Aritra Bose

Computational Genomics
IBM T.J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598

Email: a.bose@ibm.com, abose@broadinstitute.org, bosearitra08@gmail.com
Web: `https://aritra90.github.io/`
Phone: +1(518) 522-7975

Research Interests Computational Biology, Machine Learning, Statistical Genetics, Bioinformatics, Data Mining, Randomized Numerical Linear Algebra.

Education

Purdue University — West Lafayette, IN, USA
Ph.D. in *Computer Science* — Aug 2016 - Aug 2019
Advisor: Prof. Petros Drineas

Rensselaer Polytechnic Institute — Troy, NY, USA
M.S. in *Computer Science* — Aug 2014 - Jul 2016

West Bengal University of Technology — Kolkata, WB, India
B.Tech in *Information Technology* — Aug 2009 - Jun 2013

Experience

Broad Institute of MIT and Harvard — Cambridge, MA, USA
Visiting Scientist — Jun 2021 - Present

IBM T.J. Watson Research Center — Yorktown Heights, NY, USA
Research Staff Member — Feb 2021 - Present
Postdoctoral Researcher — Sep 2019 - Feb 2021
Research Intern — May 2018 - Jul 2018
Research Intern — May 2017 - Aug 2017
Research Intern — May 2016 - Aug 2016

Purdue University — West Lafayette, IN, USA
Research Assistant — Aug 2016 - Aug 2019

Rensselaer Polytechnic Institute — Troy, NY, USA
Research Assistant — Aug 2015 - Aug 2016
Teaching Assistant — Aug 2014 - May 2015

Teradata Corporation — Hyderabad, India
Analyst — Oct 2013 - April 2014

Bose Institute — Kolkata, India
Research Trainee — Sep 2012 - Oct 2013

Indian Institute of Technology — Guwahati, India
Summer Intern — May 2012 - Jul 2012

Indian Statistical Institute — Kolkata, India
Winter Intern — Dec 2011 - Mar 2012

Publications

Journals

1. **A. Bose**, F. Utro, D.E. Platt, L. Parida, *Multiple Loci Selection with Multi-way Epistasis in Coalescence with Recombination*, Algorithms 14 (5), 136, 2021.
2. **A. Bose**, D.E. Platt, L. Parida, P. Paschou, P. Drineas, *Integrating linguistics, social structure, and geography to model genetic diversity within India*, Molecular Biology and Evolution 38 (5), 1809-1819, 2021.
3. **A. Bose**, V. Kalantzis, E. Kontopoulou, M. Elkady, P. Paschou, P. Drineas, *TeraPCA: a fast and scalable software package to study genetic variation in tera-scale genotypes*, Bioinformatics, Volume 35, Issue 19, 1 October 2019, Pages 36793683.
4. G. Stamatoyannopoulos, **A. Bose**, A. Teodasiadis, F. Tsetsos, A. Plantiga, N. Psatha, N. Zogas, E. Yannaki, P. Zalloua, K.K. Kidd, B.L. Browning, J. Stamatoyannopoulos, P. Paschou, P. Drineas, *Genetics of the Peloponnesean populations and the theory of the extinction of the medieval Peloponnesean Greeks*, European Journal of Human Genetics, 25(5), pp. 637-645, 2017.

Preprints

5. **A. Bose**, D.E. Platt, N. Haiminen , L. Parida, *CuNA: Cumulant-based genotype-phenotype interaction networks in Parkinson's Disease*, medRxiv, 2021.
6. S. Dey, **A. Bose**, P. Chakraborty, M. Ghalwash, A.G. Saenz, F. Utro, K. Ng, J. Hu, L. Parida, D. Sow, *Impact of Clinical and Genomic Factors on SARS-CoV2 Disease Severity*, medRxiv, 2021.
7. S. Saha †, A.G Saenz †, **A. Bose** †, F. Utro, D.E. Platt, L. Parida, *RubricOE: a learning framework for genetic epidemiology*, medRxiv, 2021. († Equal Contributors)
8. **A. Bose**, M.C. Burch, A. Chowdhury, P. Paschou, P. Drineas, *Structure informed clustering adjusts for population stratification in association studies*, BioRxiv, 2020.
9. **A. Bose**, D.E. Platt, L. Parida, P. Paschou, P. Drineas, *Dissecting Population Substructure in India via Correlation Optimization of Genetics and Geodemographics*, BioRxiv, 2017.
10. S.Hassan, P. Pal Choudhury and **A. Bose**, (2011), *A Quantitative model for Human Olfactory Receptors*, Nature Precedings, npre20126967-2, 2012.

Conferences

11. A. Chowdhury†, **A. Bose**†, S. Zhou, D. P. Woodruff, P. Drineas, *A Fast, Provably Accurate Approximation Algorithm for Sparse Principal Component Analysis Reveals Human Genetic Variation Across the World*, RECOMB 2022. († Equal Contributors)
12. S. Dey†, **A. Bose**†, S. Saha, P. Chakraborty, M. Ghalwash, A.G. Saenz, F. Utro, K. Ng, J. Hu, L. Parida, D. Sow, *Impact of Clinical and Genomic Factors on COVID-19 Severity*, AMIA, 2021. († Equal Contributors)
13. **A. Bose**, M.C. Burch, A. Chowdhury, P. Paschou, P. Drineas, *CluStrat: a structure informed clustering strategy for population stratification*, RECOMB 2020

Abstracts (peer reviewed only)

14. **A. Bose**, M.C. Burch, A. Chowdhury, P. Paschou, P. Drineas, *Structure informed clustering for population stratification and genetic risk prediction*, Annual Meeting of the American Society of Human Genetics, 2019.
15. **A. Bose**, F. Utro, D.E. Platt, L. Parida, *Algorithms to modulate ARG by Selection*, RECOMB-Genetics, 2018. **Selected for Platform presentation.**
16. **A. Bose**, V. Kalantzis, E. Kontopoulou, M. Elkady, P. Paschou, P. Drineas, *TeraPCA: a fast and scalable software package to study genetic variation in terascale genotypes*, Annual Meeting of the American Society of Human Genetics, 2017.
17. **A. Bose**, D.E. Platt, L. Parida, P. Paschou, P. Drineas, *Correlation Optimization of Genetics and Geodemographics*, Annual Meeting of the American Society of Human Genetics, 2016. **Selected for Platform presentation.**

Dissertation

18. **A.Bose**, *Computational Methods for Population Genetics*, `https://doi.org/10.25394/PGS.9752924.v1`, Purdue University, 2019.

Patents

1. Discovering biomarkers via higher-order genotype-phenotype interactions in complex diseases.
   A. Bose, D.E. Platt, N. Haiminen and L. Parida
   Filed

News Articles

1. *Language (not geography) major force behind Indias gene flow.* `https://bigthink.com/culture-religion/indian-genetics`
2. *New study ties Indias genetic diversity to language, not geography.* `https://www.purdue.edu/newsroom/releases/2021/Q1/new-study-ties-indias-genetic-diversity-to-language,-not-geography.html`
3. *In India, People Who Speak the Same Language Have Similar DNA.* `https://theswaddle.com/in-india-people-who-speak-the-same-language-have-similar-dna-study/`
4. *Genetic testing has a data problem. New software can help.* `https://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=298521&org=NSF`.
5. *Genetic testing has a data problem. New software can help.* `https://www.purdue.edu/newsroom/releases/2019/Q2/genetic-testing-has-a-data-problem.-new-software-can-help..html`.

Invited Presentations

- Impact of Clinical and Genomic Factors on COVID-19 Severity
  - IBM Got Science! 2021 Seminar series.
- *Machine Learning framework in Genetic Epidemiology*
  - Broad Institute of MIT and Harvard, Cambridge, MA, Jun 2021.
- *Computational methods in Population Genomics*
  - Regeneron Genetics Center, Tarrytown, NY, Dec 2020.
  - Inari Agriculture Inc., Cambridge, MA, Nov 2020.
  - Allen Institute of Brain Science, Seattle, WA, Nov 2020.

Presentations

- *CluStrat: a structure informed clustering strategy for population stratification*
  - **Platform** presentation in Research in Computational Molecular Biology (RECOMB), held virtually in June 2020.
  - Poster presentation in American Society of Human Genetics (ASHG) meeting 2019, Houston, TX.
- **Platform** presentation on *Algorithms to modulate ARG by Selection* at the RECOMB-Genetics meeting, Paris, April, 2018. (This talk is given by Dr. Laxmi Parida)
- *TeraPCA: A fast and scalable method to study genetic variation in tera-scale genotypes*
  - Poster presentation in Conference of Scientific Computing and Approximation, Purdue University, West Lafayette, IN.
  - Poster presentation in ASHG 2017 meeting, Orlando, FL.
- *Integrating Linguistics, Social Structure and Geography to model genetic diversity within India.*
  - Poster presentation in Summer Intern Showcase 2017, IBM T.J Watson Research Center, NY.
  - Poster presentation in Biology of Genomes (BOG) 2017 meeting, Cold Spring Harbor Labs, NY.
  - **Platform** presentation in ASHG 2016 meeting, Vancouver, BC, Canada. (Abstract selected in top 8% of over 6000 submissions)
  - Poster presentation at Student Research Showcase in Computer Science Department, Purdue University, West Lafayette, IN.
  - Poster presentation in BOG 2016 meeting, Cold Spring Harbor Labs, NY.
  - Poster presentation in Student Research Symposium 2016 in Computer Science Department, Rensselaer Polytechnic Institute.
- **Summer school** on "Mathematics of Data", organized by **Park City Mathematics Institute (PCMI)** and the Institute for Advanced Study (IAS), held in, Midway, Utah, USA.
- **ASHG 2015**, Annual Meeting in Baltimore, MD, USA as a trainee researcher.
- **Gene Golub SIAM Summer School 2015**, held in, Delphi, Greece.

Awards of Merit

- ISCB (International Society for Computational Biology) Travel Fellowship for RECOMB 2020 in Padua, Italy.
- NSF Travel Grants to the following conferences:
  - Biology of Genomes: 2016 and 2017.
  - American Society of Human Genetics (ASHG), 2015 - 2019.
  - International Conference for Distributed Computing and Internet Technologies (ICDCIT) 2017 meeting held at Bhubaneswar, Odisha, India.
- Received a 4 year fellowship from Ministry of Human Resource Development (M.H.R.D), Government of India for significant achievement in Higher Secondary Examination
- Ranked 1 out of 65 students in Department of Information Technology in Meghnad Saha Institute of Technology, Kolkata.
- Won several school, college and corporate level Quiz competitions across India.

Professional Activities

- Reviewer for the following:
  - Journals
    - American Journal of Medical Genetics; BMC Bioinformatics
  - Conferences
    - KDD; NeurIPs; RECOMB; WABI
- Member: American Society of Human Genetics, International Society for Computational Biology.
- Peer Adviser to incoming graduate students in the Computer Science Department in Rensselaer Polytechnic Institute as well as in Purdue University.
- Co-Founder of the Robotics club of Meghnad Saha Institute of Technology which has over 400 students now.

Computer Skills

- ***Languages:*** C, C++, Python, MATLAB, Java, R, AWK, PL/SQL, Scripting(bash,etc), Perl, HTML, LaTeX
- ***Operating Systems:*** GNU/Linux, Unix, Windows
- ***Computational Biology:*** PLINK, SAIGE, GATK, Beagle, Impute2, SNPeff, samtools, bcftools, PeakSeq, Cytoscape and other computational biology and population genetic tools.
- ***Databases:*** MySQL, TERADATA, Oracle, DB2

Graduate Coursework (selected)

Machine Learning, Computational Linear Algebra, Parallel Computing, Foundations of Data Science, Algorithms Design, Frontiers of Network Science, Distributed Systems, Randomized Algorithms, Theory of Computation

Independent Coursework

Coursera.org: Algorithms: Design and Analysis; Bioinformatics I and II; Python for Genomic Data Science

References

Prof. Petros Drineas
Professor
Department of Computer Science
Purdue University
West Lafayette, IN, USA
drineas@gmail.com

Prof. Peristera Paschou
Associate Professor
Department of Biological Sciences
Purdue University
West Lafayette, IN, USA
ppaschou@gmail.com

Dr. Laxmi Parida
IBM Fellow & Manager,
Computational Genomics,
IBM T.J. Watson Research Center
Yorktown Heights, NY, USA
parida@us.ibm.com

Additional Information

- *Date of Birth*: August 8, 1990
- *Marital Status*: Married
- *Citizenship*: Indian