

# JACOB KROL | CURRICULUM VITAE

Computational Biologist/Professional RA

---

🐙 GitHub: jakekrol  
in LinkedIn: jacob-krol  
✉ Email: jacob.krol@cuanschutz.edu  
📍 Denver, Colorado & Saline, Michigan; United States

## Education

---

**Non-degree seeking** - Colorado University Anschutz School of Medicine- in-progress 2023

- Taking BIOS 7747 (Machine Learning for Biomedical Applications) course using employee tuition benefits

---

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

- GPA: 3.89/4.0
- Graduation Award: 'With Honor'
- Semester awards: Dean's List

---

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

- GPA: 3.52/4.0
- Semester awards: Honor Roll
- Transferred

## Publications

MolEvolVR: A web-app for characterizing proteins using molecular evolution and phylogeny  
**Jacob D Krol\***, Joseph T Burke\*, Samuel Z Chen\*, Lo M Sosinski\*, Faisal S Alquaddoomi, Evan P Brenner, Ethan P Wolfe, Vincent P Rubinetti, Shaddai Amolitos, Kellen M Reason, John B Johnston, Janani Ravi  
bioRxiv 2022.02.18.461833; doi: <https://doi.org/10.1101/2022.02.18.461833> (\* indicates co-primary author)

(In-progress) **Jacob D Krol**, Ethan P Wolfe, Evan P Brenner, Keenan Manpearl, Vignesh Sridhar, Joe Burke, Jill Bilodeaux, Janani Ravi Accurately predicting AR in ESKAPE pathogens using ML.

## 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

- 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>.

---

**Bioconductor** - MolevolVR a web-app for protein characterization - Boston University, Boston, MA 2023

- Presented on MolEvolVR application methodology for the Cancer and Evolution talks section

## Grants and Funding

- Applied for Fall 2023 NSF GRFP (in-review)
- Contributed to minor revisions U01 grant submission (awarded)
- Contributed revision and figures to R01 grant submission (in-review)

## 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

- 
- First author on preprint publication for a protein analysis app (<http://jravilab.org/molevolvr/?r=&p=home>); front and backend web development [R + Shiny + Slurm + batch scripts]
  - Developed a prokaryotic pangenomics machine learning (ML) pipeline for predicting antibiotic resistance in bacterial strains (publication/pre-print coming soon)
  - Implemented various ML classifiers including Random Forest, Support Vector Machine, Logistic Regression, and XGBoost
  - Used common classifier performance metrics and plots such as AUROC, AUPRC, and balanced accuracy
  - Analyzing feature across classes using fisher's test, evaluating hamming distance metrics, and unsupervised ML clustering
  - 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
  - Assisted in writing and generating figures for grant proposals and publications
  - Work closely with the Department of Biomedical Informatics software engineering team on web development, server on-boarding, installations, environment setup, and more
  - Assisted in software troubleshooting/guidance for other projects in lab

- 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
- Further preparing data with one-hot-encoding and z-score normalization
- Analyzing and visualizing model performance with Matplotlib/Seaborn & Scikit-learn performance metrics

### Professional Summary

- Developed bioinformatics tools and worked on microbial/viral omics projects using R, Python, and Bash/shell with basic familiarity of other common languages like Javascript, C, Java, and Perl
- Use large, public databases (e.g., BVBC, NCBI, InterPro) with applied machine learning to study the relationship between viral/bacterial genotypes and phenotypes
- Extensive remote server work on our lab webserver, CU Alpine HPC cluster, and MSU's Compute HPC
- Attended international conferences for workshops and presentations, hosted department programming workshops
- Developed a protein analysis web-app MolEvolvR
- Practiced code reproducibility with containers, virtual environments, and version control
- Practiced code readability with linters, documentation, and code reviews
- Exceptional knowledge of POSIX command-line tools and commonly assist in system administrator tasks on my lab's webserver
- Applied various statistical methods to analyze omics data such as supervised classification, unsupervised clustering, fisher's test, principle component analysis, ROSE oversampling, etc.
- Regularly implement and self-study statistical methods; I also like to hosts journal clubs on papers that cover essential, relevant statistical topics like over/under-sampling methods, dimensionality reduction techniques, and classification models/performance-evaluation.
- Assist in various sub-tasks throughout the lab such as one-on-one mentorship with undergraduates and graduate students

### Keywords

Python, R, machine learning (ML), High Performance Computing (HPC), Bash, Vi/Vim/Nvim, VsCode, RStudio, SLURM, PBS Torque, Linux, ssh, genomics, proteomics, domains, antibiotic resistance/drugs, pangenomics, Git/Github, Docker, Singularity, web-development, Shiny, tidyverse, tidymodels, Pandas, NumPy, Scikit-learn, Matplotlib, APIs, BVBC, bacteria, host, pathogens, phylogeny, evolution, statistical methods, principal component analysis (PCA), single value decomposition (SVD), logistic regression (LR), random forest (RF), XGboost, support vector machine (SVM), AUROC, AUPRC, fisher test, virtual environments, research, bioinformatics, computational biology, LaTeX

### References

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

2022-2023

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

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

2022-2023

- faisal.alquaddoomi@cuanschutz.edu
- IT Principal Professional

**Arjun Krishnan** - Michigan State University- East Lansing, MI

2022

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