminator J, with average mutation rates of 0.054, 0.052, and 0.043, respectively (Fig. 4B). The origin of replication had no mutations across all 16 genomes.

At a phylogenetic level, the generated phages were similar to members of the ΦX174-like *Microviridae* clade, supporting our ΦX174-templated design approach (Fig. 4C and fig. S15F). The likelihoods of generated phages under a multiple sequence alignment (MSA) of ΦX174-like phages were far lower than those of site-independently sampled phages from the same MSA (fig. S15G), indicating that the language models can go beyond simple sequence statistics during generation. To further interpret generated sequences in the context of ΦX174-like phages, we collected three datasets of natural phages to compare the generated genomes against: 134 ΦX174 variants from our unprocessed *Microviridae* data; 39 ΦX174 variants evolved over 180 days of laboratory evolution, equivalent to ~13,000 phage generations (*60*); and 16 ΦX174-like phage isolates environmentally sampled across the United States (*61*) (fig. S15H).

Relative to their nearest natural genomes, generated phages contained between 49 and 143 synonymous substitutions, 14 and 51 nonsynonymous substitutions, 0 and 3 indels, and 0 and 12 intergenic substitutions, with overall sequence similarities between 93.2 and 98.8% (Fig. 4D and fig. S15, E and I). Notably, natural genomes with

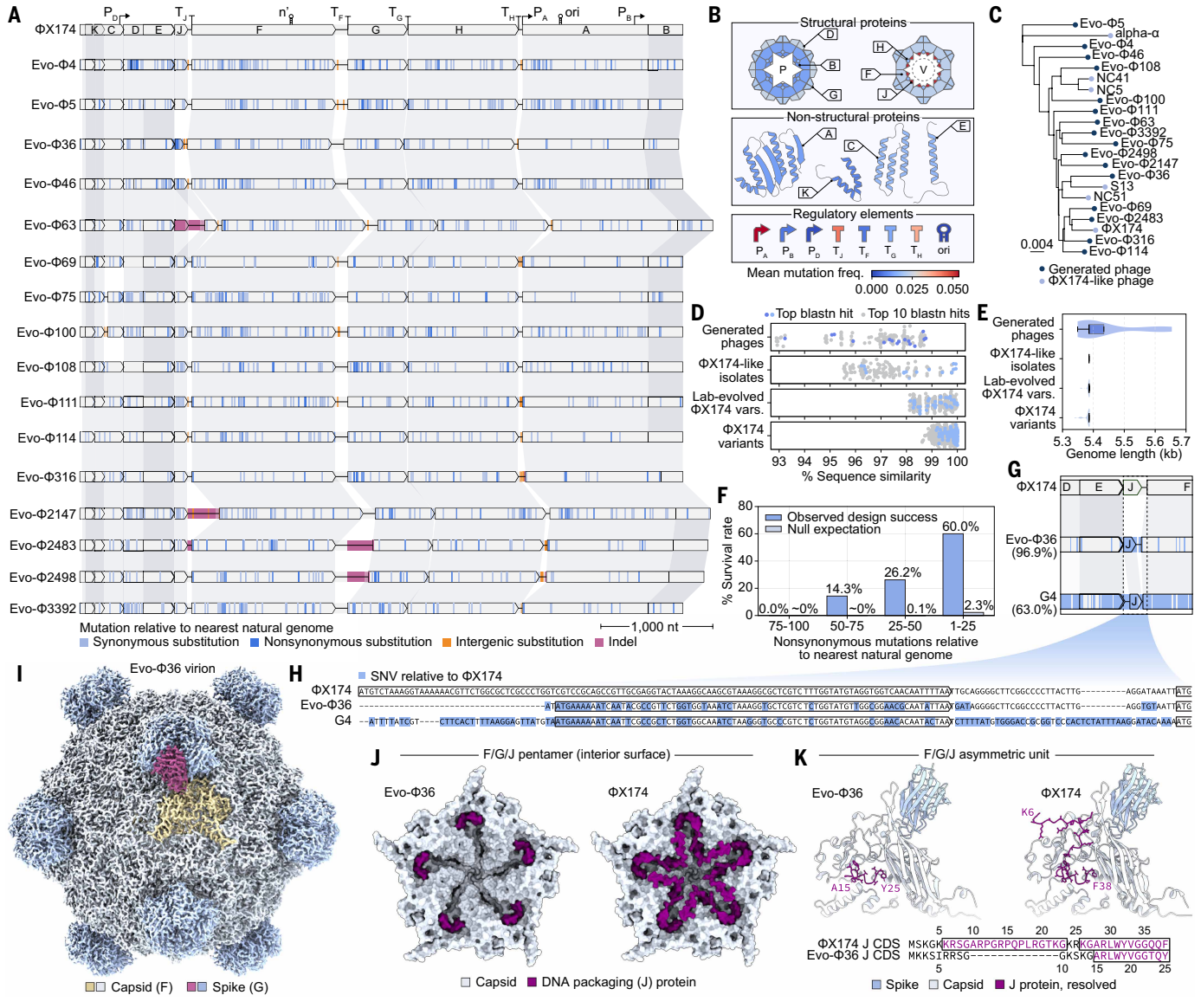


Fig. 4. Generated bacteriophages reveal sequence and structural insights. (**A**) Synteny plot of ΦX174 and functional generated bacteriophages, highlighting synonymous (light blue), nonsynonymous (dark blue), intergenic (yellow), and indel (pink) mutations compared with their nearest natural genomes. Genomes are in order by name. (**B**) Mean length–normalized nucleotide mutation frequencies of structural genes, nonstructural genes, and regulatory elements across the generated phage genomes show that gene J and regulatory elements are mutational hotspots. P, procapsid; V, virion; ori, origin of replication. (**C**) Neighbor-joining phylogenetic tree of viable generated phages (dark blue) and representative ΦX174-like phages (light blue). (**D** and **E**) Generated phages exhibit the lowest sequence similarity to nearest natural genomes (D) and the broadest range of genome lengths (E) compared with ΦX174 isolates harvested across the United States (*61*), ΦX174 variants (vars.) evolved over 180 days of laboratory evolution (*60*), and ΦX174 variants in our *Microviridae* dataset. Colored dots, top nucleotide blast (blastn) hit per phage, gray dots, top 10 blastn hits per phage. Box, interquartile range (IQR); thick black line, median; whiskers, 1.5× IQR. (**F**) The observed survival rates (dark blue) of generated phages were higher than expected survival rates (light blue) for nonsynonymous mutation counts. The null expectation was computed based on an empirically determined nonsynonymous mutation lethality of ~20% in ΦX174 (*16*) (materials and methods) averaged across the nonsynonymous mutation counts of phage genomes in each bin. (**G** and **H**) Synteny plot (G) and detailed view (H) of gene J and its surrounding intergenic regions of ΦX174, Evo-Φ36, and phage G4 with SNVs compared with ΦX174 highlighted in blue. (**I**) Cryo-EM density map of Evo-Φ36 virion highlighting individual subunits of the spike (pink) and capsid (yellow). Remaining spikes (light blue) and capsids (gray) are shown. (**J**) Interior surface view of the capsid (F, gray) and spike (G, not visible) pentamers of Evo-Φ36 (left) and ΦX174 (right) with their cognate J proteins (purple) reveals distinct capsid interactions and putative genome packaging modes between the two phages as modeled into the cryo-EM density map. (**K**) Asymmetric units, including F, G, and J of Evo-Φ36 (left) and ΦX174 (right), show resolved residues of J. Single-letter abbreviations for the amino acid residues referenced throughout this figure are as follows: Y, Tyr; A, Ala; K, Lys; F, Phe; M, Met; S, Ser; G, Gly; I, Ile; R, Arg; P, Pro; Q, Gln; L, Leu; T, Thr; W, Trp; V, Val.

<95% identity to any known phage would meet the criteria of a new species under some definitions (*62*). In comparison, ΦX174-like isolates and lab-evolved ΦX174 variants showed similar distributions of mutation types but more limited sequence similarity, whereas natural ΦX174 variants showed the lowest counts and diversity across all metrics. Generated phages showed the most diversity among their own group compared with the natural phage datasets as well as the broadest distribution of genome lengths, between 99 and 105% of the lengths of their nearest natural genomes (Fig. 4E and fig. S15J).

Previous work has shown that nonsynonymous substitutions have a ~20% lethality rate in ΦX174 and that the remaining nonlethal mutations still result in an average fitness defect of ~13% (*16*). These results

allowed us to estimate that the expected survival rate of mutated genomes would only be 2.3% on average for the generated genomes with <25 nonsynonymous mutations and close to 0.0% for higher mutational burdens (Fig. 4F and fig. S16A). Our generated phages tolerated far more nonsynonymous mutations than expected under this null model (even without considering the effects of epistasis and other mutation types, such as indels and noncoding mutations). The success rate of our tested genome designs was 60.0% (3/5) for genomes with up to 25 nonsynonymous mutations, 26.2% (11/42) for those with 25 to 50 nonsynonymous mutations, and 14.3% (2/14) for those with 50 to 75 nonsynonymous mutations. Most nonviable genomes contained well over hundreds of such mutations, making their survival probability extremely low (fig. S16, B to D).

Compared with natural ΦX174-like genomes, mutations in nonviable genomes were enriched in genes H, A, and B and depleted in gene F (fig. S16, E and F). These results suggest that Evo has implicitly learned constraints on functional phage genomes and that generative modeling may help identify success and failure modes in phage genome design.

Mutations in 13 of the generated genomes could not be recapitulated from any known natural sequences (fig. S17), including large noncoding insertions in Evo-Φ2147, Evo-Φ2483, and Evo-Φ2498. Through targeted deletion, we found context-dependent effects. Removal of the noncoding region in Evo-Φ2483 was lethal, whereas analogous deletions in Evo-Φ2147 and Evo-Φ2498 remained viable (fig. S18). We also found that deletion of the downstream gene J in Evo-Φ63 was lethal. These results suggest that model-generated insertions even in noncoding regions are not uniformly neutral, and their context-dependent effects are difficult to predict.

We next examined structural changes in the generated phages. We observed that in Evo-Φ36, gene J, a genome-packaging protein that also supports the capsid (*63*), was replaced with a shorter homologous protein found in phage G4, differing by four synonymous SNVs. Notably, Evo-Φ36 was the only tested phage with this J protein (fig. S19). Despite the similarity to G4 J, the J-F intergenic region and surrounding genomic context remained more similar to ΦX174 (Fig. 4, G and H). G4 is a distantly related *Microvirus* with only 63.0% sequence similarity to ΦX174 (fig. S20A) (*64*), whereas Evo-Φ36 shares 97.1% similarity with ΦX174. The G4 J protein is 25 amino acids long compared with the ΦX174 J protein, which is 38 amino acids long; the G4 J protein lacks portions of domains 0 and I that contribute to capsid binding in ΦX174 (fig. S20B) (*58*, *63*). Prior work has shown that swapping G4 J into ΦX174 is not viable (*57*–*59*); however, Evo-Φ36 is viable despite its similarity to ΦX174, which prompted us to test the functionality of Evo-Φ36 J in its nearest natural genome, NC51. Evo-Φ36 J was viable in NC51, whereas it remained lethal in ΦX174 (fig. S20, C to E), indicating up to 58 possible mutations across Evo-Φ36's genome that enabled its functionality (fig. S20, F and G). These results highlight Evo's ability to design genomes that integrate context-dependent genome packaging and structural compatibilities.

Given the unusually small J protein in Evo-Φ36, we sought to understand its structural consequences. AlphaFold 3 predicted different orientations of J in the capsid pentamers of Evo-Φ36 compared with those in ΦX174 and G4 (fig. S21) (*65*). To further explore these interactions, we solved the structure of ΦX174 to a resolution of 2.8 Å and that of Evo-Φ36 to a resolution of 2.9 Å by cryo–electron microscopy (cryo-EM) (Fig. 4I, figs. S22 to S24, and table S1). Consistent with previous findings (*66*), we observed that the hydrophobic C terminus of the J protein in ΦX174, domain II, interacts with the interior surface of the capsid, whereas the more basic domains 0 and I tether the genomic DNA while structurally supporting the capsid (Fig. 4, J and K, and fig. S25). Despite introducing more polarity at the capsid-J interface by substitution of the C-terminal phenylalanine (F) with a tyrosine (Y), Evo-Φ36 J maintains a similar overall binding mode at its C terminus (Fig. 4K). Notably, we could not resolve 14 residues at the N terminus of Evo-Φ36 J, whereas the corresponding residues in ΦX174 J visibly interact with the capsid. This difference could potentially be explained by particle integrity (*67*), or it could be because the Evo-Φ36 N terminus is unstructured with respect to the capsid and primarily functions in DNA binding. Taken together, these findings show that despite divergent sequence contexts, Evo-Φ36 J preserves a compatible interaction with its capsid, denoting how generative design can uncover new coevolutionary solutions.

Generated bacteriophages exhibit a broad range of fitness profiles

Generative genomics can propose genetic sequences that are not constrained by natural selection pressures. We therefore hypothesized that functional mutations in the generated phage genomes might manifest in diverse fitness profiles relative to each other and to natural phages (Fig. 5).

We first examined the infection dynamics of each phage by measuring how quickly and strongly each drove the host population to its minimum density, focusing on the timing, rate, and depth of lysis (Fig. 5A). Along with the generated phages and ΦX174, we rebooted 16 natural ΦX174-like phages (*61*) and measured their individual lytic dynamics (Fig. 5B and figs. S26 to S28) We found that phages Evo-Φ2483, Evo-Φ69, Evo-Φ108, and Evo-Φ111 tended to exhibit strong lytic phenotypes comparable to those of ΦX174-like phages WA10, NC51, S13, and NC41, among others. Evo-Φ63 generally exhibited the slowest and most shallow lysis compared with that of all other phages. Generated phages also showed a broader range of phenotypes across the measured lysis kinetics relative to that of natural phages (Fig. 5, B and C).

We next explored the dynamics of the generated phages when competing for host cells in the same population, representing relative fitness to each other in laboratory conditions (Fig. 5D). We coinfected ΦX174 and all generated phages in single *E. coli* C cultures by mixing the 16 generated phages and ΦX174 at equal multiplicity of infection (MOI) and measuring the cumulative fold change (FC) in sequencing read counts of each phage over time. Phages with the highest cumulative FC over time are those that had the largest relative increase in their own population size, indicating a relative fitness advantage.

We observed that three generated phages, Evo-Φ69, Evo-Φ100, and Evo-Φ111, outcompeted other phages in the population (Fig. 5E). Evo-Φ69 specifically had the highest FC in the mixed population compared with that of nearly all other generated phages and ΦX174, with cumulative FCs between ~10X and 41X after 6 hours of infection, whereas Evo-Φ36 and Evo-Φ108 had the largest decreases in FC over time. We found no correlation between the naturalness of each phage and their relative competitiveness (fig. S29), suggesting that naturalness is not a reliable predictor of fitness for generated phages. As a whole, these results support the ability of genome language models to design diverse mutations across complete genomes yielding a broad range of lytic kinetics and fitness profiles.

Generated bacteriophages rapidly overcome bacterial resistance

Phage therapy is emerging as a promising alternative to antibiotics, but its efficacy can be limited by rapid evolution of resistant bacteria (*33*). We hypothesized that the diverse phages produced by our genome design method could form a cocktail that more readily overcomes bacterial resistance (Fig. 6) (*29*, *31*, *68*). To investigate our hypothesis, we first evolved two ΦX174-resistant *E. coli* C cultures (Fig. 6A and fig. S30A; materials and methods). Whole-genome sequencing of isolated strains from each resistant *E. coli* C culture revealed that each strain independently developed mutations within the *waa* operon (Fig. 6B and fig. S30B), which is associated with lipopolysaccharide (LPS) synthesis. We called these strains CR1 and CR2. In particular, CR1 contained a missense mutation, L259W (Leu^{259} → Trp), in the *waaT* gene and strain CR2 contained a single base deletion at nucleotide position 485 of *waaT*, resulting in a premature stop codon followed

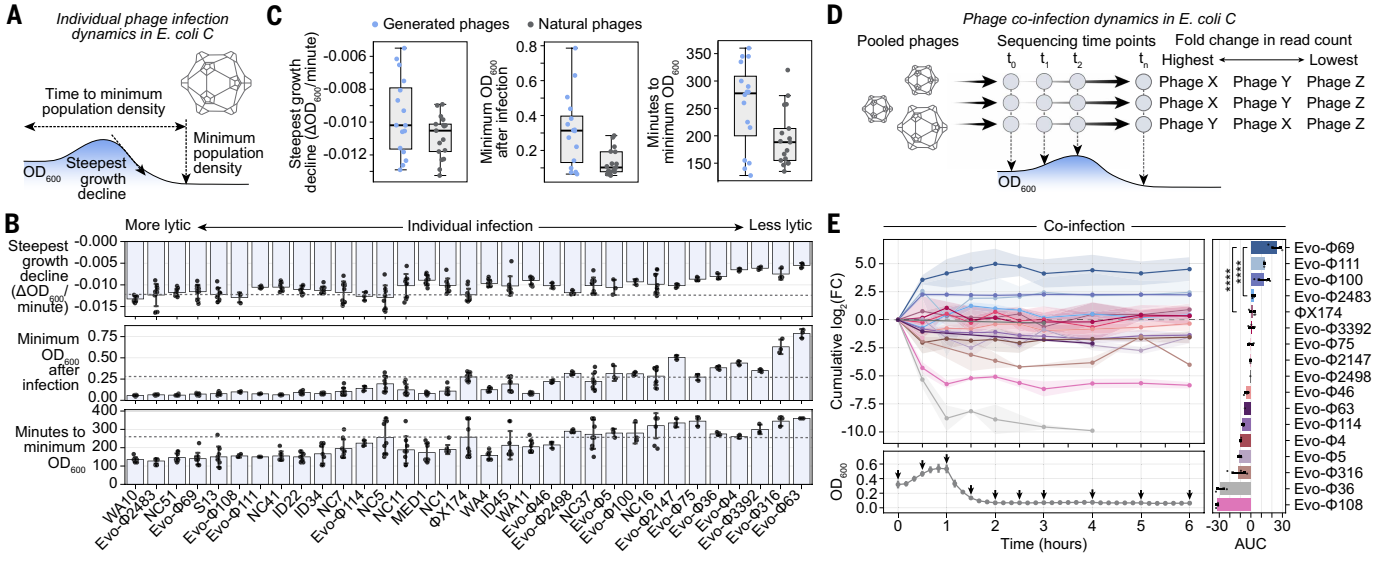


Fig. 5. Generated bacteriophages exhibit a broad range of fitness profiles. (**A**) We measured individual phage lytic dynamics in infected *E. coli* C by slope of decline in host growth rate, minimum population density (OD_{600}) after infection, and time to minimum population density. (**B**) Some generated phages exhibit strong lytic rates relative to those of ΦX174 and ΦX174-like phages. Bar height, mean; error bars, standard deviation; circles, $N = 3$ to 12 infections; dotted line, mean value of ΦX174. (**C**) Generated phages also exhibit a broad range of lytic phenotypes relative to natural phages. Data point, mean value; box, interquartile range (IQR); thick black line, median; whiskers, 1.5× IQR. (**D**) We measured phage coinfection dynamics in *E. coli* C by sequencing relative abundances of each phage in pooled cultures at various time points. (**E**) We tracked cumulative FC [$\log_2$(FC)] of sequencing read counts of generated phages and ΦX174 (top left) coinfected in *E. coli* C at equal MOI over 6 hours (bottom left). Evo-Φ69 outcompetes ΦX174 and other generated phages, as measured by area under the curve (AUC) of the cumulative $\log_2$(FC) (right). (Top left) Data point, mean cumulative $\log_2$(FC) value; shading, standard deviation. (Bottom left) Data point, mean OD_{600} value; error bars, standard deviation; arrowhead, sequencing sample extraction time point. (Right) Bar height, mean AUC; circles, competition AUC values; error bars, standard deviation. $N = 3$ competitions. Statistical significance was determined by one-way analysis of variance with Tukey's post hoc test (****adjusted $P < 0.0001$).

by potential reinitiation at a downstream in-frame start codon. Mutations in *waaT* have previously been observed to confer resistance against ΦX174 (*69*), suggesting that strains CR1 and CR2 likely exhibit resistance through modifications to LPS synthesis.

Next, we explored whether ΦX174 only, a natural phage cocktail consisting of a mixture of 16 ΦX174-like phages and ΦX174 (fig. S26), or a mixture of all 16 generated phages and ΦX174 could overcome resistance in strains CR1 and CR2. We challenged three separate cultures: ΦX174-susceptible cells, ΦX174-resistant cells, and a mixture of the two to create a selective pressure for phages replicating in susceptible cells to evolve the ability to infect resistant cells (Fig. 6C). After each challenge, we passaged the supernatant from the mixed strains into fresh cultures and observed their growth dynamics.

Upon initial infection, the generated phage cocktail, natural phage cocktail, and ΦX174 successfully inhibited growth of the susceptible strain but did not inhibit growth of the resistant strains (Fig. 6D). The generated phage cocktail was able to successfully inhibit growth of strain CR1 after a single passage and CR2 after two passages and sustain growth inhibition for at least five passages. By contrast, ΦX174 alone and the natural phage cocktail both could not inhibit growth of any resistant strains even after five passages, suggesting insufficient adaptive variation in their genetic diversity.

We isolated and sequenced individual phages that successfully grew on strains CR1 and CR2, which revealed a single predominant phage genome responsible for overcoming each resistant strain (materials and methods), which we designated Evo-ΦR1 and Evo-ΦR2, respectively. Sequence alignment revealed that all of Evo-ΦR1's genome could be collectively composed from five segments of Evo-Φ111 and Evo-Φ114 except for two SNVs in the major capsid protein (Fig. 6E). All of Evo-ΦR2's genome could be composed from four segments of Evo-Φ111, Evo-Φ114, and Evo-Φ2147 except for a single SNV in the major capsid protein (Fig. 6F). The ability to overcome resistance likely arose from recombination and acquired mutation events involving two or three of the generated phages.

Given the resistant mutations in the LPS gene synthesis operons of strains CR1 and CR2, we speculated that key mutations in the predominant resistant phages likely decorated the outer surfaces of their major capsid and spike proteins, as these proteins modulate LPS binding (*47*). Indeed, for both Evo-ΦR1 and Evo-ΦR2, we observed 15 nonsynonymous mutations across the capsid and spike proteins that were not present in ΦX174, 14 of which were generated by Evo (Fig. 6, E and F). The majority of the mutations appear on the outer surface of the virions, as predicted by AlphaFold 3 (Fig. 6G and fig. S30, C and D) (*65*). These data show that evolutionary innovations generated by Evo likely contributed to resilience against bacterial resistance and, more broadly, suggest the utility of generative models for producing genetically diverse phage cocktails that could translate into improved therapeutic efficacy.

Discussion

In this work, we leveraged genome language models to achieve the first generative design of complete bacteriophage genomes. We established a computational framework for specifying our design goals, including the development of a gene annotation method and various scoring metrics, allowing us to controllably design toward a target genomic architecture and host tropism. In contrast with fully de novo sequence generation, we chose ΦX174, a tractable, safe, and historically notable model genome (*13*, *36*, *39*, *40*) as a design template. We systematically evaluated thousands of such computationally generated sequences and experimentally tested nearly 300 designs, resulting in 16 viable phages containing considerable genomic diversity and enabling a phage cocktail that rapidly overcame bacterial resistance. The

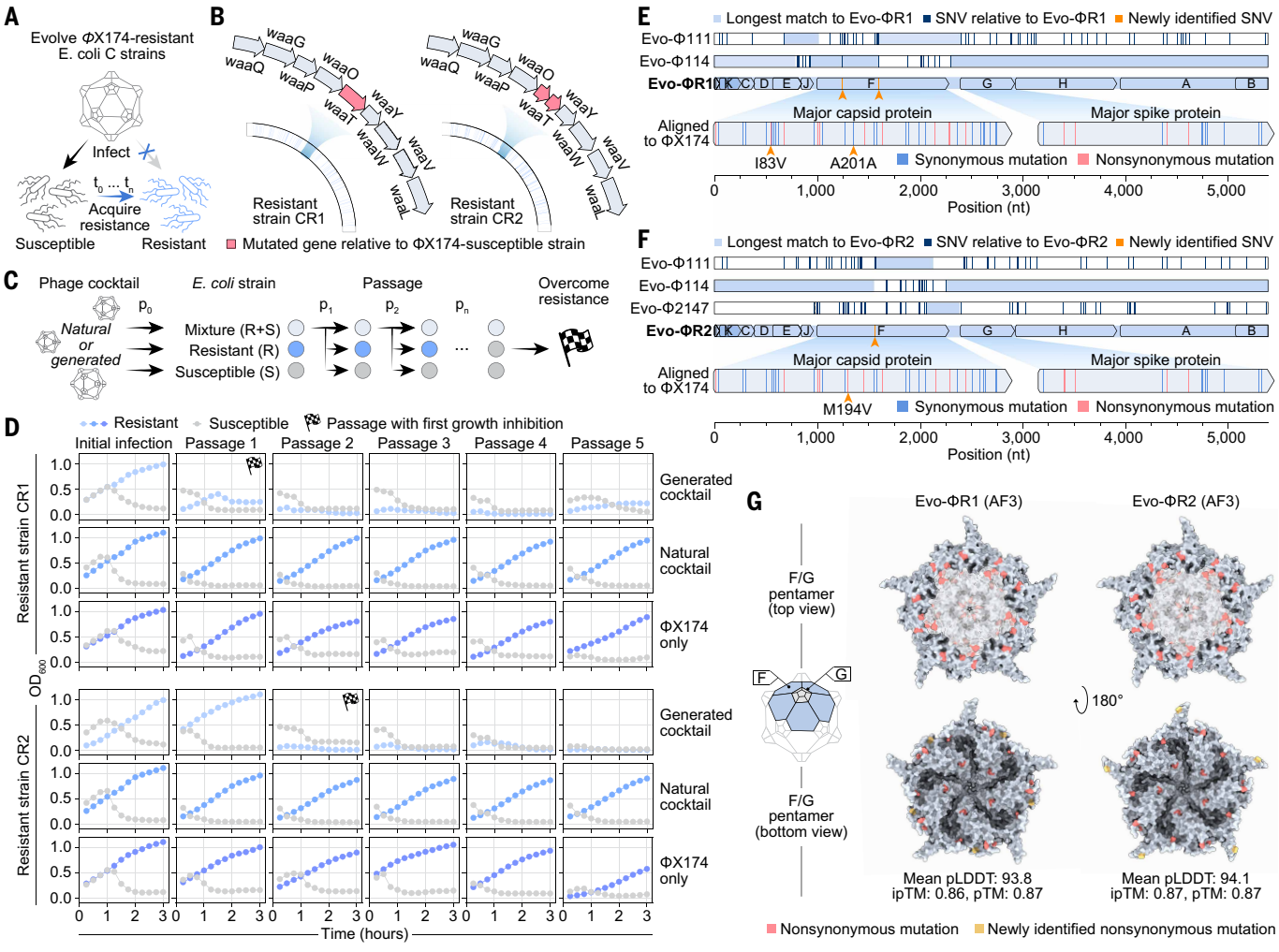


Fig. 6. Generated bacteriophages rapidly overcome bacterial resistance. (**A**) We evolved ΦX174-resistant *E. coli* C strains by long-term culture until resistant clones overtook the bacterial population. (**B**) Whole-genome sequencing of two ΦX174-resistant *E. coli* C strains revealed mutations in the *waa* operon absent in susceptible *E. coli* C, which functions in LPS receptor synthesis. (**C**) Experimental setup for evolving phage counter-resistance by serially passaging cocktails of phages on susceptible and resistant *E. coli* C. (**D**) Growth curves show that a cocktail of ΦX174-like phages (natural cocktail) and ΦX174 alone fail to overcome resistance, whereas generated phage cocktails suppress growth and sustain infection of all resistant cultures across five passages. Checkered flag, first passage with growth inhibition. (**E** and **F**) Alignments of generated phages against the predominant resistant phages Evo-ΦR1 (E) and Evo-ΦR2 (F) capable of infecting resistant strains CR1 and CR2, respectively, show that they are derived from generated genomes. Generated phages used in the alignments are those with the longest identical sequences (light blue) without SNVs (dark blue) to each resistant phage such that they collectively minimize the number of newly identified mutations (yellow) observed in the resistant phage. Major capsid and spike proteins of each resistant phage aligned to those of ΦX174 are below, with synonymous (blue) and nonsynonymous mutations (pink) relative to ΦX174. (**G**) AlphaFold 3 (AF3) predictions of capsid (light blue) and spike (light gray) pentamers show that most capsid and spike mutations appear on the exterior of resistant phages. Nonsynonymous and newly identified nonsynonymous mutations relative to ΦX174 are highlighted in pink and yellow, respectively. ipTM, interface predicted template modeling score; pTM, predicted template modeling score.

generated phages exhibited broader ranges of lysis kinetics relative to natural ΦX174-like phages, demonstrating the ability of DNA language models to generate divergent genotypes and phenotypes at the genome scale.

Synthetic genomics has historically relied on directed evolution, random mutagenesis, or rational engineering (*70*, *71*). These approaches have been limited in the scope of their achieved evolutionary novelty owing to the complexity of genome sequences and the limited throughput of methods for genome editing and synthesis. Rational engineering is further limited by incomplete human understanding of biology; for example, previous systematic efforts have struggled to increase features such as phage lysis rate or genome length (*72*–*74*). By contrast, our approach enabled designs with broad variation in both genotype and phenotype that would have been otherwise difficult to achieve, including the genome of Evo-Φ36, with a contextually functional truncated J gene, the genome of Evo-Φ63, that is 5% (268 nucleotides) longer than ΦX174; Evo-Φ69, which was highly competitive in coinfection assays; and Evo-Φ2483, which exhibited fast lysis rates among natural phages. Notably, despite being in a highly constrained evolutionary setting, the genome of Evo-Φ2147 has ~93% sequence similarity to known natural genomes. Some sequence-based taxonomic guidelines use <95% sequence similarity as one threshold to delineate new bacteriophage species (*62*).

The capability to generate new phage genomes with AI systems also raises important biosafety, biocontainment, and biosecurity considerations (*75*) necessitating discourse on the strengths and limitations of governance, policies, and misuse mitigation strategies (*76*). Groups conducting future whole-genome design work should consult both safety and security professionals throughout the project lifecycle. In

line with longstanding precedent for conducting biological research within established biosafety levels (*77*), we performed all experiments at the biosafety level appropriate for research with bacteriophages and their nonpathogenic bacterial hosts with supplementary precautions (materials and methods). Alongside ongoing efforts in developing more effective biosafety and biosecurity approaches (*78*–*80*), such frameworks can be adapted and applied to the generative design of new biological systems, especially when, as in this work, designs are constrained by well-characterized natural genomes as templates. Moreover, as we have previously demonstrated, the generative models themselves can possess inherent safeguards based on their training data; for example, we have previously shown that data exclusions successfully prevent the pretrained Evo 2 models from designing eukaryotic viruses, including pathogenic human viruses (*26*). We have provided further commentary on this in the supplementary text. By continuing to build upon robust safety and security frameworks, the field can responsibly unlock the potential of generative models to access and engineer complex biological functions for the benefit of science and society.

We envision that the further development of generative genomics for bacteriophage genome design will require additional methodological innovations and improved training datasets. Future approaches could leverage techniques for improving and accelerating model conditioning beyond SFT alone (*81*–*84*). Continual genomic and metagenomic sequencing projects could also contribute additional training data. Designing larger phages will pose additional challenges, particularly in cost-effective DNA synthesis and assembly. Techniques for multifragment genome assembly, combinatorial oligonucleotide pools, and in vitro transcription-translation systems are promising avenues for realizing more complex genome design (*52*, *85*, *86*).

Beyond biotechnological or therapeutic utility, generative design of bacteriophage genomes offers distinctive opportunities for studying evolution (*87*). By systematically sampling from learned genomic distributions, it is possible to explore large mutational landscapes and investigate the genetic basis of properties such as host tropism. Accessing adaptive variation through generative design may also prove useful (*88*). Indeed, we observed subsequent recombination and mutation of our generated phages that likely conferred resilience against host resistance. The ability to rapidly design phage genomes tuned for host range, fitness, and resistance evasion may expand biotechnological toolkits and transform phage therapy pipelines, enabling more adaptive and resilient antimicrobial strategies (*29*, *68*, *89*, *90*).

In total, our results demonstrate that generative AI can capture an underlying evolutionary design space with enough fidelity to produce viable bacteriophage genomes. Continued progress will likely bring about more generalizable models capable of designing across diverse biological systems with desirable functional properties. The rapid pace of improvement in generative biology suggests a future where genome design could become a core biotechnology alongside genome sequencing, synthesis, and editing, possibly enabling the generation of larger and more complex genomes.

The materials and methods are available in the supplementary materials.

REFERENCES AND NOTES:

1. R. C. Edgar *et al.*, Petabase-scale sequence alignment catalyses viral discovery. *Nature* **602**, 142–147 (2022). doi: 10.1038/s41586-021-04332-2; pmid: 35082445
2. D. G. Gibson *et al.*, Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319**, 1215–1220 (2008). doi: 10.1126/science.1151721; pmid: 18218864
3. C. A. Hutchison III *et al.*, Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016). doi: 10.1126/science.aad6253; pmid: 27013737
4. S. M. Richardson *et al.*, Design of a synthetic yeast genome. *Science* **355**, 1040–1044 (2017). doi: 10.1126/science.aaf4557; pmid: 28280199
5. K. Wang *et al.*, Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016). doi: 10.1038/nature20124; pmid: 27776354
6. R. M. Dedrick *et al.*, Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nat. Med.* **25**, 730–733 (2019). doi: 10.1038/s41591-019-0437-z; pmid: 31068712
7. M. G. Durrant *et al.*, Bridge RNAs direct programmable recombination of target and donor DNA. *Nature* **630**, 984–993 (2024). doi: 10.1038/s41586-024-07552-4; pmid: 38926615
8. D. J. Mandell *et al.*, Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55–60 (2015). doi: 10.1038/nature14121; pmid: 25607366
9. A. Nyerges *et al.*, A swapped genetic code prevents viral infections and gene transfer. *Nature* **615**, 720–727 (2023). doi: 10.1038/s41586-023-05824-z; pmid: 36922599
10. M. Costanzo *et al.*, A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016). doi: 10.1126/science.aaf1420; pmid: 27708008
11. S. F. Elena, R. E. Lenski, Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398 (1997). doi: 10.1038/37108; pmid: 9389477
12. F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961). doi: 10.1016/S0022-2836(61)80072-7; pmid: 13718526
13. B. G. Barrell, G. M. Air, C. A. Hutchison 3rd, Overlapping genes in bacteriophage phiX174. *Nature* **264**, 34–41 (1976). doi: 10.1038/264034a0; pmid: 1004533
14. C. A. Hutchison III *et al.*, Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **286**, 2165–2169 (1999). doi: 10.1126/science.286.5447.2165; pmid: 10591650
15. R. Sanjuán, A. Moya, S. F. Elena, The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8396–8401 (2004). doi: 10.1073/pnas.0400146101; pmid: 15159545
16. P. Domingo-Calap, J. M. Cuevas, R. Sanjuán, The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLOS Genet.* **5**, e1000742 (2009). doi: 10.1371/journal.pgen.1000742; pmid: 19956760
17. J. Dauparas *et al.*, Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022). doi: 10.1126/science.add2187; pmid: 36108050
18. J. Fredens *et al.*, Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019). doi: 10.1038/s41586-019-1192-5; pmid: 31092918
19. T. Hayes *et al.*, Simulating 500 million years of evolution with a language model. *Science* **387**, 850–858 (2025). doi: 10.1126/science.ads0018; pmid: 39818825
20. B. L. Hie *et al.*, Efficient evolution of human antibodies from general protein language models. *Nat. Biotechnol.* **42**, 275–283 (2024). doi: 10.1038/s41587-023-01763-2; pmid: 37095349
21. J. B. Ingraham *et al.*, Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023). doi: 10.1038/s41586-023-06728-8; pmid: 37968394
22. K. Jiang *et al.*, Rapid in silico directed evolution by a protein language model with EVOLVEpro. *Science* **387**, eadr6006 (2025). doi: 10.1126/science.adr6006; pmid: 39571002
23. P. Srinivasan, C. D. Smolke, Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* **585**, 614–619 (2020). doi: 10.1038/s41586-020-2650-9; pmid: 32879484
24. J. Kaplan *et al.*, Scaling laws for neural language models. arXiv:2001.08361 [cs.LG] (2020).
25. A. Vaswani *et al.*, "Attention is all you need" in *Advances in Neural Information Processing Systems* (Curran Associates Inc., 2017), pp. 6000–6010.
26. G. Brixi *et al.*, Genome modelling and design across all domains of life with Evo 2. *Nature* **652**, 1349–1361 (2026). doi: 10.1038/s41586-026-10176-5; pmid: 41781614
27. A. T. Merchant, S. H. King, E. Nguyen, B. L. Hie, Semantic design of functional de novo genes from a genomic language model. *Nature* **649**, 749–758 (2026). doi: 10.1038/s41586-025-09749-7; pmid: 41261132
28. E. Nguyen *et al.*, Sequence modeling and design from molecular to genome scale with Evo. *Science* **386**, eado9336 (2024). doi: 10.1126/science.ado9336; pmid: 39541441
29. M. K. Kim *et al.*, Bacteriophage therapy for multidrug-resistant infections: Current technologies and therapeutic approaches. *J. Clin. Invest.* **135**, e187996 (2025). doi: 10.1172/JCI187996; pmid: 40026251
30. S. Kilcher, M. J. Loessner, Engineering bacteriophages as versatile biologics. *Trends Microbiol.* **27**, 355–367 (2019). doi: 10.1016/j.tim.2018.09.006; pmid: 30322741
31. D. P. Pires, S. Cleto, S. Sillankorva, J. Azeredo, T. K. Lu, Genetically engineered phages: A review of advances over the last decade. *Microbiol. Mol. Biol. Rev.* **80**, 523–543 (2016). doi: 10.1128/MMBR.00069-15; pmid: 27250768
32. B. Shao, J. Yan, A long-context language model for deciphering and generating bacteriophage genomes. *Nat. Commun.* **15**, 9392 (2024). doi: 10.1038/s41467-024-53759-4; pmid: 39477977
33. S. A. Strathdee, G. F. Hatfull, V. K. Mutalik, R. T. Schooley, Phage therapy: From biological mechanisms to future directions. *Cell* **186**, 17–31 (2023). doi: 10.1016/j.cell.2022.11.017; pmid: 36608652
34. P. R. Jaschke, G. A. Dotson, K. S. Hung, D. Liu, D. Endy, Definitive demonstration by synthesis of genome annotation completeness. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24206–24213 (2019). doi: 10.1073/pnas.1905990116; pmid: 31719208
35. D. Y. Logel, P. R. Jaschke, A high-resolution map of bacteriophage ϕX174 transcription. *Virology* **547**, 47–56 (2020). doi: 10.1016/j.virol.2020.05.008; pmid: 32560904

36. F. Sanger *et al.*, Nucleotide sequence of bacteriophage φ X174 DNA. *Nature* **265**, 687–695 (1977). doi: 10.1038/265687a0; pmid: 870828
37. J. Shlomai, A. Kornberg, An Escherichia coli replication protein that recognizes a unique sequence within a hairpin region in phi X174 DNA. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 799–803 (1980). doi: 10.1073/pnas.77.2.799; pmid: 6444725
38. P. C. Kirchberger, H. Ochman, Microviruses: A World Beyond *phi*X174. *Annu. Rev. Virol.* **10**, 99–118 (2023). doi: 10.1146/annurev-virology-100120-011239; pmid: 37774127
39. H. O. Smith, C. A. Hutchison 3rd, C. Pfannkoch, J. C. Venter, Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15440–15445 (2003). doi: 10.1073/pnas.2237126100; pmid: 14657399
40. M. Goulian, A. Kornberg, R. L. Sinsheimer, Enzymatic synthesis of DNA, XXIV. Synthesis of infectious phage phi-X174 DNA. *Proc. Natl. Acad. Sci. U.S.A.* **58**, 2321–2328 (1967). doi: 10.1073/pnas.58.6.2321; pmid: 4873588
41. P. R. Jaschke, E. K. Lieberman, J. Rodriguez, A. Sierra, D. Endy, A fully decompressed synthetic bacteriophage øX174 genome assembled and archived in yeast. *Virology* **434**, 278–284 (2012). doi: 10.1016/j.virol.2012.09.020; pmid: 23079106
42. A. P. Camargo *et al.*, Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024). doi: 10.1038/s41587-023-01953-y; pmid: 37735266
43. C. Camacho *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). doi: 10.1186/1471-2105-10-421; pmid: 20003500
44. Z. Lin *et al.*, Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023). doi: 10.1126/science.ade2574; pmid: 36927031
45. P. Terzian *et al.*, PHROG: Families of prokaryotic virus proteins clustered using remote homology. *NAR Genom. Bioinform.* **3**, lqab067 (2021). doi: 10.1093/nargab/lqab067; pmid: 34377978
46. J. R. Brister, D. Ako-Adjei, Y. Bao, O. Blinkova, NCBI viral genomes resource. *Nucleic Acids Res.* **43** (D1), D571–D577 (2015). doi: 10.1093/nar/gku1207; pmid: 25428358
47. A. Michel, O. Clermont, E. Denamur, O. Tenaillon, Bacteriophage PhiX174's ecological niche and the flexibility of its Escherichia coli lipopolysaccharide receptor. *Appl. Environ. Microbiol.* **76**, 7310–7313 (2010). doi: 10.1128/AEM.02721-09; pmid: 20833781
48. B. W. Wright, M. P. Molloy, P. R. Jaschke, Overlapping genes in natural and engineered genomes. *Nat. Rev. Genet.* **23**, 154–168 (2022). doi: 10.1038/s41576-021-00417-w; pmid: 34611352
49. Y. Sun *et al.*, Structural changes of tailless bacteriophage ΦX174 during penetration of bacterial cell walls. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13708–13713 (2017). doi: 10.1073/pnas.1716614114; pmid: 29229840
50. R. K. Peet, The measurement of species diversity. *Annu. Rev. Ecol. Syst.* **5**, 285–307 (1974). doi: 10.1146/annurev.es.05.110174.001441
51. S. Nayfach *et al.*, CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021). doi: 10.1038/s41587-020-00774-7; pmid: 33349699
52. M. S. Faber *et al.*, Saturation Mutagenesis Genome Engineering of Infective ΦX174 Bacteriophage *via* Unamplified Oligo Pools and Golden Gate Assembly. *ACS Synth. Biol.* **9**, 125–131 (2020). doi: 10.1021/acssynbio.9b00411; pmid: 31825605
53. J. T. Van Leuven *et al.*, ΦX174 attenuation by whole-genome codon deoptimization. *Genome Biol. Evol.* **13**, evaa214 (2021). doi: 10.1093/gbe/evaa214; pmid: 33045052
54. W. Fiers, R. L. Sinsheimer, The structure of the DNA of bacteriophage φ·X174. III. Ultracentrifugal evidence for a ring structure. *J. Mol. Biol.* **5**, 424–434 (1962). doi: 10.1016/S0022-2836(62)80031-X; pmid: 13945085
55. S. Wickner, J. Hurwitz, Conversion of phiX174 viral DNA to double-stranded form by purified Escherichia coli proteins. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4120–4124 (1974). doi: 10.1073/pnas.71.10.4120; pmid: 4610569
56. C. T. Archer *et al.*, The genome sequence of E. coli W (ATCC 9637): Comparative genome analysis and an improved genome-scale reconstruction of E. coli. *BMC Genomics* **12**, 9 (2011). doi: 10.1186/1471-2164-12-9; pmid: 21208457
57. B. A. Fane, S. Head, M. Hayashi, Functional relationship between the J proteins of bacteriophages phi X174 and G4 during phage morphogenesis. *J. Bacteriol.* **174**, 2717–2719 (1992). doi: 10.1128/jb.174.8.2717-2719.1992; pmid: 1532571
58. E. T. Ogunbunmi, A. P. Roznowski, B. A. Fane, The effects of packaged, but misguided, single-stranded DNA genomes are transmitted to the outer surface of the φX174 capsid. *J. Virol.* **95**, e0088321 (2021). doi: 10.1128/JVI.00883-21; pmid: 34232738
59. A. P. Roznowski, S. M. Doore, S. Z. Kemp, B. A. Fane, Finally, a role befitting Astar: Strongly conserved, unessential microvirus A* proteins ensure the product fidelity of packaging reactions. *J. Virol.* **94**, 10–1128 (2020). doi: 10.1128/JVI.01593-19; pmid: 31666371
60. H. A. Wichman, J. Millstein, J. J. Bull, Adaptive molecular evolution for 13,000 phage generations: A possible arms race. *Genetics* **170**, 19–31 (2005). doi: 10.1534/genetics.104.034488; pmid: 15687276
61. D. R. Rokyta, C. L. Burch, S. B. Caudle, H. A. Wichman, Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J. Bacteriol.* **188**, 1134–1142 (2006). doi: 10.1128/JB.188.3.1134-1142.2006; pmid: 16428417
62. D. Turner, A. M. Kropinski, E. M. Adriaenssens, A roadmap for genome-based phage taxonomy. *Viruses* **13**, 506 (2021). doi: 10.3390/v13030506; pmid: 33803862
63. R. A. Bernal, S. Hafenstein, R. Esmeralda, B. A. Fane, M. G. Rossmann, The phiX174 protein J mediates DNA packaging and viral attachment to host cells. *J. Mol. Biol.* **337**, 1109–1122 (2004). doi: 10.1016/j.jmb.2004.02.033; pmid: 15046981
64. G. N. Godson, B. G. Barrell, R. Staden, J. C. Fiddes, Nucleotide sequence of bacteriophage G4 DNA. *Nature* **276**, 236–247 (1978). doi: 10.1038/276236a0; pmid: 714153
65. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). doi: 10.1038/s41586-024-07487-w; pmid: 38718835
66. R. McKenna *et al.*, Atomic structure of single-stranded DNA bacteriophage phi X174 and its functional implications. *Nature* **355**, 137–143 (1992). doi: 10.1038/355137a0; pmid: 1370343
67. W. Hu *et al.*, Structural basis for Salmonella infection by two Microviridae phages. *Commun. Biol.* **8**, 1166 (2025). doi: 10.1038/s42003-025-08595-7; pmid: 40770068
68. J. P. Pirnay, Phage therapy in the year 2035. *Front. Microbiol.* **11**, 1171 (2020). doi: 10.3389/fmicb.2020.01171; pmid: 32582107
69. J. Romeyer Dherbey, L. Parab, J. Gallie, F. Bertels, Stepwise evolution of E. coli C and ΦX174 reveals unexpected lipopolysaccharide (LPS) diversity. *Mol. Biol. Evol.* **40**, 154 (2023). doi: 10.1093/molbev/msad154; pmid: 37399035
70. A. L. V. Coradini, C. B. Hull, I. M. Ehrenreich, Building genomes to understand biology. *Nat. Commun.* **11**, 6177 (2020). doi: 10.1038/s41467-020-19753-2; pmid: 33268788
71. J. S. James, J. Dai, W. L. Chew, Y. Cai, The design and engineering of synthetic genomes. *Nat. Rev. Genet.* **26**, 298–319 (2025). doi: 10.1038/s41576-024-00786-y; pmid: 39506144
72. A. Aoyama, M. Hayashi, Effects of genome size on bacteriophage phi X174 DNA packaging in vitro. *J. Biol. Chem.* **260**, 11033–11038 (1985). doi: 10.1016/S0021-9258(17)39144-5; pmid: 3161888
73. D. Endy, L. You, J. Yin, I. J. Molineux, Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5375–5380 (2000). doi: 10.1073/pnas.090101397; pmid: 10792041
74. P. W. Russell, U. R. Müller, Construction of bacteriophage luminal diameterX174 mutants with maximum genome sizes. *J. Virol.* **52**, 822–827 (1984). doi: 10.1128/jvi.52.3.822-827.1984; pmid: 6092714
75. M. Wang *et al.*, A call for built-in biosecurity safeguards for generative AI tools. *Nat. Biotechnol.* **43**, 845–847 (2025). doi: 10.1038/s41587-025-02650-8; pmid: 40295784
76. D. Bloomfield *et al.*, AI and biosecurity: The need for governance. *Science* **385**, 831–833 (2024). doi: 10.1126/science.adq1977; pmid: 39172825
77. P. Berg, D. Baltimore, S. Brenner, R. O. Roblin 3rd, M. F. Singer, Asilomar conference on recombinant DNA molecules. *Science* **188**, 991–994 (1975). doi: 10.1126/science.1056638; pmid: 1056638
78. D. Bloomfield *et al.*, Securing dual-use pathogen data of concern. arXiv:2602.08061 [cs.AI] (2026).
79. B. D. Trump *et al.*, Building biosecurity for synthetic biology. *Mol. Syst. Biol.* **16**, e9723 (2020). doi: 10.15252/msb.20209723; pmid: 32692486
80. B. J. Wittmann *et al.*, Strengthening nucleic acid biosecurity screening against generative protein design tools. *Science* **390**, 82–87 (2025). doi: 10.1126/science.adu8578; pmid: 41037625
81. E. J. Hu *et al.*, LoRA: Low-rank adaptation of large language models. *Int. Conf. Learn. Represent.* (2022).
82. P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks" in *Advances in Neural Information Processing Systems* (Curran Associates Inc., 2020), pp. 9459–9474.
83. V. Mnih *et al.*, Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015). doi: 10.1038/nature14236; pmid: 25719670
84. T. Widatalla, R. Rafailov, B. Hie, Aligning protein generative models with experimental fitness via direct preference optimization. bioRxiv 2024.05.20.595026 [Preprint] (2024). https://doi.org/10.1101/2024.05.20.595026.
85. A. Levrier *et al.*, PHEIGES: All-cell-free phage synthesis and selection from engineered genomes. *Nat. Commun.* **15**, 2223 (2024). doi: 10.1038/s41467-024-46585-1; pmid: 38472230
86. J. M. Pryor, V. Potapov, K. Bilotti, N. Pokhrel, G. J. S. Lohman, Rapid 40 kb genome construction from 52 parts through data-optimized assembly design. *ACS Synth. Biol.* **11**, 2036–2042 (2022). doi: 10.1021/acssynbio.1c00525; pmid: 35613368
87. E. Shaer Tamar, R. Kishony, Multistep diversification in spatiotemporal bacterial-phage coevolution. *Nat. Commun.* **13**, 7971 (2022). doi: 10.1038/s41467-022-35351-w; pmid: 36577749
88. J. Zheng, J. L. Payne, A. Wagner, Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* **365**, 347–353 (2019). doi: 10.1126/science.aax1837; pmid: 31346060
89. V. K. Mutalik, A. P. Arkin, A phage foundry framework to systematically develop viral countermeasures to combat antibiotic-resistant bacterial pathogens. *iScience* **25**, 104121 (2022). doi: 10.1016/j.isci.2022.104121; pmid: 35402883
90. E. Wimmer, S. Mueller, T. M. Tumpey, J. K. Taubenberger, Synthetic viruses: A new opportunity to understand and prevent viral disease. *Nat. Biotechnol.* **27**, 1163–1172 (2009). doi: 10.1038/nbt.1593; pmid: 20010599

91. S. H. King, Training data for paper "Generative design of bacteriophages with genome language models," Zenodo (2025); https://doi.org/10.5281/zenodo.17101843.
92. S. H. King, Code for paper "Generative design of bacteriophages with genome language models," Zenodo (2026); https://doi.org/10.5281/zenodo.20484521.

ACKNOWLEDGEMENTS

We thank B. Ameglio, P. Bollyky, A. Borah, C. Chiu, S. Chithrananda, D. Endy, S. Evans, B. Fane, P. Hsu, S. Mille-Fragoso, B. Plosky, J. Sacher, and S. Somolinos for helpful discussions and support with the manuscript. We thank J. Ku and J. Sullivan for assistance with computational infrastructure. We thank Twist Bioscience for support with DNA synthesis. We thank Arc Institute LabOps for assistance with experimental setups. We thank A. Borah, C. Chiu, J. Noh, the Gao lab, the Hsu lab, and the Li lab for generous sharing of materials and reagents. We thank J. Noh, H. Wang, and the Stanford University Cryo-Electron Microscopy Center (cEMc) for assistance with cryo-EM. **Funding:** S.H.K., A.T.M., and G.B. acknowledge funding support from the National Science Foundation Graduate Research Fellowship Program. D.B.L. acknowledges funding support from the Fannie and John Hertz Foundation. A.T.M. acknowledges funding support from the Knight-Hennessy Graduate Scholarship Fund. B.L.H. acknowledges funding support from the Gates Foundation, the Chan Zuckerberg Initiative, the Advanced Research Projects Agency for Health (ARPA-H), Arc Institute, Schmidt Sciences AI2050, Stanford Center for Digital Health, Stanford Institute for Human-Centered Artificial Intelligence (HAI) Hoffman-Yee Research Grants. **Author contributions:** S.H.K. and B.L.H. conceived and designed the study. B.L.H. supervised the project. S.H.K. and B.L.H. collected the training data. S.H.K., G.B., and B.L.H fine-tuned the models. S.H.K. and D.G. performed the phage genome generation and evaluation. S.H.K., C.L.D., and D.G. performed the phage genome assembly transformation assays. S.H.K. and C.L.D. performed the phage sequencing, propagation, titering, and host tropism assay. S.H.K. and A.T.M. performed the phylogenetic analysis. S.H.K., C.L.D., and D.B.L. performed the cryo-EM experiments. D.B.L. processed the cryo-EM data. S.H.K., D.B.L., and M.E.W. analyzed the cryo-EM data. S.H.K. and A.T.M. performed the phage competition and bacterial resistance assays. S.H.K. and B.L.H. wrote the initial draft of the manuscript. All authors contributed to writing the final version of the manuscript. **Competing interests:** B.L.H. acknowledges outside interest in Arpelos Biosciences and Genyro as a scientific cofounder. S.H.K. and B.L.H. are named as inventors on provisional patent application 63/883,539 submitted by Stanford University and the Arc Institute that covers methods for generative design of bacteriophage genomes. The other authors declare no competing interests. **Data, code, and materials availability:** The raw and processed *Microviridae* datasets used for SFT are available for download on Zenodo (*91*). The code for generation and fine-tuning with Evo 1 is available at https://github.com/evo-design/evo/ and on Zenodo (*92*). The code for generation and fine-tuning with Evo 2 is available at https://github.com/arcinstitute/evo2 and on Zenodo (*92*). The code for phage genome design and analysis are available at both repositories and on Zenodo (*92*). The fine-tuned *Microviridae* model for Evo 1 is available at https://huggingface.co/evo-design/evo-1-7b-131k-microviridae, and the fine-tuned *Microviridae* model for Evo 2 is available at https://huggingface.co/evo-design/evo-2-7b-8k-microviridae. The annotated generated phage genomes are available on NCBI GenBank with accession numbers PZ509151 to PZ509168. The cryo-EM density maps and resulting models for ΦX174 and Evo-Φ36 have been deposited in the Electron Microscopy Data Bank (EMDB) with accession codes EMD-77390 and EMD-77391 and the Protein Data Bank (PDB) with accession codes 36CQ and 36CR, respectively. All newly created materials are available upon reasonable request to the corresponding authors.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aec2657
Materials and Methods; Supplementary Text; Figs. S1 to S30; Table S1; References (*93*–*140*); MDAR Reproducibility Checklist; Data S1

Submitted 12 September 2025; accepted 2 June 2026

10.1126/science.aec2657

Generative design of bacteriophages with genome language models

Samuel H. King, Claudia L. Driscoll, David B. Li, Daniel Guo, Aditi T. Merchant, Garyk Brixi, Max E. Wilkinson, and Brian L. Hie

Science **393** (6811), eaec2657. DOI: 10.1126/science.aec2657

Editor's summary
The ability to design complex biological systems with artificial intelligence (AI) has the potential to transform biotechnology, but progress has largely been limited to the scale of individual genes and proteins, with whole-genome design remaining out of reach. King *et al*. used generative AI models trained on millions of natural genomes to design entire bacteriophages (see the Perspective by Inglesby and Hanke). Experimental tests yielded 16 functional genomes with diverse sequences, structures, and fitness profiles. A cocktail of the generated bacteriophages rapidly overcame bacteria that had evolved resistance to a natural bacteriophage. This work lays a foundation for AI-guided design of biological function at the whole-genome scale. —Di Jiang