

CONG LIU, PhD

AI/ML Data Scientist · Computational Biology & Clinical R&D
congliu.bio@gmail.com | [Homepage](#) | [Google Scholar](#)

SUMMARY

AI/ML data scientist with a PhD and 15+ years turning large, heterogeneous biological and biomedical data into models, tools, and decisions. Combines deep statistical foundations: Bayesian inference, study design, multivariate and survival methods with hands-on modern AI engineering, including agentic LLM pipelines, MCP tool integration, and RAG systems. Currently building Bayesian optimization and simulation methods for clinical R&D at a global pharmaceutical company, and has shipped production R packages and Shiny applications now in active use by clinical statisticians. Strong genomics and computational biology background, with peer-reviewed work in Science and other leading journals (34+ papers, 680+ citations).

PROFESSIONAL EXPERIENCE

Data Scientist & AI Scientist

2025–Present

Enterprise AI, GSK (Contract)

- Designed and built a Bayesian optimization framework for clinical study design, developed as a decision-support tool and presented to senior R&D leadership.
- Developed a suite of clinical simulation models: patient enrollment and accrual, disease progression (time-to-event), and statistical power, supporting study design and feasibility assessment.
- Shipped two R packages and R Shiny applications with integrated LLM components, now in active use by clinical oncology statisticians.
- Built agentic LLM workflows using the Model Context Protocol (MCP), exposing simulation and analysis tools through natural-language interfaces for non-technical scientists.
- Developed knowledge-graph and retrieval-augmented generation (RAG) pipelines for large-scale clinical statistics literature mining and question answering.
- Designing a patient digital-twin simulation framework to model individual patient trajectories for trial monitoring and study design.
- Evaluate and responsibly deploy emerging AI systems: LLMs, agent frameworks, automated analysis, including AI safety evaluation within scientific workflows.

Research Associate in Entomology

2023–Present

Museum of Comparative Zoology, Harvard University

- Lead genomics and computational biology research, applying population genomics, phylogenomics, and Bayesian statistical methods to large-scale sequencing data (published in Science).
- Develop high-throughput pipelines (Python, R) integrating genomic, multispectral imaging, and 3D micro-CT data into unified analytical workflows.
- Built and curated GABI-I, a large open-access database integrating heterogeneous published, field, and global-database records.
- Co-mentor graduate students on study design, statistical analysis, and reproducible research practices.

E. O. Wilson Biodiversity Postdoctoral Fellow

2019–2022

Museum of Comparative Zoology, Harvard University; Advisor: Dr. Naomi E. Pierce

- Applied community genomics and Bayesian inference to quantify population-level biological change from historical specimen sequencing.
- Designed and led large-scale studies integrating genomic, environmental, and remote-sensing data; coordinated multi-institutional collaborations across the US, Japan, Fiji, and China.

EDUCATION

Ph.D. in Ecology and Evolution

2012–2018

Okinawa Institute of Science and Technology Graduate University (OIST), Japan — Advisor: Dr. Evan P. Economo

M.Sc. in Ecology **2009–2012**
University of Chinese Academy of Sciences (Xishuangbanna Tropical Botanical Garden), China

B.A. in Life Science **2005–2009**
Anhui Normal University, China

TECHNICAL SKILLS

Statistics & ML: Bayesian inference and optimization, study design, power and simulation analysis, survival/time-to-event modeling, multivariate and spatial statistics, model validation and interpretation

Generative & Agentic AI: Large language models, agentic pipelines, Model Context Protocol (MCP), RAG systems, knowledge graphs, prompt engineering, fine-tuning, text mining, AI safety evaluation

Genomics & Computational Biology: Population and comparative genomics, phylogenomics, high-throughput sequencing pipelines, high-dimensional biological data integration

Engineering & Platforms: Python, R (package development, Shiny), SQL, cloud platforms, API design, reproducible workflows (Git)

SELECTED PUBLICATIONS

Liu C*, Sarnat E, Kubota Y, Pierce NE, Mikhayev AS*, Economo EP* (2025). Genomic signatures indicate massive decline of endemic island insects. **Science**.

Vidal M*, **Liu C**, Wang SP, Segraves KA (2025). Partner dependency alters patterns of coevolutionary selection in mutualisms. **PNAS**.

Full publication list available via Google Scholar.

SELECTED HONORS

Harvard-China Faculty Grant, Harvard University **2022**

E. O. Wilson Biodiversity Postdoctoral Fellowship, Harvard University **2019**

JSPS Research Fellowship for Young Scientists, Japan **2018**