

Denver, CO & Saline,  
MI  
✉ jacob.krol@  
cuanschutz.edu  
in jacob-krol-b3b784156  
🐙 jakekrol

# Jacob Krol

## Computational Biologist/Professional RA

*I enjoy solving relevant, challenging problems in science using computational approaches.*

### Professional Experience

**Information Sciences Professional (PRA), Dept. of Biomedical Informatics, Center for Health Artificial Intelligence, University of Colorado Anschutz School of Medicine - JRaviLab, Aurora, CO. 2022-2023**

- Developed a full prokaryotic pangenomics machine learning (ML) pipeline that collects data from bvbrc, annotates genomes, and generates a pangenome of gene clusters serving as input for predictive antibiotic resistance ML classifiers
- Testing performance across various ML methods Random Forest, Support Vector Machine, and XGBoost across multiple pathogenic organisms (ESKAPE & more)
- Front and backend web development for a protein analysis app (<http://jravilab.org/molevolvr/?r=&p=home>) [R + Shiny + Slurm + batch scripts]
- Hosted multiple workshops on Bash, Git, R package development, and ssh-workflow basics
- Creating custom container environments (Docker & Singularity) for running projects on various machines
- Software used frequently: Github/Git, slurm, PBS torque, R (tidyverse, Bioconductor suite, httr, Shiny, & much more), Python (Pandas, Numpy, Scikit-learn, Matplotlib, BioPython, requests, & more), Bash, Docker, notebook environments (Rmd, Jupyter, & Quarto), various CLI tools, & many more
- Work closely with the Department of Biomedical Informatics software engineering team on web development, server on-boarding, installations, environment setup, and more

**Student Research Assistant II, Computational Mathematics Science and Engineering program, Michigan State University - Krishnan Lab & Malmstrom Lab, East Lansing, MI. 2022**

- Learned, presented on, and implemented statistical methods on viral protein datasets: fisher test, logistic regression, & principal component analysis
- 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
- Analyzing and visualizing model performance with Matplotlib/Seaborn & Scikit-learn performance metrics
- Gave frequent presentations and updates to 3 co-PIs involved in project

## Education

**B.Sc Computational Neuroscience**, *Michigan State University*, **2020-2022**  
East Lansing, Michigan.

- GPA: 3.89
- Graduation Award: 'With Honor'

**Math and Science Transfer Program**, *Washtenaw Community College*, Ann Arbor, MI. **2018-2020**

- GPA: 3.52
- Transferred

## Presentations & posters

**Great Lakes Bioinformatics Conference**, *MolevolvR a web-app for protein characterization*, McGill University, Montreal, CA. **2023**

- Discussed the development and future directions of a web-app I develop: <http://jrvilab.org/molevolvr>

**Great Lakes Bioinformatics Conference**, *How and when to build a web-app or R package?*, McGill University, Montreal, CA. **2023**

- Co-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/jrvilab/iprscanr>. My section was 1 hour out of the total 4.

**Bioconductor**, *MolevolvR a web-app for protein characterization*, Harvard University, Boston, MA. **2023**

- Latest developments on protein analysis web application, MolEvolvR

## Professional Summary

I make and build upon bioinformatics tools using R, Python, and Bash/shell with basic familiarity of other common languages like Javascript, C, Java, and Perl. I've done a lot of data wrangling with large biological datasets on and off HPC environments. Primarily, I work with Omics data and use ML methods to bridge genotype features to phenotypic outcomes. I place high importance on knowing the underlying principles of statistical methods and how to use them effectively. I've also been maintaining and upgrading a protein analysis app (<http://jrvilab.org/molevolvr/?r=&p=home>) built with the R shiny framework which now uses a Slurm scheduler (previously PBS torque) on the backend to handle job submissions. Using Git, Docker, virtual environments, and linters, I try to ensure my code is reproducible and readable. I've been using Linux-based OSs for over 3 years, and through work I've spent a good amount of time assisting undergrads and grad students in working from the command line including multiple in-person and remote workshops.

## References

**Janani Ravi**, *University of Colorado Anschutz School of Medicine*, **2022-2023**  
Aurora, CO.

- janani.ravi@cuanschutz.edu
- Principal investigator of JRViLab

**Faisal Alquaddoomi**, *University of Colorado Anschutz School of Medicine*, **2022-2023**  
Aurora, CO.

- faisal.alquaddoomi@cuanschutz.edu
- Software engineer

**Arjun Krishnan**, *Michigan State University*, East Lansing, MI. **2022**

- arjun.krishnan@cuanschutz.edu
- Principal investigator of Krishnan Lab