Jay S. Ghurye

CONTACT
INFORMATION

Website: www.cs.umd.edu/∼jayg
E-mail: jayg@cs.umd.edu
Phone: +1-240-425-2921 (M)

RESEARCH
INTERESTS

Bioinformatics, Computational Biology, Data Mining, Graph Theory

EDUCATION

University of Maryland, College Park, USA

Ph.D. Computer Science, Fall 2014 - ongoing

Advisor: Prof. Mihai Pop

M.S. Computer Science, Fall 2014- Fall 2016

GPA : 3.85 / 4.0

Veermata Jijabai Technological Institute, Mumbai, India

B.Tech. Computer Engineering, June 2010 - May 2014

CGPA : 9.3 / 10 (4th rank in the class of 81)

PROFESSIONAL
EXPERIENCE

- Special Volunteer - March 2017 - Present, National Human Genome Research Institute, NIH
- Research Intern - Summer 2016, Pacific Biosciences, Menlo Park, CA
- Graduate Teaching Asistant - Spring 2015, Database Design, UMD
- Graduate Teaching Asistant - Fall 2014, Object Oriented Programming-1, UMD
- Software Development Intern - Summer-2013, Nomura Services India Private Limited

RESEARCH
EXPERIENCE

Scaffolding and Assembly Correction Using Hi-C Data

Advisor: Jason (Chen-Shan) Chin and Adam Phillippy

This project started as an internship project at Pacific Biosciences. In this work, we are interested in using long range contact information data like Hi-C to correct long read assemblies and get chromosom e level scaffolding. To achieve this, we have developed a tool that takes into account Hi-C contact information and performs centromere to telomere level contiguous scaffolding. Currently, our focus is to find large misassemblies performed by assemblers using Hi-C data and correct them. In long term, we plan to create an end to end pipeline that would cluster the scaffolds according to the chromosomes they belong to.

Publications:

- **Jay Ghurye**, Mihai Pop, Sergey Koren, Derek Bickhart and Chen-Shan Chin "Scaffolding of long read assemblies using long range contact information", BMC Genomics
- **Jay Ghurye**, Arang Rhie, Brian Walenz, Siddarth Selvaraj, Adam Phillippy, and Sergey Koren "SALSA2 - assembly graph aided Hi-C scaffolding", *Under submission*
- **Jay Ghurye**, Sergey Koren, Arang Rhie, Brian Walenz, Siddarth Selvaraj, Anthony Schmitt and Adam Phillippy "Effect of different Hi-C library preparation on eukaryotic genome scaffolding", *Under preparation*

Code: https://github.com/machinegun/SALSA (Implemented in c++ and Python)

Graph Based Approaches for Metagenome Assembly

Advisor: Mihai Pop

This work primarily focuses on finding variants and repeat structures in metagenome data through

assembly. We are interested in finding out interesting patterns in the assembly graph that provide some useful information about biologically relevant features of the organisms present in metagenome. We are developing a metagenome scaffolder that enables scaffolding of metagenome assembles along with variant discovery.

Publications:

- **Jay Ghurye** and Mihai Pop "Better Identification of Repeats in Metagenomic Scaffolding", *Proc. Algorithms in Bioinformatics, WABI 2016, Aarhus-Denmark*
- **Jay Ghurye**, Victoria Cepeda-Espinoza and Mihai Pop "Metagenomic Assembly: Overview, Challenges and Applications", *Yale Journal of Biology and Medicine, Microbiome Issue*
- Olson, Nathan D., Todd J. Treangen, Christopher M. Hill, Victoria Cepeda-Espinoza, **Jay Ghurye**, Sergey Koren, and Mihai Pop "Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes", *Briefings in Bioinformatics (2017),bbx098.*
- **Jay Ghurye**,Todd Treangen, Sergey Koren, Marcus Fedarko, W. Judson Hervey IV and Mihai Pop "MetaCarvel: linking assembly graph motifs to biological variants", *Under preparation*
- **Jay Ghurye**,Todd Treangen, Sergey Koren, Marcus Fedarko, W. Judson Hervey IV and Mihai Pop "Assessing global rates of recombination and gene flux in human microbiomes via contig graphs", *Under preparation*
- Marcus Fedarko, **Jay Ghurye**,Todd Treangen, and Mihai Pop "MetagenomeScope: Web-Based Hierarchical Visualization of Metagenome Assembly Graphs", *Under submission*

Code: https://github.com/marbl/MetaCarvel (Implemented in c++ and Python)

Efficient clustering of phylogenetic marker genes

Advisor: Mihai Pop
In this work, we focus on theoretical and practical issues related to clustering of millions and billions of sequences corresponding to phylogenetic markers such as 16S rRNA gene. To make this process computationally feasible, we have developed an efficient approach that can cluster tens of millions of sequences within a single day.

Publications:

- Brian Brubach, **Jay Ghurye**, Aravind Srinivasan and Mihai Pop "Better Greedy Sequence Clustering with Fast Banded Alignment", *Proc. Algorithms in Bioinformatics, WABI 2017, Boston-Massachusetts*
- Brian Brubach , and **Jay Ghurye** "A Succinct Four Russians Speedup for Edit Distance Computation and One-against-many Banded Alignment", *Accepted for publication at CPM 2018*

Code: https://bitbucket.org/brubach/edit_release (Implemented in Java and Python)

Computational Methods to Analyze Human Behavior During Disasters

Advisor: Vanessa Frias-Martinez
In this work, we developed a Markov chain based model to analyze human behavior using Call Data Records (CDRs) during the floods in Rawanda in April 2012.

Publication:

- **Jay Ghurye**, Gautier Krings and Vanessa Frias-Marinez "A Framework to Model Human Behavior at Large Scale during Natural Disasters", *IEEE conference on Mobile Data Management 2016,Porto-Portugal*

Coursework   **Graduate**: Computational Systems Biology and Functional Genomics, Computational Genomics, Biological Network Modeling, Quantum Computing, Computational Linguistics, Database Management Systems, Distributed Storage System, Randomized Algorithms (Audit), Deep Learning

(Audit)

Undergraduate: Data Structures and Algorithms, Approximation Algorithms, Discrete Structures, Artificial Intelligence

HONORS AND AWARDS	

HONORS AND AWARDS

- Awarded travel grant by NSF to attend IEEE MDM 2016 conference in Porto, Portugal
- Deans Fellowship, Department of Computer Science, University of Maryland, 2014-2016
- Awarded travel grant by Computer Science Department and USENIX to attend FAST-2015 conference
- Recepient of Sir Ratan Tata Scholarship, covering entire tuition fees during undergraduate studies
- Secured 36$^{\text{th}}$ rank in ACM-ICPC Asia Amritapuri Regionals in 2012 amongst 300 teams
- Stood 33$^{\text{rd}}$ in Engineering Common Entrance Test amongst over 100,000 students

PROGRAMMING SKILLS

C, C++, Python, Java, Golang, R, Shell Scripting

SERVICES

- **Reviewer:** Gigascience, BMC Bioinformatics, Transactions on Computational Biology and Bioinformatics
- **Sub-reviewer:** WABI 2016, ACM-BCB 2016
- **Graduate Admission Committee:** Fall 2016, Fall 2017

REFERENCES

Prof. Mihai Pop
Professor
Department of Computer Science
University of Maryland
College Park, 20740
Email: mpop@umiacs.umd.edu

Dr. Adam Phillippy
Investigator
Genome Informatics Section
NHGRI, National Institute of Health
Bethesda, 20892
Email: adam.phillippy@nih.gov

Prof. Michael Cummings
Professor
CMNS-Biology
University of Maryland
College Park, 20740
Email: mike@umiacs.umd.edu

Dr. Jason (Chen-Shan) Chin
Sr. Director, Deep Learning
DNAnexus
Mountain View, CA 94040
Email: jchin@dnanexus.com

Prof. Vanessa Frias-Martinez
Affiliate Assistant Professor
Department of Computer Science
University of Maryland
College Park, 20740
Email: vfrias@umd.edu