

Chen Hsieh

chen.hsieh.uga@gmail.com | +1-706-308-7928 | chenhsieh.xyz | github.com/chenhsieh | linkedin.com/in/chen-hsieh

SUMMARY

Bioinformatics PhD candidate studying gene duplication and lineage-specific gene evolution in *Populus* through comparative transcriptomics and phylogenetics. Recently working on the robustness of genomic foundation models: conducting parameter-efficient fine-tuning on DNA/RNA language models and using mechanistic interpretability to find where they fail and whether their confidence tracks real biology. Available Fall 2026 (PhD, University of Georgia).

EDUCATION

University of Georgia, PhD Candidate, Bioinformatics	Aug 2020 – Dec 2026 (<i>expected</i>)
National Taiwan University, MSc, Horticultural Crops Science	Jun 2019
National Taiwan University, BS, Agriculture	Jun 2017

TECHNICAL SKILLS

Programming: Python · R · Bash/Shell · JavaScript · C/C++

Deep Learning & Foundation Models: PyTorch · Hugging Face · genomic and RNA language models (Nucleotide Transformer, PlantCAD2, DNABERT2, AIDO.RNA, mRNA-FM, CodonFM) · Low-Rank Adaptation (LoRA) · Parameter-Efficient Fine-Tuning (PEFT) · embedding extraction, pooling, and frozen probing

Interpretability & Robustness: per-layer linear probing · in-silico mutagenesis · per-position importance attribution · decision-depth analysis · black-box adversarial attacks & training · benchmark-validity auditing

Genomics & Classical ML: RNA-seq pipelines (STAR, HISAT2, DESeq2, edgeR) · comparative transcriptomics · phylogenetics · genome-wide annotation · CRISPR gRNA design · XGBoost & SHAP

Engineering & Deployment: Git/GitHub · Docker · CI/CD (GitHub Actions) · Python packaging (PyPI) · REST API · Snakemake · HPC/SLURM · agentic research workflows · Streamlit / R Shiny

RESEARCH EXPERIENCE

PhD Dissertation Research – Tsai Lab <i>Institute of Bioinformatics, University of Georgia</i>	2020 – Present
--	----------------

- Dissertation on lineage-specific genes and the gene-duplication landscape in *Populus*; used expert-annotated and published case studies as a ground-truth reference set to improve automated analysis pipelines.
- Reprocessed large-scale community RNA-seq data (FASTQ to differential expression) on reproducible, HPC-scalable pipelines; built genome-wide protein structure annotation workflows.
- Modeled ohnolog expression divergence with XGBoost and SHAP under leakage-aware validation (chromosome-held-out cross-validation, permuted-label controls).
- Evaluated seven DNA foundation models (Nucleotide Transformer, PlantCAD2, DNABERT2, AgroNT, HyenaDNA, Grover) on *Populus* promoter sequences, diagnosing embedding quality and establishing architecture-specific pooling.
- Contributed to VariantDB v3.2 (CRISPR gRNA verification tool) and coexpression analysis of a perennial-specific sulfate transporter subgroup associated with lignification; both published in *Tree Physiology* (2025).

Adversarial Robustness of Genomic Foundation Models <i>Algoverse AI Research Program (team research project)</i>	2026
--	------

- Built a model-agnostic robustness pipeline (replicate, attack, biological-validity gates, interpret, harden, transfer) across seven classification tasks and six DNA/RNA foundation models (AIDO.RNA-1.6B, Nucleotide Transformer, mRNA-FM, CodonBERT, CodonFM, structRFM).
- Designed constrained black-box adversarial attacks (synonymous, structure-preserving) gated by an independent biological oracle, benchmarked against matched random-edit floors.
- Built a mechanistic interpretability suite (in-silico mutagenesis, activation patching, concept probes, representation geometry); causal patching revealed sparse-but-redundant encoding.
- Delivered LoRA-based defenses that roughly halved attack success. Wrapped the workflow into a modular agentic skill. Targeting a NeurIPS 2026 workshop submission in late August.

MSc Research <i>Dept. of Horticulture, National Taiwan University</i>	2017 – 2019
---	-------------

- Comparative transcriptomics of mycorrhiza-enhanced salt tolerance in rice (*Frontiers in Plant Science*, 2022, co-first author); automated HRM genotyping pipelines for a peach chilling requirement study (*IJMS*, 2020).

SELECTED LEADERSHIP & PROJECTS

Technology Consultant & IT Group Convener

Apr 2022 – Present

Taiwanese Young Researcher Association (Project TYRA)

- Built a platform powering a mentor-mentee matching program that grew from 274 to 700+ annual participants across 4 cohorts and facilitated 1,876 researcher connections; organization later partnered with Fulbright Taiwan and government education agencies.
- Deployed an automated event pipeline (Eventbrite API + scheduling automation) for the weekly seminar series.

Director of Marketing & Communication

Jul – Dec 2025

PhD Consulting Club, University of Georgia

- Tripled club membership across multiple departments; facilitated weekly structured case practice sessions.

Team Leader

2017 – Present

Hackathons & Startup Competitions

- Led cross-functional teams of up to 5 at HudsonAlpha HATCH (2024, 2025) and UGA-Bayer Hackathon (2025); delivered working prototypes under 8 to 28-hour constraints including a plant-based diet planner for astronauts and a corn yield predictor (Snakemake, PyTorch, OpenAI API, Streamlit).
- Built the first Mandarin-language chatbot for plant disease diagnosis (Open Data Innovative Application Contest, Taiwan, 2017); pitched to venture capital panels and secured an NTD\$410,000 award.

PUBLICATIONS

Surber, S. M., **Hsieh, C.**, Na, L., Harding, S. A., & Tsai, C.-J. (2025). An updated sulfate transporter phylogeny uncovers a perennial-specific subgroup associated with lignification. *Tree Physiology*.

Zhou, R., Seth, S. R., Reeves, J., Burns, A. H., **Hsieh, C.**, Horn, T. W., Xue, L.-J., & Tsai, C.-J. (2025). *Populus* VariantDB v3.2 facilitates CRISPR and functional genomics research. *Tree Physiology*.

Tuma, T. T., Nyamdari, B., **Hsieh, C.**, Chen, Y.-H., Harding, S. A., & Tsai, C.-J. (2024). Perturbation of tonoplast sucrose transport alters carbohydrate utilization for seasonal growth and defense metabolism in coppiced poplar. *Tree Physiology*, 44(7), tpae061.

Hsieh, C., Chen, Y.-H., Chang, K.-C., & Yang, S.-Y. (2022). Transcriptome analysis reveals the mechanisms for mycorrhiza-enhanced salt tolerance in rice. *Frontiers in Plant Science*, 13, 1072171. (co-first author)

Chou, L., Huang, S.-J., **Hsieh, C.**, Lu, M.-T., Song, C.-W., & Hsu, F.-C. (2020). A high-resolution melting analysis-based genotyping toolkit for the peach (*Prunus persica*) chilling requirement. *International Journal of Molecular Sciences*, 21(5), 1543.

SELECTED PRESENTATIONS

Talk. Investigating de novo gene birth in *Populus*. SMBE Satellite Meeting on De Novo Gene Birth, Texas A&M University (2023).

Posters. Characterization of stress-responsive lineage-specific genes in *Populus*, IUFR0 Tree Biotechnology Conference (2024) · Searching for orphan genes in *Populus*, ASPB Worldwide Summit (2021) · Comparative transcriptome analysis of gibberellin-induced sex determination in bitter melon, XXX International Horticultural Congress, Istanbul (2018).

SELECTED OPEN-SOURCE TOOLS

Research: **slurm-receipt** (PyPI): CLI that turns SLURM job history into a compute-cost, energy, and cloud-equivalent report. · **sapelo2-boilerplate**: Claude Code rulesets and sbatch templates for efficient GPU/HPC jobs. · **agentic-research-toolkit**: portable agentic research workflows (SKILL.md) written for discovery over confirmation.

Creative coding: **dna2oia** (PyPI, meme-inspired DNA-to-audio sonification), **bioLOLPython** (PyPI, sequence analysis with internet-slang dialects), and a Reigns-style PhD-survival game. · **Biosecurity Atlas**: knowledge graph of the biosafety-and-AI research funding landscape aggregated from five public funding and publication APIs.

TEACHING, MENTORING & OUTREACH

Mentored undergraduate researchers in bioinformatics workflows and phenotyping (UGA, 2021–2025) and in genome-wide sequence analysis (NTU, 2017–2018), where the mentee won first prize at the departmental poster competition. · Guest lecturer, Methods in Horticultural Research (IV), National Taiwan University: taught RNA-seq analysis to an audience with no computational background; pre-lecture content doubled expected attendance. · Elected BIGSA Representative (2022–2023) and organizer of the Institute of Bioinformatics seminar series, hosting Josh Starmer (StatQuest), Robert Edgar (MUSCLE), Ben Langmead (Bowtie/Bowtie2), and Jeffrey Perkel (*Nature Technology* Editor).

AWARDS & CERTIFICATES

Government Fellowship for Studying Abroad (est. \$150K), Ministry of Education, Taiwan (2019) · Summer Research & Communication Grants, UGA (2021, 2022) · Sci4Pol Certificate, National Science Policy Network (2025).