

EDUCATION

- Ph.D. Computer Science** | University of Colorado Boulder | Boulder, CO Sep 2024 — present
- Advisor: Ryan Layer
- Non-degree** | University of Colorado Anschutz Medical Campus | Aurora, CO Sep 2023 — Dec 2023
- GPA: 4.00/4.00
 - BIOS 7747: Machine Learning for Biomedical Applications, graduate course offered by the Colorado School of Public Health
- Bachelor of Science, Neuroscience** | Michigan State University | 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, Linux, High performance computing, Git/GitHub, Data mining, Data wrangling, Data visualization, Machine learning, VsCode, Command-line, Vim, Docker, Singularity, Web development (R Shiny), C/C++, LaTeX
- Interests** Computational biology, Machine learning, Statistics, Algorithms, High performance Computing, Bioinformatics, Personalized medicine, Genomics, Software development

RESEARCH EXPERIENCE

- Information Sciences Professional** Nov 2022 — 2024
- JRavi Lab | Department of Biomedical Informatics | PI: Janani Ravi
University of Colorado School Anschutz School of Medicine | Aurora, CO
- Developed an R package (github.com/jravidlab/amr) for machine learning analysis of antimicrobial resistant (AMR) bacterial pathogens (*manuscript in preparation*)
 - Full stack web 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 graduate-level fellowships to national agencies: 1) NSF GRFP and 2) DOE CSGF
 - Presented research and software development workshops at international conferences: 1) R Bioconductor 2023 and 2) Great Lakes Bioinformatics 2023
 - Led journal club presentations and workshops on shell programming, Git version control, remote linux computing, and Docker
 - Maintained lab's GitHub organization: code review plus managing issues and pull requests
 - Peer-mentored Ph.D. (3) and undergraduate students (3) on R/Python/Shell programming, version control, data wrangling, developing ML models, hypothesis testing, web development, and presenting scientific research
 - Deployed an R Shiny dashboard's publication to shinyapps.io *The Phage-shock-protein (PSP) Envelope Stress Response: Discovery of Novel Partners and Evolutionary History*. DOI: doi.org/10.1101/2020.09.24.301986 (accepted at *mSystems*)
 - Interviewed post-doctoral and research assistant candidates
 - Performed system administrator tasks (e.g., server onboarding, dependency/user/data/resource management) for lab's server
 - Contributed ideas, revisions, and figures 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
- Developed machine learning classifiers trained on viral omics data to predict plant virus host and viral taxonomy
 - Analyzed the association of protein domains with viral host phenotype through Fisher's Exact test and
 - Reduced large viral feature spaces with principal component analysis (PCA)
 - 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 2024 submission).
3. Krol, J. D. et al. ESKAPEing antimicrobial resistance in R with open-source AMR-predictive models, feature datasets, and summarized results. *In preparation*. (Planned for 2024 submission to Bioconductor).

AWARDS AND FUNDING

- **Awarded** NSF ACCESS explore project allocation for high performance computing. Project summary "Processing sequences into feature datasets for microbial genotype to phenotype machine learning" **2024**
- **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.* University of Colorado Anschutz, Aurora, CO.
- **Apr 2024** American Society for Microbiology Rocky Mountain Branch *Robust machine learning-based classification of antimicrobial resistance in high-impact pathogens* University of Colorado Boulder, Boulder, CO.

RESEARCH PROJECTS

MolEvolvR: A web-app for characterizing proteins using molecular evolution and phylogeny **2022 — present**

- Improved code base quality by resolving over 50 GitHub issues, containerizing app components, applying linters, and modularizing redundant code
- Benchmarked and tuned parameters for protein homology search and annotation tooling
- Collected performance analytics and used statistical summaries 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
- Clustered over 100GB of genomic data on CU Anschutz Alpine computing cluster
- Developed a modular R package for project source code, data, and models

Prediction of plant virus hosts using omics data and machine learning **2022**

- Developed machine learning classifiers trained on viral omics data to predict plant virus host and viral taxonomy
- Featurization of protein sequences and 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

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 — 2024**
- Skylar Stefonowicz; 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