Omid Ardalani

Computational Biologist | Genomes at Scale, Machine Learning on Genomes, Agentic AI Systems for Biology

Kgs. Lyngby, Denmark · +45 20 20 08 67 · Omidard@biosustain.dtu.dk
omidard.github.io · github.com/omidard · ORCID 0000-0002-0064-3802

PROFILE

I am a computational biologist who works at the scale where the real patterns in biology become visible, and who builds the infrastructure that puts that scale within reach of other scientists. Over a decade I have reconstructed whole-organism metabolic models across entire bacterial families, written the pipelines behind them, and released them as tools that other researchers now use. My work sits between biology and machine learning: I train models on genomes and protein sequences, and I build agentic systems that let language models carry out real computational biology, together with the evaluation harnesses needed to know when they are wrong. I came into computation from the laboratory bench, so I know what a scientific workflow looks like in practice and where it breaks.

SELECTED ACHIEVEMENTS

- **Scale.** Built the largest strain-resolved genome-scale metabolic model collections assembled for a bacterial species and for a bacterial family: roughly 12,000 *E. coli* and 2,300 Lactobacillaceae in-silico twins, and the open pipelines that make that scale reachable.
- **A verifiable signal in biology.** Every one of those models emits falsifiable predictions: this strain grows on this medium or it does not, this gene is essential or it is not. Each is scoreable against experimental measurement, which makes it one of the few parts of biology with a clean problem-and-answer pair, and therefore usable as a training and evaluation signal.
- **Tools biologists actually use.** First author on the pangenome-scale model-generation pipeline behind **BiGG 2026** (biggr.org), the next release of the BiGG Models database, the field's reference model repository. Also built the **Flux Analysis Studio**, a constraint-based modeling workbench that runs with no server: the linear program is solved in the browser tab by a WebAssembly build of GLPK. Eight analyses (FBA and pFBA with editable media and knockouts, dynamic FBA, flux variability, production envelopes, phenotype phase planes, essentiality screens, multi-model analytics, and cohort comparison with Mann-Whitney and Benjamini-Hochberg FDR), each with a live Escher flux map, and every result reproduces COBRApy and scipy to full precision.
- **Agentic AI for biology.** Designed and operate a multi-agent research system on Claude Code that does real life-science work (bioinformatics pipelines, database mining, figure generation, literature synthesis), with grounded verifiers that check each claim against the underlying data instead of asking the model whether it was right.
- **The genetic basis of metabolic function.** First and corresponding author, *Science Advances* (2026). Now scaled into a pipeline that searches millions of genes across thousands of genomes, tests whether each gene is genuinely functional rather than merely present, and maps the full space of distinct genetic codes that encode one and the same biochemical function.
- **From bench to product.** Isolated seven lactic acid bacteria strains and led their scale-up to commercial production at a biotech startup. Lactobacillaceae-related patent. Main inventor; Notice of Invention filed, DTU. Trained in the group of **Prof. Bernhard O. Palsson**.

KEY SKILLS

- **Machine learning for biology:** protein language models and embeddings; ML prediction of enzyme kinetics (kcat, Km); unsupervised learning and non-negative matrix factorization (NMF) of pangenomes; PyTorch, TensorFlow, scikit-learn; model training, validation, and benchmarking against reference data
- **Agentic AI systems and LLM tooling:** heavy Claude Code user; designed and run a multi-agent research system for biology (tool scaffolding, subagent orchestration, MCP servers, hooks, reusable skills); evaluation harnesses and grounded verifiers; eval-driven development; context and prompt engineering
- **Software engineering and infrastructure:** production Python (advanced), R (advanced), Bash, Linux; large codebases; open-source release and maintenance (github.com/omidard); Git; Snakemake and Nextflow; HPC (SLURM) and cloud (Azure); large-scale distributed batch computing; APIs, databases, and public web deployment
- **Genomics and bioinformatics:** pangenome reconstruction across whole bacterial families; sequence search and clustering (BLAST, DIAMOND, HMMER, MMseqs2, CD-HIT); annotation (Bakta, Prokka, eggNOG,

InterProScan, Pfam); gene-family and orthology analysis; taxonomy and ANI (GTDB-Tk); horizontal gene transfer; AMRFinderPlus; Biopython; protein structure at scale (Foldseek, US-align, SWISS-MODEL)

- **Biological databases and their failure modes:** UniProt and Swiss-Prot, NCBI, GTDB, KEGG, BiGG, MetaNetX, AlphaFold DB; schema-level work, and a documented record of finding the silent errors these resources propagate (duplicate assemblies, contradictory taxonomy, homology-derived annotations that are chemically impossible)
- **Metabolic modeling:** genome-scale and pangenome models (GEM, panGEM); FBA and FVA; COBRApy, CarveMe, MEMOTE; computational strain design (OptKnock, FSEOF); in-silico media formulation
- **Laboratory and translation:** microbiology, molecular biology, strain isolation and development, fermentation and bioprocess, QA/QC, GMP/GLP, technology transfer; communicating computational results to experimental scientists and to business

PROFESSIONAL EXPERIENCE

Postdoctoral Researcher, DTU Biosustain / DTU BRIGHT (Technical University of Denmark)

Kgs. Lyngby, DK · 2025-Present

- Lead the pangenome-scale program that searches **millions of genes across thousands of bacterial genomes**, tests whether each gene is genuinely functional rather than merely annotated, and maps the complete space of genetic solutions encoding the same metabolic function.
- First author on the pipeline behind **BiGG 2026, the next release of the BiGG Models database** (biggr.org), the field's reference repository for genome-scale models. Own the production infrastructure behind it: Python pipelines and large codebases, Nextflow and Snakemake workflows, HPC and Azure batch compute, open-source releases, and the public web apps over the model collections.
- Built the **verification layer the group now reuses**, after establishing that standard annotation practice produces systematic, silent errors: duplicate assemblies inflating genome counts (16,404 records reduced to 8,826 truly unique), taxonomy that disagrees between NCBI and GTDB, gene-symbol-to-function calls that are artifacts, and homology-inherited gene-to-reaction assignments that are chemically impossible. Each was found by grounding the claim in data rather than trusting the database.
- Apply **protein language-model embeddings and ML kcat/Km predictors** to prioritize enzymes, and **non-negative matrix factorization** to recover genome structure that no other method surfaced.

Independent Work, Agentic AI Systems for Biology

2025-Present

- Designed and operate a **multi-agent research system built on Claude Code** that carries out real computational biology: bioinformatics pipelines, public-database mining, analysis notebooks, figure generation, and literature synthesis, organized as specialized agents with explicit contracts, handoffs, and persistent project state.
- Built the **evaluation layer** that makes its output trustworthy: grounded verifiers that measure the artifact itself (reading a rendered figure back and measuring font sizes, overlaps, clipping, and blank panels, rather than asking the model whether the figure is good), reviewer agents held separate from the author agent and given isolated context, and checks requiring every claim to resolve to source data.
- **Core principle: a model does not know when it is wrong, so the system must verify against the world, not against the model.** Every subsystem carries an eval harness whose pass or fail criterion is computed from data.
- Engineering delivered: tool scaffolding and integrations, MCP servers, hook-driven routing, subagent orchestration, reusable skills, and reproducible workflow graphs.

Researcher, Computational Biology, DTU

2024-2025

- Developed marker-gene and rare-gene methods, supervised and unsupervised, that classify strains and identify niche-adaptive metabolism across pangenomes.

Ph.D. Researcher, Systems Biology, DTU (group of Prof. Bernhard O. Palsson)

2021-2024

- Built the largest strain-resolved metabolic model collections assembled for a bacterial species and family, roughly **12,000 *E. coli* and 2,300 Lactobacillaceae genome-scale models**, and wrote the reconstruction pipelines that made that scale possible. Released as open-source software (github.com/omidard/EcopanGEM) and as public model browsers (omidard.github.io/panGEMs).
- Predicted strain-specific growth and metabolite production, including short-chain fatty acids, and formulated growth media in silico.

- As **first and corresponding author** on a multi-institution study, set the research direction, coordinated the co-authors, and carried the work through peer review. **Led the Lactobacillaceae pangenome study (*mSystems*)** as lead scientist.

R&D Specialist, Strain Development, Fermentation and QA/QC, Behdad Genetics Corp. (biotech startup)

Tehran, IR · 2019-2021

- **Isolated seven lactic acid bacteria strains and led their scale-up from laboratory to commercial production** as feed probiotics, running a full design-build-test-learn and technology-transfer cycle.
- Designed and optimized growth media and fermentation processes; established QA/QC and authored SOPs to GMP and GLP standards. Hands-on molecular biology, microbiology, and bioprocess work at the bench and at production scale, which is where I learned what a real scientific workflow looks like and where it breaks.

SELECTED OPEN-SOURCE TOOLS

- **BiGG 2026** (biggr.org) · **Flux Analysis Studio** (omidard.github.io/EcopanGEM/analysis.html), the in-browser modeling workbench above · **panGEMs**, a public browser over 2,313 *E. coli* and 2,346 Lactobacillaceae models (omidard.github.io/panGEMs) · **EcopanGEM** · **LactoPanGEM** · **GrowthDB**, the measured growth and uptake rates model predictions get scored against · **Media**, citation-backed growth media. All at github.com/omidard, overview at **omidard.github.io**

SELECTED PUBLICATIONS AND INTELLECTUAL PROPERTY

- **Annotating the pangenome reveals the diversity in the genetic basis for metabolic enzymes.** *Science Advances* 12(27) (2026). doi:10.1126/sciadv.aeb3363. First and corresponding author.
- **Rare metabolic gene essentiality is a determinant of microniche adaptation in *Escherichia coli*.** *PLOS Pathogens* 21(12) (2025). doi:10.1371/journal.ppat.1013775. First author. Open code: github.com/omidard/EcopanGEM
- **Decomposition of the pangenome matrix reveals a structure in gene distribution in the *Escherichia coli* species.** *mSphere* 10(1) (2025). doi:10.1128/msphere.00532-24. Second author.
- **Pangenome reconstruction of Lactobacillaceae metabolism predicts species-specific metabolic traits.** *mSystems* 9(7):e00156-24 (2024). doi:10.1128/msystems.00156-24. First author.
- **Reconstruction and validation of a genome-scale metabolic model of *L. lactis* NCDO 2118; in-silico succinate and GABA overproduction.** *Biochemical Engineering Journal* 170:107967 (2021). doi:10.1016/j.bej.2021.107967. First author.
- **BiGG 2026: a pangenome-scale metabolic-model generation pipeline and the next-generation release of the BiGG Models database.** *In preparation.* First author and pipeline architect.
- **The *Escherichia coli* pangenome is organized by co-occurring gene sets representing horizontal and vertical inheritance.** *Under review, Science Advances.*
- **Lactobacillaceae-related patent. Main inventor; Notice of Invention filed, DTU.**
- Full list of peer-reviewed papers, preprints, and PhD thesis: ORCID 0000-0002-0064-3802

EDUCATION

- **Ph.D., Systems Biology, Technical University of Denmark (DTU)** (2021-2024). Thesis: *Pangenome-scale metabolic reconstruction of the Lactobacillaceae family and Escherichia coli.* Supervisor: Prof. Bernhard O. Palsson.
- **M.Sc., Microbial Biotechnology, University of Tehran** (GPA 18.14 / 20, 2015-2018). Thesis: genome-scale model of *Lactococcus lactis* NCDO 2118; model-guided media optimization to increase GABA production.
- **B.Sc., Cellular and Molecular Biology / Biotechnology, Islamic Azad University, Arak** (GPA 17.53 / 20, 2010-2014). Trained across general biology, molecular biology, and biotechnology before moving into computational microbiology and systems biology, so the computational work rests on the underlying biology rather than on data alone.

LANGUAGES, HONORS AND REFERENCES

English (IELTS Academic 7.5) | Danish (basic, in progress) | Farsi (native) | Kurdish (native). Ranked 3rd nationally in the Iranian national entrance examination. Ranked 1st among M.Sc. students in Microbial Biotechnology, University of Tehran (2018).

References available on request: Prof. Bernhard O. Palsson (UCSD); Prof. Lars K. Nielsen (University of Queensland); Dr. Patrick V. Phaneuf (DTU).