

Anestis Gkanogiannis, Ph.D.

+30 694 283 5799

✉ anestis@gkanogiannis.com

*Senior AI/ML + Genomics Practitioner /
Agentic tools, multi-omics, Bioconductor
maintainer*

linkedin.com/in/anestis-gkanogiannis

0000-0002-6441-0688

github.com/gkanogiannis

Date of birth: 2 January 1981 ■ Nationality: Greek ■ Residence: Greece

Summary and Research Interests

Senior AI/ML and genomics practitioner with ~15 years building open-source, production-grade tools for large-scale biological data. Maintainer of multiple Bioconductor packages (fastreeR, metabinR, jvecfor), and author of agentic, LLM-driven tooling that runs reproducible bioinformatics workflows from natural-language requests.

Broad multi-omics background spanning genome assembly and annotation, population genomics, large-scale NGS and functional-genomics analysis, and metagenomics, backed by reproducible HPC software and end-to-end program leadership. Currently focused on bringing modern AI — embeddings, deep learning, and LLM-based agents — to making complex omics datasets faster and easier to interrogate.

Selected Software & Tools

- jvecfor (R / Bioconductor) — Fast KNN and shared-nearest-neighbor (SNN) graph construction for single-cell RNA-seq. Drop-in replacement for BiocNeighbors::findKNN + bluster::makeSNNGraph. Java backend on the jvector HNSW-DiskANN library, SIMD-accelerated; up to 2× faster at scale than BiocNeighbors/Annoy on the 1.3M-cell TENxBrainData benchmark. <https://bioconductor.org/packages/jvecfor/>
- fastreeR (R / Bioconductor, Bioconda, PyPI, DockerHub/GHCR, Galaxy) — Scalable distance-matrix and phylogenetic-tree computation from VCF/FASTA via a multithreaded Java backend. Recent versions add embedding-based distances driven by genomic language models (BioFM-265M, DNA-BERT, Nucleotide Transformer) and windowed/streaming output. <https://bioconductor.org/packages/fastreeR/>
- metabinR (R / Bioconductor) — Reference-free binning of metagenomic sequences by k-mer composition and abundance, Java-backed. <https://bioconductor.org/packages/metabinR/>
- ClawBio (open-source Claude Code agent for bioinformatics) — Contributor: authored the fastreeR skill enabling the agent to run fastreeR workflows from natural-language requests. <https://github.com/ClawBio/ClawBio/tree/main/skills/fastreer>
- PPD Insight (Mobile app, Android) — AI-powered plant phenotyping app for cassava, developed in collaboration with the CGIAR Cassava Breeding Program. Uses computer vision to estimate cassava tuber post-harvest physiological deterioration from farmer-submitted images. <https://play.google.com/store/apps/details?id=com.gkanogiannis.ppdinsight>

Professional Experience

Independent

Bioinformatics & Genomics Consultant

October 2025 - Present

- Providing research, analysis, and software solutions across population genomics, genome assembly/annotation, machine learning, and NGS pipeline development.
- Supporting research institutions and biotech teams with project design, data analysis workflows, reproducible pipelines, and genomic data interpretation.
- Delivering results-focused, publication-grade outputs and advisory support on computational strategy and infrastructure.

ICBA

Dubai, United Arab Emirates

Bioinformatics Scientist

December 2022 - September 2025

- Led genomics programs at national scale, integrating data analysis, infrastructure, and AI components across multidisciplinary teams.
- Scientific and technical lead for national-scale genomics initiatives in the UAE, including population genomics and genetic diversity analysis of 200+ Ghaf tree individuals.
- Owned end-to-end project delivery for long-read genome assembly and annotation projects (PacBio HiFi, Hi-C, Iso-Seq), from experimental design to publication.
- Defined technical roadmaps, selected methods and tools, and coordinated work across bioinformatics, wet-lab scientists, and external collaborators.
- Mentored and supervised junior researchers and collaborators, reviewed analyses, and resolved technical blockers.
- Accountable for reproducible infrastructure, including Nextflow pipelines, Docker/Singularity containers, and HPC operations.
- Led development and release of fasttreeR, an open-source toolkit for scalable genomic distance computation and SNP-based embedding representations, enabling efficient similarity analysis and downstream machine learning applications (Bioconductor, PyPI, DockerHub).
- Designed and deployed machine learning and deep learning systems for biological data analysis and phenotyping, bridging genomics workflows with production-grade AI applications.
- Ensured projects resulted in peer-reviewed publications, reusable software, and production-ready pipelines.

CRG

Barcelona, Spain

Bioinformatician

April 2022 - November 2022

- Core bioinformatics contributor on cancer-metabolism projects in a mammalian system, designing and implementing reproducible analysis pipelines (Nextflow, Bioconductor) across functional-genomics assays (RNA-Seq, ATAC-Seq, ChIP-Seq, CRISPR screens, SLAM-Seq, Nanopore, and WGS) and co-authoring two resulting papers (Nature Communications; EMBO Molecular Medicine)

Personal goal pursuit

Evia, Greece

Career break

April 2021 - March 2022

- Career break focused on skills refresh, and self-directed training.

Alliance of Bioversity International and CIAT

Cali, Colombia

Head Bioinformatics Scientist

January 2017 - March 2021

- Led a bioinformatics team (3-5 researchers) supporting multiple CGIAR breeding programs (Cassava, Rice, Bean, Forages).
- Defined analysis strategy and priorities, delegated tasks, reviewed outputs, and ensured delivery across concurrent projects.
- Served as primary interface between bioinformatics, breeders, and program leadership, translating research goals into actionable analytics.
- Led development of AI and machine-learning solutions adopted in breeding and phenotyping workflows, including the PPD Insight mobile app.
- Owned and managed HPC infrastructure, data pipelines, and shared computational resources.
- Contributed to high-impact research including GWAS, genomic selection, population genomics, and domestication studies.

- Delivered challenge-winning ML models (Zindi/CGIAR Wheat Challenge 2020, Syngenta Crop Challenge 2018).

CIRAD

Montpellier, France

Post-Doctorate Researcher

June 2015 - December 2016

- Led development of web-based tools for analysis and visualization of Cassava NGS data.
- Acted as technical bridge between data producers and end-users.

Genoscope

Paris, France

Post-Doctorate Researcher

June 2013 - May 2015

- Designed and implemented novel algorithms for metagenome sequence binning.
- Owned method development, benchmarking, and integration into analysis workflows.

University of New Brunswick

Fredericton, Canada

Post-Doctorate Researcher

October 2011 - March 2013

- Managed a research team within the I-AID (Intelligent Analysis of Information and Dissemination) project.
- Coordinated ML software development for NLP-based document classification and information extraction.
- Defined project milestones, distributed development tasks, and oversaw delivery of research outputs.
- Contributed to algorithm design and software implementation for large-scale text mining.

Awards

Zindi/CGIAR Wheat Growth Stage Challenge

2020

First Place Winner (as CGIAR participant)

- Developed computer vision artificial intelligence model to automate wheat growth stage estimation from farmer-submitted images.

Syngenta Crop Challenge in Analytics

2018

First Place Winner

- Participated in developing machine learning models to predict how corn hybrids will perform in untested locations, in an effort to help plant breeders improve hybrid performance predictions.

Discovery Challenge, European Conference on Machine Learning (ECML)

2008

First prize, Spam Detection Task

Greek General Secretariat for Research and Technology

2005 - 2008

PENED Scholarship

Greek Mathematical Society

1995

Regional Award

Education

AUEB

Athens, Greece

Doctorate, Computer Science

December 2005 - May 2011

Thesis: *Information Retrieval and Text Classification. Linear Classifiers and Modified Perceptron.*

Research: Machine Learning, Artificial Intelligence, Information Retrieval, Text Classification

Supervisor: Prof. Emer. Theodore Z. Kalamboukis

AUEB

Master's, Computer Science

Thesis: *Text Classification with k-NN.*

Research: Machine Learning, Artificial Intelligence, Information Retrieval, Text Classification

Supervisor: Prof. Emer. Theodore Z. Kalamboukis

Athens, Greece

October 2003 - June 2005

University of Crete

Bachelor's, Physics

Research: Lasers specialization (emphasis on laser-superconductor interaction)

Extra: Computer programming and algorithms

Heraklion, Greece

October 1998 - June 2003

Teaching & Mentorship

Plant and Animal Genome

Mentorship Program

- Mentor to 3 early-career researchers in genomics and bioinformatics.

San Diego, USA

Jan 2025

AUEB

Teaching Assistant

- Supervised and mentored undergraduate and graduate students in programming, statistics, and information retrieval
- Designed and led hands-on laboratory sessions and project-based coursework

Athens, Greece

Sep 2005 - Jul 2010

Skills & Core Strengths

AI & Machine Learning (Applied)

- LLM-based agents and agentic tooling for bioinformatics (Claude Code skills/agents that drive reproducible workflows from natural-language requests)
- Single-cell graph construction and approximate-nearest-neighbor methods for scRNA-seq (KNN/SNN graphs, HNSW-DiskANN); genomic-language-model embeddings for variant representation.
- Genomic data representations and embedding strategies for large-scale similarity, clustering, and ML applications.
- Machine learning and deep learning for genomics and phenotyping
- Natural Language Processing and text classification (PhD-level expertise), computer vision (YOLO), and predictive modeling
- Frameworks: TensorFlow, PyTorch, scikit-learn, HuggingFace
- Proven record of production and challenge-winning models

Technical Leadership & Delivery

- Technical leadership of genomics and AI projects from concept to publication and deployment
- Project ownership, roadmap definition, prioritization, and delivery under real constraints
- Mentoring and technical supervision of junior researchers and cross-functional collaborators
- Translating scientific goals into scalable, reproducible computational solutions

Bioinformatics & Multi-omics

- Genome assembly and annotation (PacBio HiFi, Hi-C, Iso-Seq, ONT)

- Population genomics, genetic diversity, GWAS, LD analysis, PCA, ADMIXTURE
- Functional and multi-omics analysis (RNA-Seq, ATAC-Seq, ChIP-Seq, Iso-Seq, SLAM-Seq, CRISPR screens)
- Variant calling, QC, and large-scale NGS data processing
- Metagenomics and comparative genomics

Software Engineering & Infrastructure

- Reproducible pipelines: Nextflow, Docker, Singularity
- Linux, HPC (Slurm, SGE), cloud environments (AWS, GCP, Azure)
- Bioconductor package author and maintainer (fasttreeR, metabinR, jvecfor); open-source development and long-term maintenance
- Version control, CI/CD, collaborative development practices

Programming

- Python, R (Bioconductor), Bash
- Java, C/C++ (performance-critical tools, long-term codebases)
- Scientific software prototyping and production deployment

Workshops

- ELIXIR BioHackathon Europe: BUSCO genes for phylogenomics (2025)
- EMBL-EBI: Plant genomes: from data to discovery (2024)
- ELIXIR BioHackathon Europe: Genome annotation workflows (2023)
- EiB: AI & ML for Genomic Selection (2019)
- Cornell / BTI / IRRI / NIAB (2017-2018)

Languages

English: Professional / Working proficiency. Working experience since 2011

French: Intermediate level. Working experience since 2013

Spanish: Intermediate level. Working experience since 2016

Greek: Mother tongue

Interests

Building practical IoT / ML prototypes (Raspberry Pi, sensors, applied AI)

Scuba diving, road tripping, and expedition planning