



Upendra Kumar Devisetty

Greenlight Biosciences
Durham, North Carolina - 27709

E-mail: upendrakumar.devisetty@gmail.com
Website: <http://upendrak.github.io/>
Phone: (520)-392-2482

SKILLS

Languages: Python, R, SQL, Bash

Tools: Pandas, Matplotlib, Seaborn, NumPy, Keras, PySpark, Scikit-learn, Jupyter, Gith, Snakemake, Shiny, Flask, Big-Query, MySQL, Postgres, AWS, Ansible, Openstack, Heroku, Docker and familiarity with Kubernetes

Machine Learning: Data wrangling, Feature engineering, Regression, Classification, Clustering, Decision Trees, Ensemble methods, Convolutional Neural Networks, NLP, Recommendation systems

Techniques: DevOps, Hadoop, HPC, GPU and distributed computing

WORK EXPERIENCE

Greenlight Biosciences
Senior Data Scientist

Jan 6th 2020

- Delivering machine learning models, and data-driven insights, that will drive the company's data-based value creation
- Next-Generation Sequencing analysis of the transcriptome/genomic data to help deliver high-quality targets for RNAi
- Setting up the computational infrastructure that meets the computational requirements of the Data Science Team

CyVerse, University of Arizona
Science Informatician

2016 — 2019

- Integrated Work-Queue MAKER genome annotation pipeline in the Jetstream cloud
- Optimized the process of bringing tools and workflows into CyVerse Discovery Environment using Bioconda and Biocontainers
- Constructed evolutionary analysis of Long-intergenic Non-Coding RNA pipeline from RNA-Seq data
- Integrated Gene expression matrix Pegasus pipeline in the Jetstream cloud
- Integrated more than 150 Bioinformatics apps and images into Cyverse

Insight Data Science, Seattle, WA
Data Science Fellow

Jun 2019 — Sept 2019

- Built PlantMD (<http://plantmd.onrender.com/>), an image-based web app that can rapidly and accurately diagnose plant diseases with 99% accuracy achieving an ROC-AUC score of 0.92.
- Trained and validated Alexnet and VGG16 CNN architectures. Used (100K, 500GB) diseased and healthy plant leaf images using Keras on Google Collabs GPU nodes.
- Used Docker, Github, Dockerhub and AWS and to automatically manage building and deploying PlantMD

Insight Data Science, Seattle, WA
Technical Consultant

Sept 2019 — Feb 2020

Developed and delivered workshops such as Introduction to AWS for Data Scientists, Big Data processing platforms (Hadoop and Spark), Flask web development, ML model deployment using Heroku to Data Engineering fellows.

University of Arizona

2018 — 2019

Co-Principal Investigator

- Mine public RNA-seq from 15 agriculturally significant or model systems to create a dynamic lncRNA dataset using Evolinc
- Enhance Evolinc's lncRNA identification and comparative analyses capabilities

Datacamp

2018 — Present

DataScience Instructor

- Designed and developed course content for Big Data Fundamentals via PySpark (<https://www.datacamp.com/courses/big-data-fundamentals-via-pyspark>) using Apache PySpark and its components (RDD, DataFrames, SparkSQL and MLlib). The course has over 8000 students till date.
- Developed Top 10 programming languages for Data Scientists project for Datacamp to analyze 100K responses from Stack Overflow 2019 Developer Survey using PySpark and Seaborn

Oregon State University

2015 — 2016

Research Associate

- Developed cheap and high-throughput DNA extraction and DNA library construction protocols for structural polymorphism discovery in Populus
- Implemented TASSEL GBS pipeline for Populus
- Constructed GBS libraries and analyzed GBS data for understanding Aspen phylogeography

University of California Davis

2010 — 2015

Postdoctoral Fellow

- Detected molecular genetic markers and constructed the novel comprehensive transcriptome assembly pipeline using deep RNA-Seq data of Brassica rapa.
- Developed pipelines and novel tools for assembly validation and assemblers comparison of Brassica rapa.
- Optimized the high-throughput RNA-Seq protocol for making ~1000 and ~2000 libraries from mapping populations of Brassica rapa and Arabidopsis thaliana respectively.
- Determined the genotype of mapping population of Brassica rapa using RNA-Seq data and constructed first ever genetic map using coding genetic markers
- Conducted RNA-Seq expression analysis on Phytochrome mutants of B. rapa.
- Conducted research to uncover novel differentially expressed genes and pathways associated with shade avoidance response in Brassica rapa.
- Provided bioinformatics research support to other projects in the lab including Arabidopsis RNA-Seq, Tomato RNA-Seq, Bacterial genome assembly and annotation.

University of Nottingham

2006 — 2009

Research Technician

- Cloned and isolated RNAase and SLF alleles of Petunia hybrida using 3' & 5' RACE, Southern blotting and western blotting for molecular analysis.
- Performed Gateway cloning, yeast two-hybrid assays of petunia and cherry anther and style library screens with identified baits.
- Conducted Agrobacterium infection of RNAi constructs into petunia lines to generate transgenic plants lacking incompatibility alleles. Propagated the transformants to establish and maintain glasshouse grown transgenic stocks.

- Provided technical support to line manager and guided two Master's students and a B.Sc., student towards completing their dissertation.

National Chemical Laboratory

2003 — 2005

Junior Research Fellow

- Developed tissue culture regeneration protocol for rapid multiplication
- Developed Agrobacterium mediated transformation protocols

GRANTS AND AWARDS

Grants:

- 2018 MINE-PGR: "Mining public RNA-seq data to identify and annotate long non-coding RNAs in fifteen diverse angiosperms". Duration: 2 Year. Budget - \$232,462.00. Role Co-PI
- 2017 Extreme Science and Engineering Discovery Environment grant: "Sociality in larval insects: the underlying genomics", Duration: 1 Year. Budget - \$166,918.00. Role PI
- 2017 Extreme Science and Engineering Discovery Environment grant: "WQ-MAKER: A flexible and scalable genome annotation platform on Jetstream cloud". Duration: 1 Year. Budget - \$166,918.00. Role Co-PI
- 2015 International Joint Research and Development grant: "Build-up of plant breeding prediction pipeline: Based on big data of Genomics, Phenomics, Phenotyping and Environment". Duration: 3 Years. Budget - 500M KRW/yr. Role: Co-PI
- 2013 Extreme Science and Engineering Discovery Environment grant: "Genotyping by Sequencing and Detection of eQTLs in a Recombinant Inbred Line Population of Brassica rapa". Duration: 1 Year. Budget - \$166,918.00. Role: PI
- 2013 Data Intensive Academic Grid grant from University of Maryland for performing high-throughput data analysis. Role: PI

Awards:

- 2015 - Post Doctoral Student Association (PSA) travel award (\$400) from University of California, Davis
- 2015 - Travel grant (\$400) from iPlant Collaborative for attending Plant and Animal Genome Conference, SanDiego
- 2014 - University of California, Berkeley GSL Ion Torrent Proton Grant winner. Funding for Proton RNA-Seq pilot studies at the Genomics Sequencing Laboratory, University of California, Berkeley.
- 2005 - Doctoral Training fellowship (£9000) from the University of Nottingham UK. Towards living expenses for PhD studies in UK.
- 2005 - Full Tuition Fee Research scholarship (£45,000) from University of Nottingham, U.K. Towards pursuing PhD studies in UK.
- 2001 - Dept. of Biotechnology (Govt. of India) Scholarship. Towards pursuing Master's studies in India
- 2002 - Joint CSIR-UGC Junior Fellowship and Eligibility for Lectureship –National Eligibility (NET) from CSIR, India

PUBLICATIONS AND BOOK CHAPTERS

Publications:

- Sateesh Peri, Sarah Roberts, Isabella R Kreko, Lauren B McHan, Alexandra Naron, Archana Ram, Rebecca L Murphy, Eric Lyons, Brian D Gregory, **Upendra K Devisetty**, Andrew DL Nelson Read Mapping and Transcript Assembly: A Scalable and High-Throughput Workflow for the Processing and Analysis of Ribonucleic Acid Sequencing Data. *Frontiers in Genetics* 10:1361
- Muhammad Arslan, **Upendra Kumar Devisetty**, Martin Porsch, Ivo Große, Jochen A Müller, Stefan G Michalski. RNA-Seq analysis of soft rush (*Juncus effusus*): transcriptome sequencing, de novo assembly, annotation, and polymorphism identification: *BMC Genomics* 20:489
- Johannes Köster Björn Grüning, Ryan Dale, Andreas Sjödin, Brad A. Chapman, Jillian Rowe, Christopher H. Tomkins-Tinch, **Upendra Kumar Devisetty** et al.,

(2018). Bioconda: sustainable and comprehensive software distribution for the life sciences: *Nature Methods* 15:475-476

- Hazekamp, Nicholas L., **Upendra K Devisetty**, Nirav Merchant and Douglas Thain. "MAKER as a Service: Moving HPC applications to Jetstream Cloud." (2018).
- Markelz, R J Cody; Covington, Michael F; Brock, Marcus T; Devisetty, **Upendra K**; Kliebenstein, Daniel J; Weinig, Cynthia; Maloof, Julin N. Using RNA-seq for Genomic Scaffold Placement, Correcting Assemblies, and Genetic Map Creation in a Common *Brassica rapa* Mapping Population. *G3: Genes | Genomes | Genetics* g3.117.043000 (2017)
- Joyce, B., Haug-Baltzell, A. K., Hulvey, J. P., McCarthy, F., **Devisetty, U. K.**, Lyons, E. Leveraging CyVerse Resources for De Novo Comparative Transcriptomics of Underserved (Non-model) Organisms. *J. Vis. Exp. ()*, e55009, doi:10.3791/55009 (2017)
- Wang, Z, R. Yang, **U. Devisetty**, J. Maloof, Y. Zuo, J. Li, Y. Shen, J. Zhao, M. Bao and G. Ning (2017). "The divergence of flowering time modulated by FT/TFL1 is independent to their interaction and binding activities." *Frontiers in Plant Science* 8(697).
- Andrew D. Nelson*, **Upendra K. Devisetty***, Kyle Palos, Asher K. Haug-Baltzell, Eric Lyons, Mark A. Beilstein (2017). "Evolinc: a comparative transcriptomics and genomics pipeline for quickly identifying sequence conserved lincRNAs for functional analysis". *Frontiers in Genetics*. 1(10). * These authors contributed equally to this manuscript
- Andrew D. L. Nelson, Evan S. Forsythe, **Upendra K. Devisetty**, David S. Clausen, Asher K. Haug-Batzell, Ari M. R. Meldrum,* Michael R. Frank, Eric Lyons, and Mark A. Beilstein. A Genomic Analysis of Factors Driving lincRNA Diversification: Lessons from Plants. *G3* (2016)
- **Devisetty UK**, Kennedy K, Sarando P et al. Bringing your tools to CyVerse Discovery Environment using Docker [version 1; referees: awaiting peer review]. *F1000Research* 2016, 5:1442 (doi: 10.12688/f1000research.8935.1)
- Marcus T. Brock, Lauren K. Lucas, Nicholas A. Anderson, Matthew J. Rubin, R. J. Cody Markelez, Michael F. Covington, **Upendra K. Devisetty**, Clint Chappel,‡ Julin N. Maloof and Cynthia Weing. Genetic architecture, biochemical underpinnings, and ecological impact of floral UV patterning. *Molecular Ecology*: (2016) 25, 1122–1140
- Robert L. Baker, Wen Fung Leong, Marcus T. Brock, Robert C. Markelz, Mike Covington, **Upendra K. Devisetty**, Julin Maloof, Stephen Welch, and Cynthia Weinig: Modeling leaf development enables quantitative trait mapping mapping of novel loci and reveals independent genetic modules for leaf size and shape in *Brassica rapa*. *New Phytology* 2015:257-268
- Kazunari Nozue, An Tat, **Upendra Kumar Devisetty**, Matt Robinson, Maxwell Mumbach, Yasunori Ichihashi, Saradadevi Lekkala, and Julin N. Maloof: Shade Avoidance Components and Pathways in Adult Plants Revealed by Phenotypic Profiling. *PLOS GENETICS* 2015: DOI: 10.1371/journal.pgen.1004953
- **Upendra Kumar Devisetty**, Mike Covington, An V. Tat and Julin N. Maloof: Using deep RNA-Seq for polymorphism detection and improving genome annotation of *Brassica rapa* – *G3* 2014 0:g3.114.012526v1-g3.114.012526
- Daniel Koenig, José M. Jiménez-Gómez, Seisuke Kimura, Daniel Fulop, Daniel H. Chitwood, Lauren R. Headland, Ravi Kumar, Michael F. Covington, **Upendra Kumar Devisetty**, and Julin N. Maloof: Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *PNAS* 2013: 1309606110v1-201309606
- Polly Yingshan Hsu, **Upendra K Devisetty** and Stacey Harmer: Accurate timekeeping is controlled by a cycling activator in *Arabidopsis*. *eLife* 2:e00473
- **Upendra Kumar Devisetty**, Katie Mayes and Sean Mayes: The RAD51 and DMC1 homoeologous genes of bread wheat: Cloning, molecular characterization and expression analysis. *BMC Research Notes* 2010, 3:245

Book chapters:

- **Upendra Kumar Devisetty** and Susan Bush. Cloud Computing for Bioinformatics: Meeting the Challenges of Big Data Analysis, Storage and Integration. Amazon Digital Services, Inc.
- **Upendra Kumar Devisetty** and Demudu Naidu Lekkala. The role of Next-generation sequencing in shaping the future of life science research in the 21st century, PP 5-26. In: A.K.Roy (eds.). Emerging Technologies of the 21st century. New India Publishing Agency, 101, Vikash Surya Plaza, New Delhi-110034, India, ISBN 978-93-83305-33-9:890p

SYNERGISTIC ACTIVITIES

Invited presentations:

- 2018 - Containerization and Workflows in CyVerse (Keynote speaker at Rocky Mountain Genomics Hackathon)
- 2017 - WQ-Maker: A Flexible and Scalable Genome Annotation Pipeline on Jetstream Cloud (Plant And Genome conference)
- 2016 – Biostatistics: 5th International Conference on Biometrics & Biostatistics
- 2015 - Development of ICT-based Novel Plant Breeding Prediction Pipeline integrating Genomics, Transcriptomics, Phenotypic and Environment data from Synthetic Allopolyploids of Brassica at Fungi and Plant, Korean company in South Korea
- 2015 - Development and Application of Genomic Resources in Brassica rapa at Brassicas workshop (Plant And Genome conference)
- 2015 - A Hybrid Approach to Assemble and Annotate the Brassica rapa Transcriptome in the Cloud through the iPlant Collaborative and XSEDE (Plant And Genome conference)
- 2012 - Department of Plant Biology Annual Retreat – “Agroecological annotation of gene function and computational analysis of gene networks”
- 2009 - Molecular Cloning, Characterization and Functional Analysis of Meiotic recombination gene homoeologues in polyploid Wheat (Plant and Animal Genome Conference)

Hackathons:

- 2018 - Developing a Machine Learning Framework for identifying and classifying noncoding RNAs (<https://github.com/NCBI-Hackathons/LNCPEP>) (Rocky Mountain Genomics Hackathon). Role: Team leader
- 2018 - Gerber: Generalized Easy Reproducible Bioinformatics Environment wRapper (<https://github.com/NCBI-Hackathons/ContainerInception>) (UCSC Genomics Hackathon).
- 2016 - UltraFast Expressed Variant Calling (https://github.com/NCBI-Hackathons/Ultrafast_Mapping_CSHL) (Cold Spring Harbor Laboratory)

Workshops:

- Certified Software and Data carpentry instructor
- Lead instructor for Software Carpentry workshop in Tucson (<https://uhilgert.github.io/2017-08-26-Tucson>)
- Lead instructor for Bash shell scripting and GIT: Data and Software Carpentry workshop, Tucson, Arizona
- Lead instructor for Software and Data Carpentry workshop in Stanford (<https://loriling96.github.io/2018-03-29-stanford/>)
- Lead instructor for CyVerse Container Camp workshop at Tucson, Arizona (<http://www.cyverse.org/cc>)
- Lead instructor for ANGUS 2-week workshop at the University of California, Davis (2017 and 2018)
- Lead instructor for workshop on bringing your Bioinformatics tools to CyVerse’s Discovery Environment using Docker at Biostatistics and Biometrics conference
- Lead instructor for Data carpentry Genomics workshop at Noble Research Institute (Oklahoma)
- Lead instructor for Cybercarpentry workshop at North Carolina University, Chappel Hill
- Focus forum webinar: Evolinc: Identification and Evolutionary Analysis of Long

non-Coding RNA

- Focus forum webinar: WQ-MAKER: A flexible, scalable genome annotation pipeline on Jetstream cloud
- Mentored California State University and local community college students from under-represented backgrounds in bioinformatics analysis of RNA-Seq data
- Mentored 5 Undergraduate students, 3 Master student and 2 PhD students.
Coordinated projects utilizing resources and personnel across multiple institutions

Memberships:

- Member of University of Arizona's Data Science Institute
(<https://datascience.arizona.edu/person/upendra-kumar-devisetty>)

EDUCATION

Ph.D. (Crop Genetics) 2005 — 2009

University of Nottingham, UK

- Investigated the meiotic recombination in wheat using molecular biology, molecular genetics and field-based crop analysis.
- Cloned RAD51 & DMC1 meiotic homoeologous genes of wheat and investigated their role in meiotic recombination pathway in Arabidopsis using variety of functional genomic approaches

M.Sc., (Molecular Biology and Biotechnology) 2001 — 2003

G.B.P.U.AT, India

Master's thesis: Mode of action of biocontrol agent, *Trichoderma harzianum* on fungi pathogen

B. Sc., (Agriculture) 1996 — 2000

A.N.G.R.A.U, India