Jacob Krol | curriculum vitae

Computational Biologist/Professional RA

📍 Denver, CO & Saline, MI

✉ jacob.krol@cuanschutz.edu

Summary

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

- Evaluated performance across various ML methods Random Forest, Support Vector Machine, and XGBoost across multiple pathogenic organisms (ESKAPE & more)

- Used common classifier performance metrics and plots such as AUROC, AUPRC, and balanced accuracy

- Front and backend web devlopment 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

- 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

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

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

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://jravilab.org/molevolvr

Great Lakes Bioinformatics Conference - How and when to build a web-app or R package?- McGill Universtiy, 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/jravilab/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, MolEvovlR

Keywords

Python, R, machine learning (ML), High Performance Computing (HPC), Bash, Vi/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, BVBRC, 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

Professional Summary

I work on bioinformatics tools and omics projects using R, Python, and Bash/shell with basic familarity 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 use ML methods to bridge genotypic features to phenotypic outcomes of pathogenic organisms and their hosts. I place high importance on knowing the underlying principles of statistical methods and knowing when and how to use them effectively. I've also been maintaining and upgrading a protein analysis app (`http://jravilab.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 various programming metaskills through workshops: writing R/Py packages, command-line/shell-scripting, virtual environments, containers, Git/Github, HPC, & ssh.

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
- Software engineer

Arjun Krishnan - Michigan State University- East Lansing, MI   2022

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