

Curriculum Vitae

Sunit Jain

1333 Palos Verdes Dr. Email: mail4sunit@gmail.com
 San Mateo, California 94403 Phone: +1-734-353-2434
 Visa Status: Permanent Resident

SUMMARY

I have over 8 years of experience in Meta'omic data analyses, algorithms design and pipeline development. During this time, I have developed a number of high-throughput cloud scale data-mining pipelines and biological databases all with the goal of answering the 'Who?', 'Why?' and 'What?' about microbial communities in various environments. More recently, I have focused on identifying and nominating proteins from these communities for various areas of interest related to human and plant health. I currently hold 12 publications and 2 patents for my efforts ([Google Scholar](#)).

SKILLS**Computer Skills**

- Languages of choice: Perl, R, Cypher, Shell (bash)
- Familiar with: Python 2/3, C/C++, Java, SAS, SPSS.
- Databases: Neo4j (NoSQL, graph database), DynamoDB, AuroraDB, PostgreSQL, BioSQL, MySQL, Oracle, SQLite.
- Platforms: AWS, Unix/Linux, Windows, Mac.

Bioinformatics

- Public Databases: EMBL, JGI-IMG, KBase, NCBI, KEGG, Meta-Cyc, MG-RAST and many more.
- Tools and Pipelines: AMOS, Arb, Blast, Bowtie, BWA, Consed, Concoct, ESOM, Galaxy, GATK, GMOD, MaxBin2, Megan, Meta-AMOS, MetaPathways, MG-RAST, Mothur, PrinSeq, samtools and many more.
- Assemblers: Celera, Newbler, Velvet, Oases, Trinity, Meta-Velvet, IDBA, Megahit, SOAP, SPAdes, Ray and many more.

Public Code Samples:

[1] <https://github.com/Geo-omics/scripts> (2016) [2] <https://github.com/tetramerFreqs/Binning> (2015)

RELEVANT EXPERIENCE**Meta'omics Scientist | AWS Pipelines Lead****Dec, 2015 – Present***R&D Bioinformatics, Second Genome, Inc.***Management:**

- Oversee the development and deployment of all AWS based pipelines.
- Enforce basic SDLC best practices like the use of git, regular release cycles, CI/CD etc.

Cloud Development:

- Lead Developer for '**Gallifrey**': The cloud based meta-omics analysis pipeline (Perl, BASH, R). This pipeline performs de novo assembly, read alignment to reference databases and Host-RNASeq analysis.
- Lead Developer for '**Holmes**': The universal genome QC pipeline. Perform basic quality checks on genomes at scale, and on the cheap. Feeds directly into the proprietary GreenGenes database.
- Lead Developer for '**SALUT!**': A NLP tool to mine public literature for associations between disease areas and entities (organisms, pathways, etc.).
- Lead Developer for '**MessengerPigeon**': A general purpose tool to scale any unit task on the cloud.
- Developed Shiny GUI interfaces for biologists to be able to launch bioinformatics pipelines (R, Shiny).

Bioinformatics Analysis:

- Bioinformatics Lead for Immuno-oncology bio-therapeutics discovery (2015-2017). Patent filed.
- Developed graph database (Neo4j) pipeline solutions as a means to integrate various types of omics data and nominate protein/peptides candidates to be tested as therapeutics in various disease assays.
- Directly contributed to substantially increasing the size of the proprietary GreenGenes database by scraping and organizing data from public databases.

Data Science:

- Developed an algorithm to discover insecticidal proteins/peptides with novel modes of action (Neo4j, R).
- Developed a RF model to predict the 'expressibility' of a nominated protein. Achieved 86% accuracy* (R).

Research Bioinformatics Specialist**May, 2011–Dec, 2015***Department of Earth and Environmental Sciences, University of Michigan*

- Managed and oversaw all bioinformatics projects for the lab.
- Developed method to discover gene cluster for the production of a cancer treating natural product, **ET-734**, from an environmental sample. **U.S. Patent No. 20150361470**.
- Scaled the method developed for the patent, to find, classify and score novel secondary metabolites from meta-omics data as potential natural products of interest.
- Performed binning, metagenomic and metatranscriptomic analysis on various metagenomic datasets that directly lead to the discovery of numerous novel microorganisms. See 'Publications' for more.
- Developed a pipeline for *de-novo* metatranscriptome assembly. See Baker et al 2013 in Publications.
- Mined meta-omic data from some of the deepest hydrothermal vents in the world. This was challenging especially because of the amount of data and novel organisms present in the datasets.
- Developed the graph database (neo4j) back-end to integrate microbial genomics, environmental chemistry, and ecosystem processes data to understand harmful algal blooms. *Still under active development: [Cyanohub](#)*.
- Developed algorithms to identify and characterize viruses from metagenomic datasets.
- Provided bioinformatics and HPC support to the students, researchers and collaborators of the lab.
- Built custom analysis and *de-novo* assembly tools/pipelines for the various versions of Illumina and PacBio datasets.

Bioinformatics Analyst | Graduate Student Researcher**Aug 2009–May 2011***Michigan Geomicrobiology Lab, University of Michigan*

- Designed and developed Metagenome and Metatranscriptome NGS data analysis pipelines.
- Contributed to the binning and comparative genomic analyses for samples collected from the Guaymas Basin Hydrothermal plume. See Lesniewski et al 2012 in Publications.
- Comparative Genomics study on the Cyanobacteria found at the bottom of Lake Huron. This region is believed to be an analog to study the Proterozoic Eon. See Voorhies et al 2012 in Publications

EDUCATION**University of Michigan***Ann Arbor, Michigan, USA***MS, Bioinformatics**

2011

Amity University*Noida, India***B.Tech, Bioinformatics**

2008

PATENTS**2017**

Provisional patent applied related to the discovery of microbial peptides for immuno-oncology.

2016

Title: Nonribosomal Peptide Synthetases.
 Inventors: David Sherman, Michael Kaufman-Schofield, **Sunit Jain**, and Gregory J. Dick.
 Publication date: 2015/12/17
 Issue date: 2016/11/08
 Patent office: US
 Patent number: 9,487,763 B2
 Application number: 14/713662

PUBLICATIONS

2017

- Stadler, L.B., Delgado Vela, J., **S. Jain**, Dick, G.J., Love, N.G. "Elucidating the impact of microbial community biodiversity on pharmaceutical biotransformation during wastewater treatment." *Microbial Biotechnology*. doi: 10.1111/1751-7915.12870.
- Berry, M.A., White, J.D., Davis, T.W., **S. Jain**, Johengen, T.H., Dick, G.J., Sarnelle, O. and Denef, V.J. "Are oligotypes meaningful ecological and phylogenetic units: a case study of Microcystis in freshwater lakes". *Frontiers in Microbiology*. doi: 10.3389/fmicb.2017.00365.

2016

- Li, M., **S. Jain**, Dick, G.J. "Genomic and Transcriptomic Resolution of Organic Matter Utilization Among Deep-Sea Bacteria in Guaymas Basin Hydrothermal Plumes." *Frontiers in Microbiology*. doi:10.3389/fmicb.2016.01125.
- Den Uyl, P. A., Richardson, L. L., **S. Jain**, and G. J. Dick "Unraveling the physiological roles of the cyanobacterium Geitlerinema sp. BBD and other black band disease community members through genomic analysis of a mixed culture." *Plos One*. doi: 10.1371/journal.pone.0157953.
- Flood, B.E., Fliss, P., Jones, D.S., Dick, G., **S. Jain**, Kaster, A.K., Winkel, M. Mußmann, M., Bailey, J.V., "Single-Cell (Meta-) Genomics of a Dimorphic Candidatus Thiomargarita nelsonii Reveals Genomic Plasticity." *Frontiers in Extreme Microbiology*. doi: 10.3389/fmicb.2016.00603.

2015

- Li, M., B. J. Baker, K. Anantharaman, **S. Jain**, J. Breier, and G. J. Dick "Genomic and Transcriptomic Evidence for Scavenging of Diverse Organic Compounds by Widespread Deep-Sea Archaea", *Nature Communications*. doi: 10.1038/ncomms9933
- Schofield, M. M.*, **S. Jain** *, D. Porat, G. J. Dick and D. H. Sherman. "Identification and analysis of the bacterial endosymbiont specialized for production of the chemotherapeutic natural product ET-743." *Environmental Microbiology*. doi: 10.1111/1462-2920.12908 **[Co-first author]**

2014

- Vorobev, A, S. Jagadevan, **S. Jain**, K. Anantharaman, G. J. Dick, S Vuilleumier, and J Semrau. "How Do Facultative Methanotrophs Utilize Multi-Carbon Compounds for Growth? Genomic and Transcriptomic Analyses of Methylocystis Strain SB2 Grown on Methane vs. Ethanol." *Applied and Environmental Microbiology*. doi: 10.1128/AEM.00218-14.
- Li, M., **S. Jain**, B.J. Baker, C.A. Taylor and G.J. Dick. "Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume". *Environmental Microbiology*. doi: 10.1111/1462-2920.12182.
- Sheik, C., **S. Jain** and G. J. Dick. "Metabolic flexibility of deep-sea Sar324 revealed through metagenomic and transcriptomic analysis." *Environmental Microbiology*. doi: 10.1111/1462-2920.12165.

2013

- Baker, B.J., C. S. Sheik, C. A. Taylor, **S. Jain**, A. Bhasi, J.D. Cavalcoli and G.J. Dick. "Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling." *The ISME Journal*, doi: 10.1038/ismej.2013.85.

2012

- Lesniewski, R.A., **S. Jain**, P.D. Schloss, K. Anantharaman, and G.J. Dick. "The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs." *The ISME Journal*, doi:10.1038/ismej.2012.63.
- Voorhies, A.A., B.A. Biddanda, S.T. Kendall, **S. Jain**, D.N. Marcus, S.C. Nold, N.D. Sheldon, and G.J. Dick. "Cyanobacterial life at low O₂: Community genomics and function reveal metabolic versatility and extremely low diversity of a cyanobacterial mat". *Geobiology* 10: 250-267.

TEACHING

Number	Title	Format	Contributions	Credits	Semesters Taught
EARTH 513	Geomicrobiology	Lecture	Planning, Guest Lectures	3	W14
EARTH 523	Microbial Community Omics	Lecture + Lab	Planning, HPC Omics Labs	2	F14