Rajan Kapoor

PhD Student | 979-398-2987 | r.kapoor@tamu.edu
linkedin.com/in/rkapr | github.com/rkapr

Summary

Detail oriented electrical engineer, with 3+ years of experience in data analysis of bioinformatics datasets in Crop Science with focus on wheat and sorghum. Worked with plant breeders and professionals in Texas A&M AgriLife to identify research problems of interest, creating bioinformatics pipelines to collect, clean and process large publicly available gene expression datasets, and identify and evaluate predictive statistical models from published research for genome wide gene network analysis. The data analysis and resulting hypotheses contributed to three interdisciplinary research proposals to NSF and USDA.

Research Experience

Understanding nitrogen transport pathways related to grain protein content in wheat

• Understanding gene regulatory pathways that transport nitrogen from senescing leaves to grain with goal of understanding mechanisms affecting grain protein content trait.
• Performed WGCNA based gene clustering for seed gene expression data on computing cluster, which discovered multiple MADS box TF as key candidate genes among modules involving amino acid transporters.
• Performed gene ontology analysis which identified two modules involved in activating storage molecules including starch, lipids and protein.
• Identified UMAMI gene family identified using Interpro search, and used publicly available datasets to create heatmaps and homeolog expression plots for exploratory analysis.
• Collected and cleaned publicly available alignment files (CRAM) from multiple SRA projects. using globus command line, extracted reads mapping to these genes using SAMtools, for verifying gene models.
• Performed exploratory analysis of UMAMI gene expression in Arabidopsis to verify orthologs and identify phylogenetically conserved expression patterns.

Key achievement: Results formed the basis of an interdisciplinary USDA funding proposal with Soil and Crop Sciences.

Open source software contributions

• Boolean gene regulatory network estimation using minimum description length principle (MATLAB).
• *Zeroinfl*: A Python package for zero inflated Poisson regression (with Eric Chuu).

Education

PhD, Electrical and Computer Engineering (GPA: 3.82) 2017 – 2021
Texas A&M University, College Station
MS, Electrical and Computer Engineering (GPA: 3.75) 2015 – 2017
Texas A&M University, College Station

Skills

• R, Python, Shell scripting, SLURM and LSF batch processing, MATLAB, SQL, Shiny, Jupyter notebooks, R Notebooks. ML models SVM, clustering, tree-based methods, boosting/bagging.
• BLASTing on cluster, bioinformatics tools (HISAT2, featureCounts, SAMtools, BAMtools, DESeq2), phylogenetic analysis, experience using EMBL-EBI and Uniprot REST APIs, globus-cli, GitHub.
• Unsupervised network learning, identifying hidden patterns from large scale datasets, surveying statistical /bioinformatics /machine learning literature to identify best approaches to solve problem. Some exp. with mixed models, time series analysis, neural networks, image processing

Relevant Courses

Engineering/Statistics: Regression Analysis (A), Statistical computing in R & Python (A), Applied Statistics & Data Analysis (B), Pattern Recognition (A), Distribution Theory (A), Information Theory (A).
Biology/Bioinformatics: Bioinformatics (S), Bioinformatics Command Line (A), Metagenomics (A), Differential Gene Expression (A), Genome Assembly (A).
Online Certifications: Introduction to SQL (Data Camp), Learn the Command Line (Code Academy).

Publication

R. Kapoor, A. Datta, C. Sima, J. Hua, R. Lopes and M. L. Bittner, "A Gaussian Mixture-Model Exploiting Pathway Knowledge for Dissecting Cancer Heterogeneity," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 17, no. 2, pp. 459-468, 1 March-April 2020, doi: 10.1109/TCBB.2018.2869813.

Blogs/Tutorials

Unix scripting tutorial for processing public RNAseq datasets for gene model verification in wheat.

Honors/Awards

• Texas Engineering Experiment Station (TEES) Research Assistantship, Sept 2015 - Dec 2016, Summer 2017 – Fall 2020.
• Texas International Student Scholarship, Fall 2015, Spring 2016.
• Runner up to Best BTech project award in Electrical Engineering (IIT Patna), 2014.
• Gandhian Young Technology Award, 2014.
• State of Rajasthan (India) Academic Excellence Award, 2009.

Hobbies

Line drawing and cooking.

Work Authorization

Legally authorized to work under CPT on F1 visa.

References

Available upon request.