*Molecular Biology and Evolution*, 2025, **42**, 1–2
https://doi.org/10.1093/molbev/msaf108
Advance access publication 6 June 2025
**Highlights**

MBE OXFORD

# Highlight: MEGA into the New Generation of Computational Genetics

**Pedro Andrade** *

***Corresponding author:** E-mail: pedroamandrade@gmail.com.

Few applications in computational genetics have had such an influence as MEGA—Molecular Evolutionary Genetics Analysis. Still going strong in its 12th version (Kumar et al. 2024), the impact of this legacy toolset is easily captured when one considers that the publications of two of its releases—MEGA4 (Tamura et al. 2007) and MEGA7 (Kumar et al. 2016)—now feature in the top 100 list of the most highly-cited papers across all fields of research (Van Noorden 2025). Reasons for this success include the comprehensive list of tools provided to the user (e.g. sequence alignment, model testing, phylogenetic inference), its regular development, and community acceptance. But above all, one can say that MEGA shines in its accessibility, providing beginners in bioinformatics with a point-and-click graphical user interface that helps smooth the entry into this difficult field.

But can MEGA go one step further? A new Protocol article in Molecular Biology and Evolution by Allard and Kumar (2025) presents a novel tool that promises to overcome remaining barriers in accessibility. As stated by the authors, the growing array of tools available through MEGA has led to a growing complexity in the operation of the software. When faced with so many analytical possibilities, wouldn't you want someone to discuss it with? So, why not just talk to MEGA? This is their proposition with MEGA-GPT.

Artificial intelligence (AI) chatbots, such as the well-known ChatGPT by the company OpenAI, have taken the world by storm in the past 2 years. These large-language models (LLMs) are a type of neural network trained on large text datasets to predict and generate human-like language, based on text input by users. Most of these tools have been trained with large, generalistic text datasets and are quite efficient in providing answers to queries and acting as custom virtual assistants in a variety of fields, including molecular biology and bioinformatics. However, they can also be criticized for their vague and unrefined output.

A solution to these limitations of LLMs is to train models on more specific sets of instructions. To create MEGA-GPT, Allard and Kumar trained OpenAI's GPT model with the MEGA documentation and some of the most relevant articles on the original tool and its methods. Their objective was to create a chatbot model that could guide users through MEGA's growing array of options and models and help devise protocols tailored for different research questions. The authors then tested the model's performance by comparing its answers to the generalist GPT-4o model of ChatGPT in three example use cases: mutation impact assessment, timetree building, and recombination detection.

As a whole, MEGA-GPT performed better than ChatGPT across the different tasks, providing more specific and accurate answers on MEGA's tools. More importantly, MEGA-GPT did not seem to suffer from hallucinations—a well-known issue with LLMs associated with the fabrication of answers (Ji et al. 2023). ChatGPT generated erroneous, fabricated answers in 2 of the 3 tasks, going as far as generating a false protocol for recombination detection, a feature that does not exist yet in MEGA. In all 3 tasks, MEGA-GPT was able to provide more accurate and precise information, pinpointing not only MEGA's options but also its limitations.

Is this just the beginning for tailored chatbot assistants in evolutionary genetics research? Recent decades have seen an explosion in computational pipelines and tools for analyzing DNA sequence data, but one could argue that the learning curve has not exactly become smoother. Most new tools are experimental, lack detailed documentation, and are operated through a command-line interface that, for many, requires extensive training and adjustment. In other professional fields, LLMs are well on their way towards being adopted into workflows to ease multiple tasks (e.g. Lu et al. 2024; Yan et al. 2024), so it should come as no surprise that the same is likely to happen in molecular evolution. Soon, new bioinformatic tools (even those designed for command line usage) may be released accompanied by their own custom LLM assistants. In theory, these LLMs may even assist with technical issues, like solving dependency barriers. MEGA is leading the way with its new GPT-enabled tool, just as it has led the way for large-scale molecular evolutionary analyses since its first introduction back in 1994.

**Want to learn more?** Check out these other articles focusing on accessible computational tools in genetics recently published in *Molecular Biology and Evolution*:

- "Squirrel: Reconstructing semi-directed phylogenetic level-1 networks from four-leaved networks or sequence alignment" (Holtgrefe et al. 2025)

**Received:** May 5, 2025. **Accepted:** May 5, 2025

- "phyloBARCODER: A web tool for phylogenetic classification of eukaryote metabarcodes using custom reference databases" (Inoue et al. 2024)

## Data availability

No data was used in this manuscript.

## References

Allard JB, Kumar S. MEGA-GPT: artificial intelligence guidance and building analytical protocols using MEGA software. *Mol Biol Evol*. 2025:42(6):msaf101. https://doi.org10.1093/molbev/msaf101.

Holtgrefe N, Huber KT, van Iersel L, Jones M, Martin S, Moulton V. Squirrel: reconstructing semi-directed phylogenetic level-1 networks from four-leaved networks or sequence alignments. *Mol Biol Evol*. 2025:42(4):msaf067. https://doi.org/10.1093/molbev/msaf067.

Inoue J, Shinzato C, Hirai J, Itoh S, Minegishi Y, Ito SI, Hyodo S. phyloBARCODER: a web tool for phylogenetic classification of eukaryote metabarcodes using custom reference databases. *Mol Biol Evol*. 2024:41(8):msae111. https://doi.org/10.1093/molbev/msae111.

Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P. Survey of hallucination in natural language generation. *ACM Comput Surv*. 2023:55(12):1–38. https://doi.org/10.1145/3571730.

Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K. MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. *Mol Biol Evol*. 2024:41(12):msae263. https://doi.org/10.1093/molbev/msae263.

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016:33(7):1870–1874. https://doi.org/10.1093/molbev/msw054.

Lu Z, Peng Y, Cohen T, Ghassemi M, Weng C, Tian S. Large language models in biomedicine and health: current research landscape and future directions. *J Am Med Inform Assoc*. 2024:31(9):1801–1811. https://doi.org/10.1093/jamia/ocae202.

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007:24(8):1596–1599. https://doi.org/10.1093/molbev/msm092.

Van Noorden R. These are the most-cited research papers of all time. *Nature*. 2025:640(8059):591. https://doi.org/10.1038/d41586-025-01124-w.

Yan L, Sha L, Zhao L, Li Y, Martinez-Maldonado R, Chen G, Li X, Jin Y, Gašević D. Practical and ethical challenges of large language models in education: a systematic scoping review. *Br J Educ Technol*. 2024:55(1):90–112. https://doi.org/10.1111/bjet.13370.