

バイオ生成AI研究会

大規模言語モデル(LLM)を基盤とする生成AIの技術的進展は、ライフサイエンス領域へも急速に拡大しています。情報・システム研究機構(ROIS)では、国立遺伝学研究所(NIG)、国立情報学研究所(NII)および全国の大学・研究機関とともに**バイオ生成AI研究会 (GenAI Bio)** を設立し、バイオ生成AIの可能性やゲノム言語モデル構築の課題、さらには実験検証との統合など、新しい研究領域を促進することを目指しています。

バイオ生成AI研究会

前回までのまとめ

■第1回(2025年6月2日)

- 研究会発足の経緯、参加者自己紹介、浜田先生(早大)&白石先生(国がん)発表、文部科学省
- 68名（現地参加30名、オンライン参加38名）

■第2回(2025年6月30日)

- ALIS拠点概要説明、Ledsam氏(Google)発表、開発方向性に関する議論
- 124名（現地参加37名、オンライン参加87名）

■コアメンバー会議(7月9日、16日、23日)

- 開発方向性に関する議論

■第3回(2025年7月30日)

- 開発方向性に関する議論、追川先生(東大)&齋藤先生(北里大)発表
- 112名（現地参加29名、オンライン参加83名）

バイオ生成AI研究会

前回までのまとめ

■第4回(2025年8月20日)

- ・ 開発進捗報告、浅井先生(東大)発表
- ・ 120名（現地参加29名、オンライン参加91名）

■第5回(2025年9月17日)

- ・ 開発進捗報告、廣田先生(科学大)& 築山先生(科学大)発表、 計算機資源に関する議論(小笠原先生)、文部科学省
- ・ 102名（現地参加26名、オンライン参加76名）

■バイオ生成AI研究開発センター新設(2025年11月1日)

■第6回(2025年11月14日)本日

- ・ 開発進捗報告、西先生(東北大)発表

モデル開発・モデル評価進捗状況

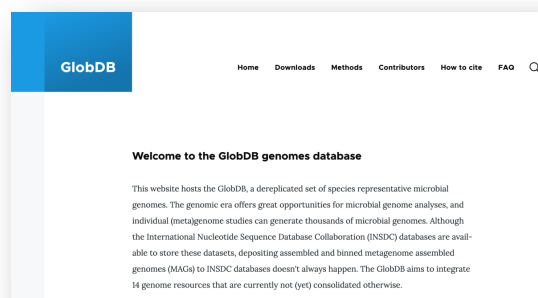
モデリング

モデル性能評価

データセット構築

GlobDBデータセットから高品質MAG
(Metagenome-assembled genomes)
約15万ゲノムを抽出。
クラスタリングして冗長性を排するとともに、
系統が偏らないように順序づけ
EVO1クラス開発向けデータセット構築完了

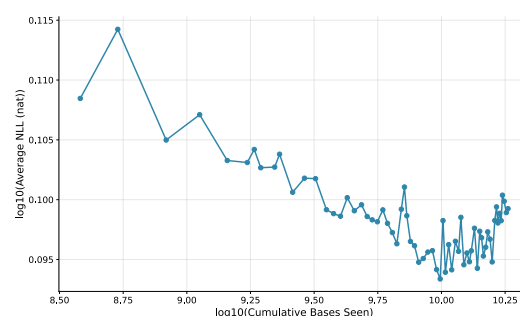
<https://globdb.org>



小規模モデル学習の検証

構築したデータセットを用いて
HyenaDNAモデルの学習を実装。
与えるゲノム配列の数に応じて
テストセット予測精度が上昇することを確認した。

観測トークン（塩基）数 vs. NLL

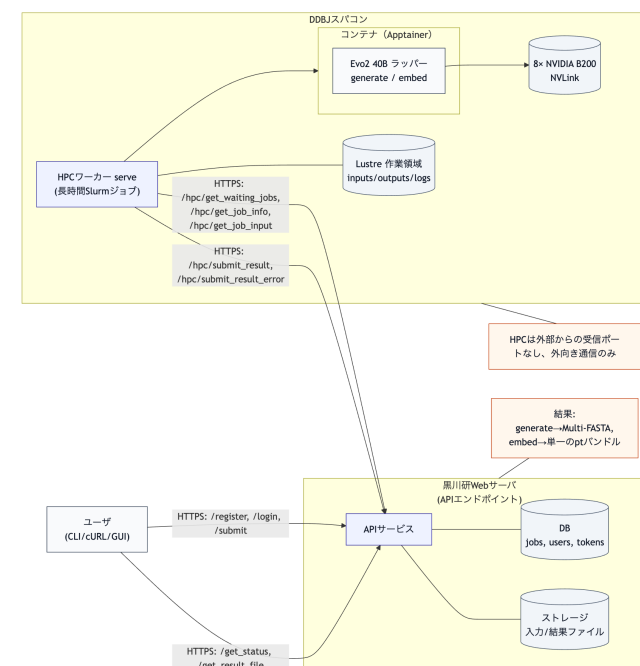


大規模モデル学習の環境構築

Evo1/Evo2レベルコンテキストサイズのモデル
学習のためには、Hyena型モデルに特化した
コンテキストパラレルの分散学習基盤が必要。
そのための方法論の調査と計算環境を構築中。

Evo2モデル推論APIの構築

モデル性能評価の効率を上げるために、
遺伝研スパコンマルチノード環境で
Evo2推論を高速に計算し、ユーザは
Web APIアクセスで配列生成・埋め込み
取得を実行できる環境を構築した。



現在の開発用計算機（遺伝研スパコン）

アクセラレータ最適化ノード

アクセラレータ最適化ノード Type 1 (4台)

NVIDIA B200 GPU を各ノードに 8基搭載した計算ノードです。AI用の計算に適したGPU搭載計算ノードです。

NVIDIA DGX B200



構成要素	型番	員数	ノードあたりの性能など
CPU	Intel Xeon Platinum 8570 (56 cores) Base 2.1GHz, Max 4.0GHz, 1.97TFlops	2	合計 112 コア, 3.94TFlops
Memory	合計2TB		合計 2TB (CPU コアあたり 17.9GB)
GPU	NVIDIA Blackwell B200	8	
Storage (OS)	1.9TB NVMe SSD	2	合計3.8TB
Storage (Data)	3.84TB NVMe SSD	8	合計30.7TB
Network	InfiniBand NDR	1	400Gbps

EVO1 H100 × 128枚 × 1ヶ月

→B200 × 32枚 × 1ヶ月

EVO2 H100 × 2,000枚 × 4ヶ月

→B200 × 32枚 × 10年

モデル開発には大規模計算機資源が必須

- ・ 使いたい時に使える環境
- ・ アジャイル開発可能な環境

基盤モデル開発用：

DGX-GB200(B200 72基)×10台

基盤モデル提供用：

DGX-GB200(B200 72基) ×1台



我が国のライフサイエンスにおける
計算機資源の増強計画が必要

バイオ生成AI研究開発センター (AI-Empowered Life Science Initiative: ALIS)

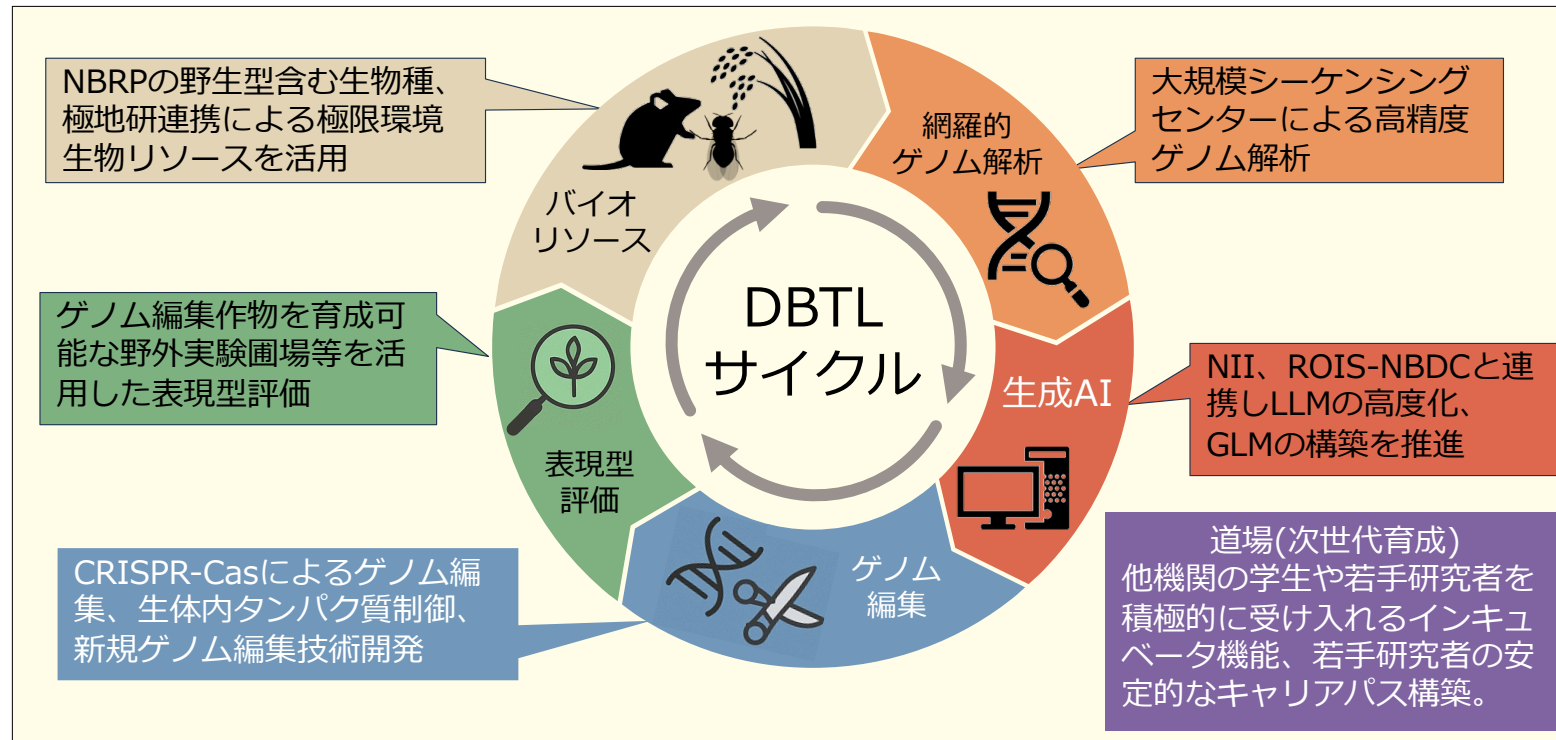
情報・システム研究機構 国立遺伝学研究所は、これまで培ってきた研究基盤をさらに発展させるため、国立情報学研究所およびROIS全研究所が連携して、急速に進展する生成AI技術を生命科学へと導入し、新たな知の創出を目指す研究拠点「バイオ生成AI研究開発センター(ALIS)」を2025年11月1日に開設しました。

ALISセンターをハブとして、全国の研究者とともに、バイオ生成AIの可能性やゲノム言語モデル構築、さらには実験検証との統合など、新しい研究領域を切り拓きます。



<https://ds.rois.ac.jp/center10/>

遺伝研では、gLMによって得られた予測結果を実験的に検証し、その成果をモデルにフィードバックすることで、AIと実験が相互に進化する循環型研究を推進します。



- ・ 「AI × バイオリソース × 表現型解析」による実験プラットフォームを構築し、生命現象や環境応答の制御を可能
- ・ 次世代型の共同研究を全国の研究者とともに推進

ALISセンター開発コアメンバー

氏名	チーム	所属	職位・学年
東 光一	モデリング	国立遺伝学研究所	助教
廣田 佳亮	モデリング	東京科学大学	博士後期課程一年
張 一鳴	モデリング	東京大学	修士課程一年
増田 元希	モデリング	東京科学大学	博士前期課程一年
松本 淳弥	データ	東京科学大学	修士課程一年
豊田 大樹	データ	東京科学大学	学士課程四年
中居 風雅	データ	東京科学大学	学士課程四年
築山 翔	評価	東京科学大学	助教
望月孝子	評価	国立遺伝学研究所	研究員
坂本 美佳	評価	国立遺伝学研究所	研究員
鈴木翔介	評価	筑波大学	博士後期課程一年
毛 涵杰	評価	国立遺伝学研究所	博士後期課程三年

開発に興味のある方、連絡下さい

第6回バイオ生成AI研究会

2025年11月14日

【プログラム】

16:00-16:05 開会挨拶（情報・システム研究機構 喜連川優 機構長）

16:05-16:20 説明（前回までのまとめとお知らせ）（ALISセンター/遺伝研 黒川顕 教授）

16:20-17:10 ゲノム言語モデル開発進捗状況（ALISセンター/遺伝研 東光一 助教）

17:10-17:20 休憩

17:20-17:50 研究発表（東北大学大学院情報科学研究科 西羽美 准教授）「AIのAIによるAIのための国際会議：Agents4Science2025参戦記」

17:50-17:55 事務連絡（事務局）

17:55-18:00 閉会挨拶（ALISセンター/遺伝研 黒川顕 教授）