

ゲノム言語モデル「Evo 1／Evo 2」による機能性バクテリオファージゲノムの生成と実験的検証：技術的機序、治療的応用可能性、およびバイオセキュリティ・ガバナンス

Gemini 3.6 Flash

概要と研究的背景

ライフサイエンスにおける人工知能 (AI) の適用は、AlphaFoldやESMに代表される単一タンパク質の3次元構造予測や局所的なアミノ酸配列の設計を中心として進展してきた¹。しかし、生物学的機能の多くは単一の遺伝子やタンパク質だけで決定されるものではなく、複数の遺伝子、プロモーターやエンハンサーなどの調節エレメント、非翻訳領域、複製起点などが複雑に相互作用する「ゲノム全体」の協調システムによって創発される²。

スタンフォード大学 (Stanford University) およびArc Instituteのブライアン・ハイ (Brian L. Hie) 助教授の研究チームは、ゲノム言語モデル (genome language models) 「Evo 1」および「Evo 2」を活用し、

大腸菌 (*Escherichia coli*) に感染する溶菌性バクテリオファージ Φ X174 を設計テンプレートとして、完全な機能を持つファージゲノムを生成・検証する研究を行った⁴。

計算機的に選別された302種類の候補ゲノムのうち、285候補が化学合成およびDNAアセンブリを経て、非病原性の *E. coli* C を用いた in vitro (試験管内・細菌培養系) 実験において試験された⁵。その結果、16候補から宿主細胞への感染、ゲノム注入、宿主細胞内での複製、および細胞溶菌の一連のウイルスライフサイクルを完遂する増殖可能な機能性バクテリオファージが回収された¹。

本研究は、ゲノム言語モデルによって完全な機能を持つファージゲノムを自律設計・創出できることを実証した概念実証 (Proof of Concept) として極めて重要な成果である⁴。一方で、本研究の直接的な実証範囲は実験室株を用いた in vitro 試験にとどまり、抗菌薬耐性菌 (AMR) 患者に対する臨床的有效性、動物個体やヒトでの安全性、腸内細菌叢への影響、ならびに商用DNA合成スクリーニングの回避性能を直接実験証明したものではない点に留意する必要がある⁴。

ゲノム言語モデル「Evo 1 / Evo 2」の技術的基盤とアーキテクチャ

ゲノム言語モデルの学習データセットと規模

Evoモデル群は、大規模言語モデル (LLM) と同様の自己回帰的 (Autoregressive) 次トークン予測アルゴリズムを、DNAの4つのヌクレオチド塩基 (A, T, G, C) に適用した基礎モデルである¹。モデルは事前学習を通じて、自然選択によって獲得されてきた遺伝子構造、プロモーター配置ルール、コドン使用バイアス、およびゲノム長距離相互作用の統計的パターンを学習する¹。

事前学習モデルの仕様とデータ規模は以下の通りである：

1. **Evo 1:** 原核生物およびファージのゲノムデータ(270万種、約3000億トークン)を含む OpenGenomeデータセットで学習された初期モデルであり、最大コンテキストウィンドウは 131,072塩基対である⁶。
2. **Evo 2:** 細菌・古細菌・真核生物・バクテリオファージを含む全生命ドメインの全ゲノムデータセット (OpenGenome2、合計8.8兆以上のヌクレオチド)に基づき学習された汎用ゲノム基盤モデルである¹¹。モデルサイズとして7B(2.4兆トークンで学習)、20B、および40B(9.3兆トークンで学習)が構築されている¹²。

StripedHyena 2 アーキテクチャと長文脈処理

従来型のTransformerアーキテクチャは、系列長 N に対して計算量およびメモリ消費量が二次関数的 ($O(N^2)$) に増大するため、長大なゲノム配列処理に制約が存在した¹³。Evo 2では、短距離・中距離・長距離の依存関係进行处理するHyena型畳み込み演算子とAttention機構を組み合わせた畳み込み型マルチハイブリッドアーキテクチャ「StripedHyena 2」が採用されている¹¹。StripedHyena 2は、事前学習の後に段階的な長文脈ミッドトレーニングを行う2段階学習を採用することで、単一塩基解像度を保持したまま最大1,000,000トークン(約100万塩基対)のコンテキストウィンドウ処理を可能にしている¹¹。これにより、離れた調節領域と遺伝子コード領域の相互作用やゲノム全体の構造的制約を高精度に捉えることができる²。

モデル名称	パラメータ数	主な訓練データセット	コンテキストウィンドウ	主なアーキテクチャ特性
Evo 1	7B	OpenGenome (原核生物・ファージ270万種 / 約3000億トークン) ⁶	131,072 塩基対 ¹²	ゲノム言語モデルの初期実装、原核生物・ファージ特化 ⁶
Evo 2 (7B)	7B	OpenGenome 2(全ドメイン / 2.4兆トークン) ¹²	最大 1,000,000 塩基対 ¹¹	StripedHyena 2 (Hyena畳み込み + Attention のマルチハイブリッド) ¹¹
Evo 2 (40B)	40B	OpenGenome 2(全ドメイン / 9.3兆トークン) ¹²	最大 1,0,000 塩基対 ¹¹	高精度なDe Novo生成および臨床バリエーション診断に対応 ¹³

ゲノム生成・選別パイプラインと実験的検証プロセス

設計テンプレートと計算機的生成パイプライン

全ゲノム生成の概念実証モデルとして、一本鎖DNAゲノムを持つ Microviridae 科の溶菌性ファージであるバクテリオファージ Φ X174 が選定された⁵。 Φ X174 は約5,386塩基対のゲノムサイズの中に11個の遺伝子、調節エレメント、および重複遺伝子領域 (Overlapping genes) が高密度に存在する複雑な遺伝的アーキテクチャを有している³。

計算機的なゲノム生成および選別プロセスは、以下の段階を経て実施された:

1. 追加微調整 (SFT): 事前学習済みの Evo 1 および Evo 2 モデルに対して、Microviridae 科の 14,466 件のファージゲノム配列を用いて教師あり微調整 (Supervised Fine-Tuning: SFT) を行い、ファージ特有の配列文法を最適化させた⁵。
2. 条件付き自己回帰生成: SFT 済みのモデルに対し、 Φ X174 の保存された短い塩基配列をプロンプトとして与え、条件付き自己回帰生成によって全長ゲノム候補配列を出力させた⁵。
3. 計算機的フィルタリング: 生成された配列群に対し、オープンリーディングフレーム (ORF) の連続性・コード密度の評価、ESMFold によるタンパク質構造予測 (高 pLDDT スコアの確認)、PHROGs および geNomad データベースによる機能注釈の照合、ならびに宿主吸着に関わるスパイクタンパク質の保存性を評価する指向性 (Tropism) フィルターを適用し、302 の候補配列を選別した⁵。
4. **in vitro** 構築および試験: 選別された候補配列のうち、Gibson assembly 等によって正常に組み立てられた 285 候補を *E. coli* C のコンピテントセルへ導入し、96 ウェル形式のプラークアッセイおよび増殖実験 (in vitro 細菌培養系) にて生存可能性を評価した⁵。

段階	プロセス	実施内容および評価基準	定量的成果・通過数
Stage 1	追加微調整 (SFT)	事前学習済み Evo モデルの Microviridae 科配列 (14,466 件) による追加学習 ⁵	ファージ特有の配列文法および再現精度の向上 ⁵
Stage 2	条件付き自己回帰生成	短い Φ X174 由来保存配列をプロンプトとして入力し、全長ゲノム候補を出力 ⁵	多数の全長ゲノム候補配列を生成 ⁵
Stage 3	推論時フィルタリング	コード密度、ESMFold 構造予測、PHROGs 機能適合性、指向性スクリーニング ⁵	302 個の高品質ゲノム候補を選別 ⁵

Stage 4	in vitro 構築・アッセイ	DNA合成・アセンブリ、E. coli Cへの導入および培養系での増殖評価 ⁵	285候補を試験し、16種類の機能性ファージを回収 [cite: 5, 8, 9]
---------	------------------	--	---

実験的検証と構造・ゲノム解析結果

実験の結果、テストされた285候補から16種類の完全な機能を持つファージ (Evo-ファージ) が回収された¹。

ゲノム解析によると、得られた16種類のファージはいずれも最近縁の天然ゲノムと比較して67～392か所の塩基変異を有しており、自然界の単一変異体を越える遺伝的多様性を示した。具体的には、

遺伝子の欠損 (Evo- $\Phi 114$ における遺伝子Kの欠損)、短縮化 (Evo- $\Phi 75$ および Evo- $\Phi 100$ における遺伝子Cの短縮化、Evo- $\Phi 316$ における遺伝子Bの短縮化)、あるいは新規領域の挿入 (Evo- $\Phi 63$ における遺伝子J領域への挿入) など動的なゲノム再編成が確認された⁵。

特に Evo- $\Phi 2147$ は、最近縁の天然ファージ NC51 との平均塩基配列同一性 (ANI) が93.0%であり、一部の分類基準においては新種に相当する可能性が示唆された (ただし、ICTVIによる正式な新種認定が行われたわけではない)。

さらに、クライオ電子顕微鏡 (Cryo-EM) による構造解析 (Evo- $\Phi 36$ [PDB: 36CR]) を実施した結果、カプシド内部に進化的に離れた系統 (*Escherichia* phage G4等) に由来するDNAパッケージングタンパク質が組み込まれ、正常に機能していることが確認された⁵。

生物学的適応度とファージ耐性克服、および将来の治療的展望

生物学的適応度の評価

in vitro 培養系における溶菌動態アッセイ (Lysis Kinetics) および競合増殖実験において、一部の Evo-ファージは天然の野生型 $\Phi X174$ と同等またはそれを上回る溶菌速度および増殖適応度 (Fitness) を示した⁵。これは、ゲノム言語モデルが単に生存可能な配列を模倣するだけでなく、機能的に高適応度なゲノムバリエーションを探索・生成できる可能性を示している⁵。

実験室作製ファージ耐性株に対するカクテル効果

細菌のファージ耐性獲得に対する有効性を検証するため、研究チームは実験室内で天然型 $\Phi X174$ に対する耐性を進化させた3種類の *E. coli* 変異株 (細胞表面受容体に関与する *waa* オペロン等に変異を持つ株) を準備した⁵。

天然型の $\Phi X174$ 単独投与では耐性大腸菌の増殖を抑制できなかったのに対し、多様な塩基配列構造を持つ Evo-ファージを組み合わせたカクテルを投与したところ、1～5回の継代培養を経て耐性が克服され、溶菌作用が確認された⁵。

なお、本研究が直接実証したのは「実験室内で作製された Φ X174 耐性株の克服」であり、臨床現場の多剤耐性菌（AMR）に対する直接的な治療効果を実証したものではない⁵。抗菌薬耐性菌感染症に対するファージ療法（Phage Therapy）への応用は、今後の応用・検証課題として位置付けられる⁴。

宿主域（特異性）の検証

生成された16種類のファージの宿主域を試験した結果、標的株である *E. coli* C および近縁の *E. coli* W に対して増殖能を示した一方、その他の試験した6株（*E. coli* K-12等を含む）においては増殖が確認されなかった⁵。

実験室の限定された菌株パネルの範囲においては宿主域が維持されていることが示されたが、これをもって「複雑な人体内の腸内細菌叢（マイクロバイオーーム）全体に対して影響を与えない」と断定することはできず、生体内での安全性評価には今後の動物実験や臨床試験が必要である⁵。

評価項目	天然型 Φ X174	AI生成ファージ (Evo-Phages)	実験結果と将来の展望
ゲノム変異・多様性	単一の標準野生型配列	最短ゲノムとの比較で67～392か所の変異、遺伝子欠損・挿入を含む多様性 ⁵	多様性を持つファージ群のプログラマブルな設計可能性 ⁴
溶菌速度・増殖力	基準値	一部株で野生型を上回る溶菌速度および競合増殖能を実証 ⁵	高適応度ゲノムの効率的探索手法として期待 ⁵
ファージ耐性克服	耐性株(3株)に対して感染不能 ⁶	カクテル投与により1～5継代で Φ X174耐性を克服 ⁶	将来的な抗菌薬耐性（AMR）対策ファージ療法への応用可能性 ⁴
宿主域（特異性）	<i>E. coli</i> C等に感染	<i>E. coli</i> Cおよび近縁の <i>E. coli</i> Wで増殖、その他6株では不活性 ⁵	実験室パネル内での宿主域維持（腸内細菌叢全体への影響は未検証） ⁵

デュアルユースの懸念とバイオセキュリティのガバナンス

オープンモデルの公開と潜在的リスク

Evo 2の訓練済みモデル（7B、40B）および関連コードは、GitHubリポジトリおよびNVIDIA BioNeMoプラットフォーム上において、商用利用が可能なApache License 2.0の下でオープンソース公開されている¹¹。モデルの公開は計算生物学研究の推進に寄与する一方で、悪意ある主体による病原体改

変や生成への利用というデュアルユース(軍民両用)上のリスクも議論されている¹。

モデル事前学習段階での安全対策とその限界

開発チームは、事前学習データセット(OpenGenome2)から真核生物に感染するウイルスゲノム(ヒト病原性ウイルス等)を除外するデータフィルタリングを実施した¹。これにより、評価された標準条件において、ヒト感染性ウイルスに対するモデリング・変異効果予測・生成性能は著しく低下することが報告されている¹⁸。

ただし、これはデータ除外による性能低下であり、生成不能を恒久的に数学的保証するものではない²。サードパーティによるタスク特異的な追加学習(Post-training / Fine-tuning)が施された場合、これらの安全対策が緩和される可能性について注意が喚起されている²。

DNA合成スクリーニングとガバナンス上の論点

AI生成配列の相同性が自然界の既知データベースと比較して低下した場合、既存の配列相同性(BLAST等)に基づく合成DNAスクリーニングの検知ルールが十分に機能しなくなる懸念が理論的に指摘されている⁴。(※ただし、本研究自体が商用DNA合成事業者のスクリーニング回避実験を実施・証明したわけではない)⁵

これに対し、専門家の間では以下のようなガバナンスおよび対抗技術の議論が進められている：

1. 機能予測型スクリーニング: 配列一致率だけでなく、AIモデルの尤度スコアやタンパク質構造予測を用いて潜在的な毒性・病原性機能を自動検知するシステムの開発⁴。
2. ハードウェア認証の導入: 卓上型DNA合成機(Benchtop DNA Synthesizers)におけるデジタル証明書認証とリアルタイムスクリーニングの義務化⁴。
3. CRISPR等による緊急防衛アプローチ: 未知のAI生成病原体の流出リスクに対し、ガイドRNAを即座に設計して標的ゲノムを切断・無効化する緊急配備型防衛パイプラインの構想⁴。

結論と将来の展望

Evo 1およびEvo 2を用いたバクテリオファージ Φ X174 の全ゲノム生成と in vitro での機能実証は、ゲノム言語モデルが分子レベルから全ゲノムスケールにわたる複雑な遺伝的アーキテクチャを理解・生成できることを示した重要な研究成果である⁴。

将来的には、感染症に対するオーダーメイド型ファージ療法の設計や、有用物質を生産する合成生物学用の人工ゲノム構築などへの応用可能性が期待されている¹。

一方で、臨床応用に向けては、動物モデルやヒトを対象とした安全性・有効性の検証、および腸内細菌叢への影響評価が不可欠である⁴。同時に、AIによるゲノム設計技術の進展に合わせたバイオセキュリティ・ガイドラインおよび合成スクリーニング制度の更新が求められている⁴。

引用文献

1. From Prompts to DNA: AI Generates 16 New Viruses and Kicks Off the Genomic Model Era, <https://www.mocchis.com/en/blog/ai-creates-new-viruses-evo-genomic-models>
2. Generative AI for Biosciences: Emerging Threats and Roadmap to Biosecurity - arXiv, <https://arxiv.org/html/2510.15975v1>
3. Generative design of novel bacteriophages with genome language models -

- bioRxiv, <https://www.biorxiv.org/content/10.1101/2025.09.12.675911v1.full.pdf>
4. Generative design of novel bacteriophages with genome language models - ResearchGate, https://www.researchgate.net/publication/395667081_Generative_design_of_novel_bacteriophages_with_genome_language_models
 5. Generative design of novel bacteriophages with genome language models - bioRxiv, <https://www.biorxiv.org/content/10.1101/2025.09.12.675911v1.full>
 6. Stanford's AI virus designs cross from theory to lab reality - Startup Fortune, <https://startupfortune.com/stanfords-ai-virus-designs-cross-from-theory-to-lab-reality/>
 7. Generative design of novel bacteriophages with genome language models - bioRxiv, <https://www.biorxiv.org/content/10.1101/2025.09.12.675911v1>
 8. Quantifying evolutionary novelty and design efficiency in generative genome design, <https://www.biorxiv.org/content/10.64898/2026.06.12.731871v1.full-text>
 9. Stanford and ARC Institute Scientists Used AI to Design 16 New Viruses That Actually Work, <https://gadgetsnow.indiatimes.com/tech-news/stanford-and-arc-institute-scientists-used-ai-to-design-16-new-viruses-that-actually-work/articleshow/133034711.cms>
 10. CRISPR Treatments for AI-Designed Synthetic Viruses: Rapid Programmable Countermeasures for Emerging and Engineered Viruses - PMC, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12737548/>
 11. スタンフォード大のAIが未知のウイルスゲノムをゼロから生成、16種類が大腸菌に感染 | 財経新聞, <https://www.zaikei.co.jp/article/20260809/865035.html>
 12. Linus/docs/paper-notes/2025.09.12.675911v1.md at main - GitHub, <https://github.com/dbrowneup/Linus/blob/main/docs/paper-notes/2025.09.12.675911v1.md>
 13. 生命の設計図をAIが書く時代～Evo 2プラットフォーム解説レポート | AI生活研究所 - note, https://note.com/ai_stroll/n/n6eef36ba82fe
 14. DNA language model: Evo2の基礎と応用 | olachinkei - note, <https://note.com/olachin/n/n08e15b12afd2>
 15. AI used to design novel bacteriophage genomes in the lab - Xinhua, <https://english.news.cn/20260807/1a6a013baf9b4d7ea8ba844f78599291/c.html>
 16. Ailurus Journal club: AI Designs Viable Genomes: A New Era for Synthetic Biology, <https://ailurus.bio/post/ai-designs-viable-genomes-a-new-era-for-synthetic-biology>
 17. AIで大腸菌に特異的に感染するウイルスが生成されました - crisp_bio, <https://crisp-bio.blog.jp/archives/40671006.html>
 18. Securing Dual-Use Pathogen Data of Concern - arXiv, <https://arxiv.org/html/2602.08061>
 19. ADAPT: a programme for the advanced detection of AI-enabled pathogenic threats, <https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2026.1819372/full>
 20. Protein design, generative AI and biological security - Frontiers,

<https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2026.1817535/full>