

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, Neuroscience**, Michigan State University (MSU), East Lansing, MI **Sep 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

SKILLS AND ACADEMIC INTERESTS

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

RESEARCH EXPERIENCE

- Information Sciences Professional** **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
- Developed an R package (github.com/jravidlab/amr) to apply machine learning, hypothesis tests, and rank-based methods on publicly available bacterial genome data in order to classify antimicrobial resistance (AMR) bacteria and identify gene/protein-domain AMR contributors (*manuscript in preparation*)
 - Full stack development for *MolEvolvR: a web app for characterizing proteins using molecular evolution and phylogeny.*, Krol, et al., 2023; bioRxiv, DOI: doi.org/10.1101/2022.02.18.461833 (jravidlab.org/molevolvr)
 - Submitted 2 graduate-funding proposals to national agencies: 1) NSF GRFP and 2) DOE CSGF
 - Assisted in the public deployment of a Shiny App for 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. (3) and undergraduate students (2) on R/python/shell programming, functional and object-oriented programming, version control, data wrangling, developing ML models, hypothesis testing, web development, presenting scientific research, and more.
 - Interviewed post-doctoral and research assistant candidates.
 - Hosted workshops on shell programming, Git version control, linux computing over SSH, and containerization with Docker and Singularity.
 - Performed system administrator tasks (e.g., server onboarding, dependency/user/data/resource management) for the JRaviLab node on the CU computing cluster.
 - Extensive GitHub code review and debugging assistance peers' projects.
 - Revisions and figure generations for grant proposals.

- Student Research Assistant** **May 2022 — Nov 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.
 - Implemented machine learning (logistic regression and PCA) and non-machine-learning statistical methods (Fisher's Exact test) to study viral genomic features associated with host specificity.
 - 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

- Awarded National Institute of Health (NIH) Intramural Research Training Award (IRTA) Summer Student Traineeship **2024**
- Submitted Fall 2023 NSF Graduate Research Fellowship Program (GRFP) titled "*Microbial phenotype prediction with graph machine learning methods.*" (Not awarded) **2023**
- Submitted Winter 2023 Department of Energy Computational Science Graduate Fellowship (DOE CSGF) titled *A multi-modal deep learning technique predict antibiotic resistance via computational chemistry and bacterial genomics.* (Not awarded) **2023**
- Contributed minor revisions to NIH NIAID U01 grant submission, *Awarded; PI: Janani Ravi.* **2023**

PRESENTATIONS

Research and technical talks

- **Jul 2023.** R/Bioconductor: Cancer and Evolution track. *MolevolvR a web-app for protein characterization.* [Web app: jrvilab.org/molevolvR]. Boston University, Boston, MA.
- **May 2023.** Great Lakes Bioinformatics Conference. *MolEvolvR a web-app for protein characterization* [Web app: jrvilab.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 University, Montreal, CA.

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 **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.
- Collecting performance analytics from the application to implement a job queuing system which estimates run-times from user inputs.

Accurately predicting AR in ESKAPE pathogens using machine learning **2022 — present**

- Implemented various ML approaches and evaluation techniques: logistic regression, gradient boosting machines, random forests, linear discriminant analysis, stratified-cross-validation, class-weighting, hyperparameter 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 **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 **2023**
- Jill Bilodeaux; Microbiology program, CU Anschutz **2023**
- Charmie Vang; Biomedical Sciences program, CU Anschutz **2023**

Undergraduates

- Ethan Wolfe; B.S. Biochemistry & Molecular Biology with CMSE and additional minors, Michigan State University **2022 — present**
- Skylar Stefanowicz; B.S. Biology, Metropolitan State University of Denver **2024**
Denver Denver, Colorado GPA Unweighted - 4.0

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