IVYSPRING INTERNATIONAL PUBLISHER

International Journal of Biological Sciences

2027; 23(1): 572-574. doi: 10.7150/ijbs.143024

Commentary

AI-guided bacteriophage genome design: from generation to adaptive evolution

Jongwon Byun[1,2], Hakjin Kim[2,3,✉], Seon-Kyu Kim[3,4,✉], Taeho Kwon[1,2,✉]

1. Quantum AI Bio Research Laboratory (KJQI-JQL), AI-Bio Solution Team, Futuristic Animal Resource and Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju, Chungbuk 28116, Republic of Korea.
2. Advanced Bioconvergence Department, Bioscience, Department of Bioinformatics, KRIBB School, University of Science and Technology (UST), Daejeon 34113, Republic of Korea.
3. Quantum AI Bio Research Laboratory (KJQI-JQL), In Quantio, Gene on Biotech, Daejeon 35229, Republic of Korea.
4. AI-Bio Solution Team, Digital Bio Center, Genomic Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea.

✉ Corresponding authors: Hakjin Kim, Ph.D., e-mail: brofessor@ust.ac.kr (H.K.). Seon-Kyu Kim, Ph.D., e-mail: seonkyu@kribb.re.kr (S. K. K.). Taeho Kwon, Ph.D., e-mail: kwon@kribb.re.kr (T. K.).

© 2027 The Author(s). This is an open access article distributed under the Creative Commons Attribution 4.0 International Licence, subject to the Publisher's Terms of Use at https://ivyspring.com/terms.

Received: 2026.08.24; Accepted: 2026.10.19; Published: 2027.01.01

Related article: Science. 2026 Aug 6;393(6811):eaec2657. doi: 10.1126/science.aec2657.

Introduction

In a recent *Science* study, King *et al*. generated complete ΦX174-like bacteriophage genomes and recovered 16 viable phages after experimental validation (1). The workflow links genome generation to laboratory adaptation while highlighting limitations in functional novelty and biosafety.

Complete genomes require coordinated coding sequences, regulatory elements, and host interactions. King *et al*. selected ΦX174, a well-characterized lytic *Microviridae* phage with a 5.4-kb genome and 11 genes, as a tractable template for testing genome-scale generative design (Fig. 1A).

At the model-specialization stage (Fig. 1A), King *et al*. fine-tuned Evo 1 7B 131K and Evo 2 7B 8K on approximately 15,000 *Microviridae* genomes (1-3). Evo 1 uses StripedHyena and was pretrained on prokaryotic and phage genomes, Evo 2 uses StripedHyena 2 and broader training data spanning all domains of life (2, 3). Their previously demonstrated genome-scale sequence-generation capabilities support their use for this task. Different architectures and training histories can yield different sequence distributions from the same prompt.

Generation was guided by ΦX174-derived prompts (Fig. 1B), followed by filtering for genome quality, gene architecture, predicted host tropism, and diversity (Fig. 1C). This was template-guided sequence design: subsequent DNA synthesis physically realized the proposed sequences for functional testing, rather than implying unrestricted generation independent of natural genomic templates.

During validation (Fig. 1D), 285 of 302 selected genomes were synthesized and assembled, yielding 16 viable phages: 12/227 Evo 1 designs (5.3%) and 4/58 Evo 2 designs (6.9%). The denominators, 227 and 58, represent the successfully assembled designs evaluated for each model after computational selection and DNA synthesis (1). They therefore describe the experimental cohorts rather than different predictions for a single identical input. These unequal samples provide descriptive yields rather than a controlled comparison of model superiority. Nine phages retained their designed sequences, seven acquired mutations, some essential for viability. Experimental validation therefore remained decisive.

Host-range testing revealed inhibition of *E. coli* W by 15 viable phages, whereas all 285 assemblies lacked detectable inhibition of *E. coli* K-12. These results illustrate the limits of sequence-based host-range prediction.

The phenotypic outcomes (Fig. 1E) require separating generated diversity from functional novelty. Computationally selected candidates spanned a broad range of nucleotide identities to ΦX174 (1). However, the 16 viable generated phages showed 93.2–98.8% overall sequence similarity to their nearest natural genomes (1). These comparisons use different reference sequences, but the experimentally

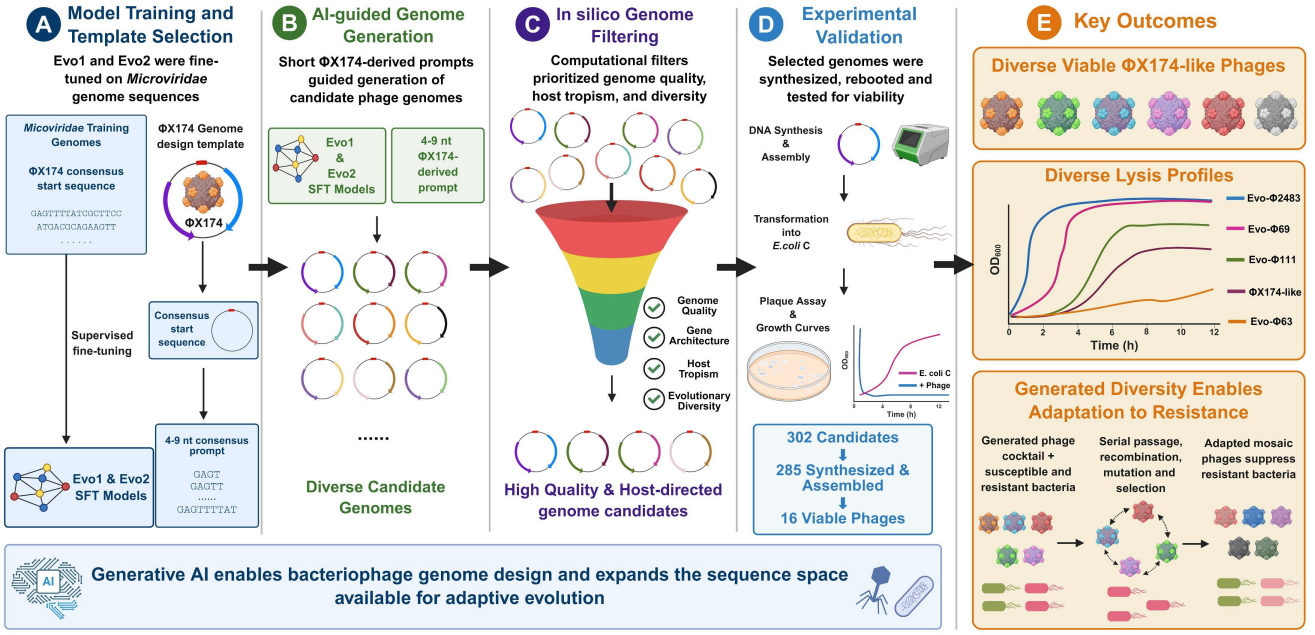

Figure 1. AI-guided generation, experimental validation, and adaptive evolution of ΦX174-like bacteriophages. A Evo 1 and Evo 2 were specialized using approximately 15,000 *Microviridae* genomes, while sequences derived from the ΦX174 template provided short prompts for genome generation. **B** Specialized models autoregressively generated diverse candidate phage genomes from ΦX174-derived prompts. **C** Candidate genomes were computationally filtered according to genome quality, gene architecture, predicted host tropism, and evolutionary diversity. **D** King *et al.* selected 302 candidates for experimental evaluation, successfully synthesized and assembled 285, and recovered 16 viable phages capable of inhibiting *Escherichia coli* C. **E** The viable phages exhibited substantial sequence diversity and distinct lysis and competitive-fitness profiles. The generated phage cocktail was initially ineffective against ΦX174-resistant bacteria. Following serial passage in mixed susceptible and resistant bacterial populations, recombination, mutation, and selection produced mosaic phages capable of suppressing two resistant *E. coli* C strains. The study demonstrates that genome language models can generate complete, functional bacteriophage genomes and expand the sequence diversity available for subsequent adaptive evolution. Created with BioRender.com.

validated phages nevertheless remained closely related to known natural genomes. This limits claims of functional innovation far from established genomic backgrounds. Within the viable set, Evo-Φ36 exhibited context-dependent variation in packaging protein J, and Evo-Φ69 showed strong competitive fitness. These observations describe the tested outcomes rather than establishing a universal sequence-similarity threshold for viability.

The adaptation experiment tested two ΦX174-resistant *E. coli* C strains, designated CR1 and CR2 (1, 4). ΦX174 and a natural-phage cocktail remained ineffective after five passages. The generated cocktail, comprising the 16 generated phages and ΦX174, was initially ineffective but suppressed CR1 after one passage and CR2 after two. The dominant adapted phages were recombinant mosaics with additional mutations. Generated diversity therefore supplied material for laboratory evolution.

Adaptation emerged through recombination, mutation, and selection under the tested conditions, rather than direct model prediction of evolutionary trajectories. These observations suggest candidate mechanisms, identifying causal determinants and their environmental dependence requires dedicated experimental evidence. Model-derived sequence scores may inform hypotheses, whereas causal explanations of adaptation require evidence beyond sequence likelihood.

Generative design may complement rational engineering and directed evolution (5). However, functional validation remained close to the ΦX174 template, and therapeutic efficacy awaits assessment. Larger phages pose additional modelling, synthesis, and assembly challenges. Clinical translation also requires reliable host-range assessment, manufacturing, and evaluation of evolution in vivo. The reported laboratory outcomes therefore represent an early proof of concept.

Biosafety considerations span generation through validation (Fig. 1A–D). The reported safeguards included a well-characterized phage–host system and biosafety precautions appropriate for bacteriophages and nonpathogenic bacterial hosts (1). Evo training excluded viruses infecting eukaryotic hosts (2, 3), and Evo 2 showed poor performance on human viral sequences (3). King *et al*. demonstrated infectivity in a bacteriophage–bacterial host system (1). Extension of this workflow to viruses infecting eukaryotic hosts remains outside the scope of that study. A validated model-internal hazard-warning module remains outside the demonstrated scope of these studies. External risk screening and institutional review should address host range, potential pathogenicity, and environmental consequences, containment should follow the assessed risk rather

than a generic BSL-1 or BSL-2 designation.

These safeguards also clarify the human decisions underlying genome design. Functional phages arose from natural templates, learned sequence patterns, synthesis, and laboratory selection. Researchers, institutions, synthesis providers, and regulators retain responsibility for evaluating uncertainty and governing subsequent use. The significance of generative genomics rests on functional achievement together with accountable oversight.

Acknowledgements

This research was supported by the KRIBB Research Initiative Program (KQM0042611, KQS0092611, KGM1172611, JHM0102611, JHM0092611), NRF funded by the Korean government (MSIT) (RS-2021-NR057659, RS-2026-25478744), National Research Council of Science & Technology (NST) grant funded by the Korean government (MSIT) (No. GTL24023-000) and the Quantum Platform Program (RS-2025-25460035) through the Korea Joint Quantum Institute of KRISS and the National Research Foundation of Korea (NRF) funded by the Korean government (Ministry of Science and ICT (MSIT)).

Author Contributions

J.B. H.K., S.K.K. and T.K. designed, researched, and wrote the manuscript. J.B. H.K., S.K.K. and T.K. participated in the discussion. H.K., S.K.K. and T.K. supervised and reviewed all the research. All authors have read and agreed to the published version of the manuscript.

Competing Interests

The authors have declared that no competing interest exists.

References

1. King SH, Driscoll CL, Li DB, Guo D, Merchant AT, Brixi G, et al. Generative design of bacteriophages with genome language models. Science. 2026;393(6811):eaec2657.
2. Nguyen E, Poli M, Durrant MG, Kang B, Katrekar D, Li DB, et al. Sequence modeling and design from molecular to genome scale with Evo. Science. 2024;386(6723):eado9336.
3. Brixi G, Durrant MG, Ku J, Naghipourfar M, Poli M, Sun G, et al. Genome modelling and design across all domains of life with Evo 2. Nature. 2026;652(8112):1349–61.
4. Strathdee SA, Hatfull GF, Mutalik VK, Schooley RT. Phage therapy: From biological mechanisms to future directions. Cell. 2023;186(1):17–31.
5. James JS, Dai J, Chew WL, Cai Y. The design and engineering of synthetic genomes. Nat Rev Genet. 2025;26(5):298–319.