SIDDHARTH S. TOMAR

PERSONAL INFORMATION

Born in India, 08 May 1994

email	siddharth.tomar@scilifelab.se
phone	(Swe) +46-702759580 · (Ind) +91-8894636427
website	www.scilen.se
github	siddharthst

EDUCATION

2016-2018	**KTH Royal Institute of Technology**

Master of Science

Master's programme in Molecular Techniques in Life Science, a joint program with Karolinska Institutet and Stockholm University
Duration: 2 years (full time)
Date of graduation: 31-May-2018
Country: *Sweden*

2012-2016	Jaypee University of Information Technology

Bachelor of Technology

GPA: 8.6 out of 10 · *Specialization in Bioinformatics*
Duration: 4 years (full time)
Date of graduation: 09-July-2016
Country: *India*

PROGRAMMING SKILLS

Programming	C++/C (including CUDA and MPI), Java, Python, Perl, R
Libraries	Pandas, Dask, Matplotlib, Altair
Databases	MySQL
Machine learning	Scikit-learn, Tensorflow
Documentation	LaTeX, Markdown
Version control	GIT (Including Automated testing and CI)
Workflows	Time series analysis, Weighted correlation network analysis

COURSEWORK - KEYWORDS

Computer science	Algorithm design and analysis, High performance computing, Cryptography, Object Oriented Programming, Relational Database Management Systems, Machine Learning - Support Vector Machines and Artificial Neural Networks
Molecular biology	Systems biology, homology search, multiple sequence alignment, phylogeny reconstruction, protein topology prediction, molecular dynamics, comparative genomics, proteomics, mass spectrometry, short and long read sequencing

RESEARCH EXPERIENCE

Jun'18–present   KAROLINSKA INSTITUTET - Sweden

SciLifeLab - Stockholm

I am currently working as a research assistant in the Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm. My current work involves large scale analysis of RNA sequencing data. This includes efficient processing of data in a clustered environment and downstream analysis using R and Python. I mainly use Pandas for storing and operating on sequence alignment data, coupled with matplotlib and Altair for static and interactive visualisations. I have also repurposed algorithms and methods used in signal processing for the genomic annotation and discovery of novel elements in mice genome.

Jun–Aug '17   KAROLINSKA INSTITUTET - Sweden

SciLifeLab - Stockholm

My work at SciLifeLab focused on annotating various RNA classes from mice liver at different developmental stages by using total RNA sequencing data. This project involved methods for discovering novel developmental stage-specific noncoding RNAs and correlation with other meta-resources to mark functionality. I mainly used Pandas along with Python REST interface to automatically acquire and create relational data links for my genes, thereby annotating them using multiple meta-resources in using a robust and reproducible framework. During this time I also explored the use of Python interactive visualisation library Bokeh for exploratory analysis.
Reference: Hassan FOROUGHI · hassan.foroughi.asl@ki.se

Jun–Aug '15   INSTITUTE OF ONCOLOGY RESEARCH - Switzerland

IOSI-IOR

I worked under Dr Francesco Bertoni's group at Institute of Oncology Research, Bellinzona, Switzerland on performance profiling of short read aligners (using Nvidia CUDA framework) for testing scalability in GPGPU environment. I studied the impact of GPU based tools on the next generation sequencing analysis pipeline and included cost versus performance analysis based on scaling.

Master's thesis

Micro-evolution of regulatory RNAs
In this master thesis project, I investigated the functional contribution of regulatory RNA in speciation in closely related mice species.

Bachelor's thesis

Improving the computation of positive selection using codon models
In my bachelor's project, I was working on improving the computation of positive selection using codon models. In this project, I implemented the Branch-Site Model on GPU architecture, and test the efficiency gains which are potentially possible. I changed the underlying algorithm to reduce the redundancy and increase the sensitivity by optimising the subtree pruning. I mainly used (CUDA)C++ for this project. The primary challenge of this project was to implement the current model as it is, without any heuristics involved, and still archive exceptionally low runtimes.

PROJECTS

2017   KTH ROYAL INSTITUTE OF TECHNOLOGY

Course Project

Assembling reference transcriptome for *P. abies* using massively parallel and PacBio real time sequencing data. This project involved utilization of large amount of data (>700 million reads) and optimization techniques for transcriptome assembly.
Reference: Lars ARVESTAD · +4687906436 · lars.arvestad@scilifelab.se

2017   STOCKHOLM UNIVERSITY

Course Project

Constructing a SVM based predictor for classifying topological features of β barrels based on their sequence.

2014   Jaypee University of Information Technology

Course Project

Identification of novel leads against dimeric interface(DBP) in *P. vivax* using computer-aided drug design. For this project, I investigated the structural dynamics of Duffy binding proteins with respect to the phylum Apicomplexa.

2013	Jaypee University of Information Technology
Course Project	Created an application written in Java for searching repeat content in a given genome. Simple machine learning technique was used, applying the Hidden Markov Model. Lempel–Ziv–Welch algorithm was also used and modified according to requirements. Perl was used as a wrapper and as text acquisition/annotation search engine.

POSTER PRESENTATION AND CONFERENCES

2018	2nd Uppsala Transposon Symposium
Poster	Micro-evolution of regulatory RNA in M. domesticus and M. castaneus

GRANTS

Nvidia	*2015* - Nvidia GPU Grant Program

OTHER INFORMATION

Scholarships	2017 · Science for Life Laboratory summer fellowship
	2017 · Karolinska Institutet Foundation scholarship

Languages	ENGLISH	· Fluent
	HINDI	· Native

Last updated on April 23, 2019