

Sequence modeling and design from molecular to genome scale with Evo

Nguyen, Eric, et al.

Speaker: Mumtaz Hussain (3902657)

Jan 30, 2025

Introduction

1. DNA: The Blueprint of Life
2. Using Machine Learning with DNA
3. Current Limits of Machine Learning in Biology

Pretraining a genomic foundation model across prokaryotic life

- ▶ A: Long-Context Genome Foundation Model
- ▶ B: StripedHyena Model - Processing Genomic Sequences
- ▶ C and D: Dataset Overview
- ▶ E: Dataset Composition by Kingdom and Phylum
- ▶ F: Model Performance Comparison (Perplexity vs. Model Size)
- ▶ G: Compute-Optimal Scaling

Pretraining a genomic foundation model across prokaryotic life

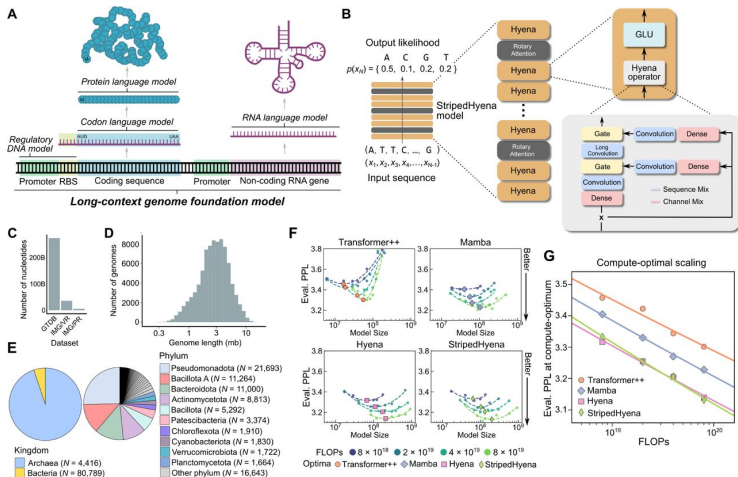


Figure: 1

Evo performs zero-shot function prediction for proteins, non-coding RNAs, and regulatory DNA

- ▶ Panel A: Language Model-Likelihood and Fitness Prediction
- ▶ Panel B: Zero-Shot *E. coli* Protein Fitness Prediction
- ▶ Panel C: Language Model Likelihood for Non-Coding RNA
- ▶ Panel D: Zero-Shot ncRNA Fitness Prediction
- ▶ Panel E: Promoter-RBS Pair Predictions (Gene Regulation)
- ▶ Panel F: Zero-Shot mRNA Expression Prediction
- ▶ Panel G: Zero-Shot Protein Expression Prediction

Evo performs zero-shot function prediction for proteins, non-coding RNAs, and regulatory DNA

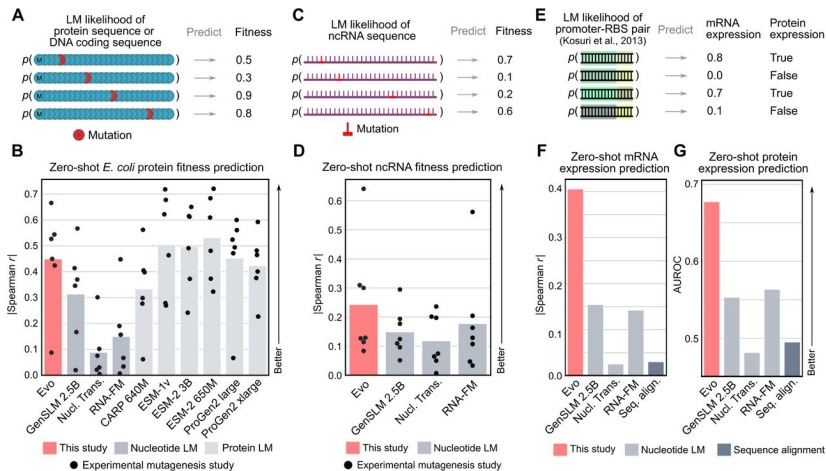


Figure: 2

- ▶ Panel A: CRISPR-Cas System Components
- ▶ Panel B: Training Evo for CRISPR-Cas Design
- ▶ Panel C: CRISPR Subtype Generation Preferences
- ▶ Panel D: Sequence Identity Between Evo-Generated and Training Data
- ▶ Panel E: Structural Comparisons of Natural vs. Evo-Generated Cas9 Proteins
- ▶ Panel F: Evolution of crRNA-tracrRNA Duplex Structures

Finetuning on CRISPR-Cas sequences enables generative design of protein-RNA complexes

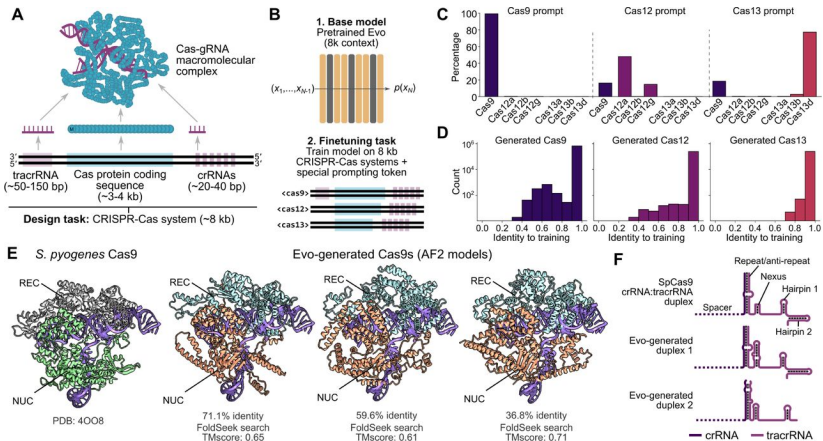


Figure: 3

Finetuning on IS200/IS605 sequences enables generative design of transposable biological systems

- ▶ Panel A: IS605 System Overview and Design Task
- ▶ Panel B: Training Evo for IS605 Generation
- ▶ Panel C: Sequence Identity Distribution of Generated IS605 Components
- ▶ Panel D: Structural Comparisons of Evo-Generated IS605 Proteins
- ▶ Panel E: Sequence Conservation and Entropy Analysis

Finetuning on IS200/IS605 sequences enables generative design of transposable biological systems

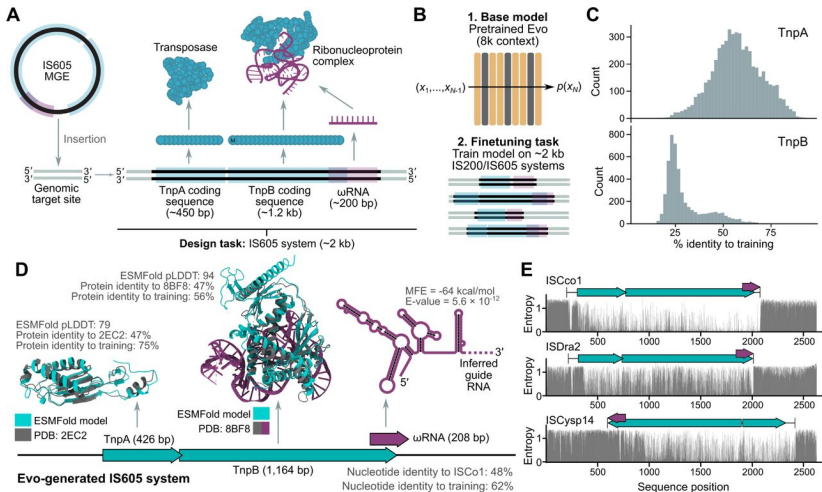


Figure: 4

Evo performs zero-shot gene essentiality prediction across diverse bacterial and phage genomes

- ▶ Panel A: Expanding Context for Gene Essentiality Prediction
- ▶ Panel B: How Evo Predicts Gene Essentiality
- ▶ Panel C: Zero-Shot Gene Essentiality Prediction Performance
- ▶ Panel D: Gene Essentiality Scores Across Different Organisms

Evo performs zero-shot gene essentiality prediction across diverse bacterial and phage genomes

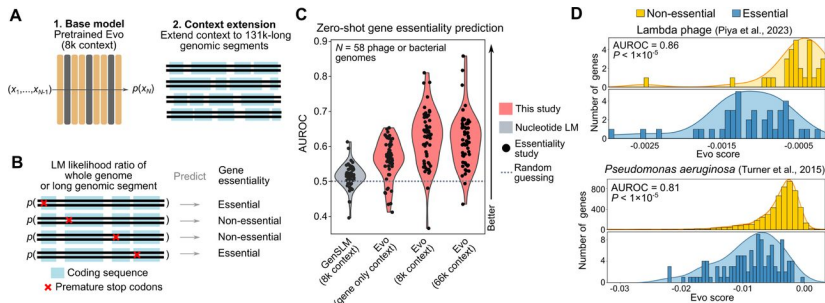


Figure: 5

Evo generates genome-scale sequences with dense coding architecture

1. Panel A: Sequence Generation Task
2. Panel B: Coding Density Comparison
3. Panel C: Coding Sequence Comparison
4. Panel D: Functional Classification of Generated Proteins
5. Panel E: Functional Classification of Generated Genome

Evo generates genome-scale sequences with dense coding architecture

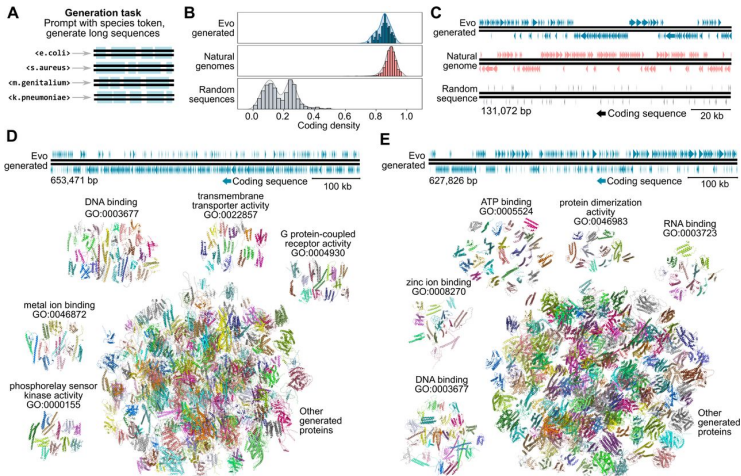


Figure: 6

3. Discussion

1. Evo: A Breakthrough in Genomic Modeling

3. Discussion

1. Evo: A Breakthrough in Genomic Modeling
2. Ethical and Biosafety Considerations

3. Discussion

1. Evo: A Breakthrough in Genomic Modeling
2. Ethical and Biosafety Considerations
3. Technical Challenges and Limitations

3. Discussion

1. Evo: A Breakthrough in Genomic Modeling
2. Ethical and Biosafety Considerations
3. Technical Challenges and Limitations
4. Future Research Directions

Thank You