Jacob D Krol

jacob.krol@cuanschutz.edu
github.com/jakekrol

linkedin.com/in/jacob-krol-b3b784156

EDUCATION

Non-degree seeking, University of Colorado Anschutz Medical Campus (CU Anschutz), Aurora, CO　　**Sep 2023 — Dec 2023**
- BIOS 7747: Machine Learning for Biomedical Applications, graduate course offered by the Colorado School of Public Health

Bachelor of Science, Computational Neuroscience, Michigan State University (MSU), East Lansing, MI　　**May 2020 — Aug 2022**
- GPA: 3.89/4.00, Semester awards: Dean's List
- Graduated 'With Honor'

Math and Science Transfer Program, Washtenaw Community College, Ann Arbor, MI　　**Sep 2017 — Apr 2020**
- GPA: 3.52/4.00, Semester awards: Honor Roll
- Transferred

SKILLS AND ACADEMIC INTERESTS

Computation	Python, R, Bash/shell, Java, Perl, High-performance computing (HPC), Linux, Vi/Vim/NVim, Git/GitHub, Containerization (Docker, Sigularity, etc.), Web development, C/C++, LaTeX
Interests	Computational Biology, Bioinformatics, Genomics, Software development, Statistics, Machine Learning, High Performance Computing, Electrophysiology

RESEARCH EXPERIENCE

Information Sciences Professional (PRA), full-time　　**Nov 2022 — present**
JRavi Lab, Department of Biomedical Informatics, Center for Health Artificial Intelligence
University of Colorado School of Anschutz School of Medicine, Aurora, CO

- Redesigned the front and backend of MolEvolvR, a web app for protein characterization., Krol, et al., 2023; bioRxiv, DOI: doi.org/10.1101/2022.02.18.461833. Web app: jravilab.org/molevolvr
- Developed a machine learning (ML) pipeline for classifying antimicrobial resistance (AMR) in bacterial strains (publication in preparation for early 2024 submission).
- Listed in the acknowledgement section of the publication: The Phage-shock-protein (PSP) Envelope Stress Response: Discovery of Novel Partners and Evolutionary History. DOI: doi.org/10.1101/2020.09.24.301986 (provisionally accepted at mSystems.)
- Mentored Ph.D. students and undergraduate to assist in omics data featurization, model development, and model outcome analysis.
- Assisted in hiring process (interviewing and feedback) for open lab positions: post-doctoral and research assistants.
- Hosted deparment wide workshops on Bash, Git, R package development, and ssh-workflow basics.
- Assisted in writing and generating figures for lab grant proposals and publications.
- Performed system administrator duties (e.g., server onboarding, dependency/user/data/resource management) for our lab's web server.

Student Research Assistant, part-time　　**May 2022 — Aug 2022**
Krishnan Lab & Malmstrom Lab, Department of Computational Mathematics Science and Engineering (CMSE)
Michigan State University, East Lansing, MI

- Led a collaborative project between the Krishnan and Malmstrom labs on developing a machine learning model to predict plant virus hosts' using omics data.
- Learned, and implemented statistical methods on viral protein datasets: fisher test, logistic regression, & principal component analysis.
- Presented the approach and results in project meetings.

PUBLICATIONS

1. **Krol**, J. D. *et al.* MolEvolvR: A web-app for characterizing proteins using molecular evolution and phylogeny. *bioRxiv. DOI: doi.org/10.1101/2022.02.18.461833* (2023).
2. **Krol**, J. D. *et al.* Accurately predicting AR in ESKAPE pathogens using machine learning. *In preparation.* (Planned for early 2024 submission).
3. **Krol**, J. D. *et al.* Source code for developing antimicrobial resistance ML models. *In preparation.* (Planned for early 2024 submission to Bioconductor).
4. **Krol**, J. D. *et al.* Summarized data for the AMR project including feature matrices, results tables, and metadata. *In preparation.* (Planned for early 2024 submission to ExperimentHub).

Awards and Funding

- Submitted a graduate research proposal titled "*Prediction of phenotypes from multiscale omics data via graph machine learning methods.*" to the Fall 2023 NSF Graduate Research Fellowship Program (GRFP), *In review.*
- Contributed to minor revisions NIH NIAID U01 grant submission, *Awarded; PI: Janani Ravi.*
- Contributed revision and figures to NIH NIAID R01 grant submission, *In-review; PI: Janani Ravi.*
- Submitting Winter 2023 Department of Energy Computational Science Graduate Fellowship, *In preparation.*

Presentations

Research and technical talks
- **Jul 2023.** R/Bioconductor: Cancer and Evolution track. *MolevolvR a web-app for protein characterization.* [Web app: jravilab.org/molevolvr]. Boston University, Boston, MA.
- **May 2023.** Great Lakes Bioinformatics Conference. *MolEvolvR a web-app for protein characterization [Web app: jravilab.org/molevolvr].* McGill University, Montreal, CA.
- **May 2023**. Great Lakes Bioinformatics Conference. *How and when to build a web-app or R 2023 package?* McGill Universtiy, Montreal, CA. (Hosted a 4 hour in-person workshop on how to build an R package using automation: devtools and usethis. A github repo for a sample R package I wrote is located at http://www.github.com/jravilab/iprscanr).

Posters
- **Sep 2023.** CU Department of Biomedical Informatics Annual retreat. *Robust machine learning-based classification of antimicrobial resistance in high-impact pathogens.* CU Anschutz, Aurora, CO.

Research Projects

MolEvolvR: A web-app for characterizing proteins using molecular evolution and phylogeny   **Nov 2022 — present**
- Maintaining the 4 year old code base by resolving over 50 GitHub issues, applying code linters, simplifying complex code, adding extensive documentation, and UI updates.
- Performed bench-marking and tuning of parameters for homology search and annotation tooling in the application's sequence characterization pipeline to address previous issues in computational complexity.
- Working with Faisal Alquaddoomi on containerizing the front- end, back-end, and slurm instance for the MolEvolvR web-application.

Accurately predicting AR in ESKAPE pathogens using machine learning   **Nov 2022 — present**
- Implemented various ML approaches and evaluation techniques: logistic regression, gradient boosting machines, random forests, linear discriminant analysis, stratified-cross-validation, class-weighting, hyperparamter searching, and evaluating auROC, balanced accuracy, etc., on hold-out dataset.
- Used the CU Anschutz Alpine high performance computing cluster (HPC) to aggregate, transform, and train ML models on over 100GB of omics data.
- Developing two R packages for the project: submitting source code to Bioconductor and data to ExperimentHub (see publications and software).
- Containerizing AMR data collection code and the two R packages under development.

Prediction of plant virus hosts using omics data and machine learning   **May 2022 — Aug 2022**
- Trained machine learning classifiers to predict plant virus' host types; also, trained models to predict plant virus taxonomy.
- Featurization of protein sequences and data wrangling with Pandas, Biopython, NumPy, and R (Tidyverse + Bioconductor) packages for biological feature extraction.
- Further preparing data with one-hot-encoding and z-score normalization.
- Analyzed and visualized model performance with Matplotlib/Seaborn & Scikit-learn performance metrics.

Peer Mentees

PhD Students
- Keenan Manpearl; Computational Bioscience program, CU Anschutz   **Spring 2023**
- Jill Bilodeaux; Microbiology program, CU Anschutz   **Spring 2023**
- Charmie Vang; Biomedical Sciences program, CU Anschutz   **Nov 2023 — present**

Undergraduates
- Ethan Wolfe; B.S. in Biochemistry & Molecular Biology with CMSE and additional minors, MSU   **Nov 2022 — present**

References

Janani Ravi, Ph.D. ; janani.ravi@cuanschutz.edu
— Assistant Professor and Principal Investigator of the JRavi Lab, Department of Biomedical Informatics, CU Anschutz
Faisal Alquaddoomi, Ph.D. ; faisal.alquaddoomi@cuanschutz.edu
— IT Principal Professional, Department of Biomedical Informatics, CU Anschutz
Arjun Krishnan, Ph.D. ; arjun.krishnan@cuanschutz.edu
— Associate Professor and Principal Investigator of the Krishnan Lab, Department of Biomedical Informatics, CU Anschutz