

## 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 **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
  - Transferred

## 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 (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
- Leading a machine learning and gene-rank based study of antibiotic resistance across the ESKAPE pathogens (*manuscript in preparation*)
  - Full stack developer 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 (jrvilab.org/molevolvr)
  - Acknowledged in *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*) publication for assisting in re-deployment of the R Shiny dashboard.
  - Mentoring Ph.D. and undergraduate students on project directions and common data science tasks: data wrangling, model development, and model analysis.
  - Interviewing post-doctoral and research assistant candidates.
  - Hosted workshops on shell programming, Git version control, and working on linux machines over SSH.
  - Performing system administrator duties (e.g., server onboarding, dependency/user/data/resource management) for the JRaviLab group's private web-server.
  - Reviewing code and debugging various lab projects.
  - Generating data and figures for grant proposals.
- 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.
  - 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

---

- Submitted Fall 2023 NSF Graduate Research Fellowship Program (GRFP) titled "*Microbial phenotype prediction with graph machine learning methods.*" (In review)
- 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.* (In review)
- Contributed minor revisions to NIH NIAID U01 grant submission, *Awarded; PI: Janani Ravi.*

## PRESENTATIONS

---

### Research and technical talks

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

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, 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

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
- Jill Bilodeaux; Microbiology program, CU Anschutz
- Charmie Vang; Biomedical Sciences program, CU Anschutz

Spring 2023

Spring 2023

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](mailto: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](mailto:faisal.alquaddoomi@cuanschutz.edu)

— IT Principal Professional, Department of Biomedical Informatics, CU Anschutz

**Arjun Krishnan, Ph.D.** ; [arjun.krishnan@cuanschutz.edu](mailto:arjun.krishnan@cuanschutz.edu)

— Associate Professor and Principal Investigator of the Krishnan Lab, Department of Biomedical Informatics, CU Anschutz