

Derek Cole Aguiar

Curriculum Vitae

July 2019

📍 Computer Science & Engineering Department, University of Connecticut, Storrs, CT 06269
🏠 www.derekaguiar.com
☎ (508) 642-8082
✉ derek.aguiar@uconn.edu

Research Interests

I am interested in advancing CS foundations of machine learning, Bayesian modelling, and applications including Bayesian nonparametrics and hierarchical latent variable models, probabilistic computing, variational inference, deep learning, graph theory and algorithms, population genomics, cancer genomics, transcriptomics, disease genetics, and medical informatics.

Academic Experience

UNIVERSITY OF CONNECTICUT, Computer Science and Engineering Department, Storrs, CT 2018-
Assistant Professor

PRINCETON UNIVERSITY, Department of Computer Science, Princeton, NJ 2014-2018
Postdoctoral Researcher, Advised by Barbara Engelhardt

BROWN UNIVERSITY, Department of Computer Science, Providence, RI 2008-2014
Ph.D., Computer Science, Advised by Sorin Istrail, 2014
Master of Science, Computer Science, 2010

UNIVERSITY OF RHODE ISLAND (URI), Department of Computer Science, Kingston, RI 2003-2007
Bachelor of Science, Computer Engineering, 2007
Bachelor of Science, Computer Science, 2007
summa cum laude, President's Awards for Excellence (CE & CS)

Teaching Experience

UNIVERSITY OF CONNECTICUT, Storrs, CT 2018-2019
Instructor, *Algorithms and Complexity*: 2019

- ▶ Taught CSE3500 – Algorithms and Complexity, which entailed creating lesson plans, syllabus, problem sets, programming assignments, tests, in-class demonstrations (FFT, GPU programming, optimization algorithms), honors projects, and worksheets.
- ▶ Topics included: asymptotic notations, solving summations, recurrences, invariants, proof techniques, dynamic programming, divide-and-conquer, data structures, probabilistic algorithms, competitive analysis, greedy algorithms, graph algorithms, intractability and NP-completeness, parallel computing, approximation algorithms, and optimization algorithms.

Instructor, *Bayesian Machine Learning*: 2018

- ▶ Developed the course CSE5095 – Bayesian Machine Learning, which entailed creating lesson plans, a syllabus, and project assignments.
- ▶ We covered the three fundamental components of Bayesian Machine Learning: probabilistic modeling, inference algorithms, and model checking.

- ▶ The course culminates in research-quality projects related to each student's research interests that implements some probabilistic framework from problem specification to model checking.

BROWN UNIVERSITY, Providence, RI

2009-2014

Head Teacher's Assistant, *Algorithmic Foundations of Computational Biology*: 2011, 2013-2014

- ▶ Assisted teaching of *Algorithmic Foundations of Computational Biology* to classes of intellectually diverse undergraduate students from the Computer Science and Applied Mathematics departments.
- ▶ Developed and delivered a week of lectures on Haplotype Assembly. Administered online discussion platform to encourage student participation.
- ▶ Mentored students participating in interdisciplinary CYRENE project on cis-regulatory genomics.
- ▶ Co-developed the syllabus, created the course webpage, developed homework assignments and exams, curated lecture notes.

Head Teacher's Assistant, *Medical Bioinformatics*: 2009-2010, 2012-2013

- ▶ Co-created the *Medical Bioinformatics* course designed for computational and life sciences graduate students.
- ▶ Created the *Genomathica*, a set of interactive genomics algorithms written in Mathematica, and used *Genomathica* in lectures and homeworks enabling life science students to tinker and develop data analysis software pipelines without the burden of learning a programming language.
- ▶ Prepared and presented a week of lectures on haplotype phasing and genomic deletion inference in genome-wide association study data.
- ▶ Co-developed the syllabus, created the course webpage, developed homework assignments and exams, constructed lecture notes, and graded homework and exams.

Teacher's Assistant, *Algorithmic Foundations of Computational Biology*: 2010

- ▶ Taught *Algorithmic Foundations of Computational Biology* to an intellectually diverse class of undergraduate students from the Computer Science and Applied Mathematics departments.
- ▶ Maintained course webpage, graded homework assignments and exams, and curated lecture notes.

UNIVERSITY OF RHODE ISLAND, Kingston, RI

2007

Teacher's Assistant, *Computer Communications*: 2007

- ▶ Provided instruction for *Computer Communications*, a computer engineering and computer science senior undergraduate course.
- ▶ Independently researched the topics covered in the course with the objective of assisting in the redesign of the curriculum. The five lesson plans I created, including programming and homework assignments, incorporated: working at the frame level with Application, Transport, and Internet layer protocols (DNS, SMTP, TELNET, ARP, TCP/IP), and C Socket programming.

Preprints and Unpublished Work

1. **Aguiar, D.**, B. Dumitrascu, and B. E. Engelhardt (2018). *Robust and interpretable statistical genetic modelling. (manuscript in final stages)*.
2. **Aguiar, D.**, L. T. Elliott, Y. W. Teh, and B. E. Engelhardt (2018). *Haplotype phasing and imputation with haplotype cluster graphs. (manuscript in preparation)*.
3. **Aguiar, D.**, G. Gliner, L. T. Elliott, Y. W. Teh, and B. E. Engelhardt (2018). *The fragmentation coagulation process haplotype cluster graph model. (manuscript in final stages)*.

Ph.D. Thesis

1. Aguiar, D. (2014). "Genome-wide algorithms for haplotype assembly, haplotype phasing, and IBD inference". PhD thesis. Brown University.

Refereed Papers

2. **Aguiar, D.**, L.-F. Cheng, B. Dumitrascu, F. Mordelet, A. Pai, and B. Engelhardt (2018). Bayesian nonparametric discovery of isoforms and individual specific quantification. *Nature Communications* **9**(1), 1681.
3. **Aguiar, D.**, E. M. Morrow, and S. Istrail (2014). Tractatus: An Exact and Subquadratic Algorithm for Inferring Identical-by-Descent Multi-shared Haplotype Tracts. In: *RECOMB*. Springer, pp.1–17.
4. **Aguiar, D.**, W. S. Wong, and S. Istrail (2014). Tumor haplotype assembly algorithms for cancer genomics. In: *Pacific Symposium on Biocomputing*. NIH Public Access, pp.3.
5. McDowell, I. C., C. Nikapitiya, **D. Aguiar**, C. E. Lane, S. Istrail, and M. Gomez-Chiarri (2014). Transcriptome of American oysters, *Crassostrea virginica*, in response to bacterial challenge: insights into potential mechanisms of disease resistance. *PLoS One* **9**(8), e105097.
6. **Aguiar, D.** and S. Istrail (2013). Haplotype assembly in polyploid genomes and identical by descent shared tracts. *Bioinformatics* **29**(13). *also in proceedings of ISMB 2013*, i352–i360.
7. Tulin^{*}, S., **D. Aguiar**^{*}, S. Istrail, and J. Smith (2013). A quantitative reference transcriptome for *Nematostella vectensis* early embryonic development: a pipeline for de novo assembly in emerging model systems. *EvoDevo* **4**(1), 16.
8. **Aguiar, D.**, B. V. Halldórsson, E. M. Morrow, and S. Istrail (2012). DELISHUS: an efficient and exact algorithm for genome-wide detection of deletion polymorphism in autism. *Bioinformatics* **28**(12). *also in proceedings of ISMB 2012*, i154–i162.
9. **Aguiar, D.** and S. Istrail (2012). HapCompass: a fast cycle basis algorithm for accurate haplotype assembly of sequence data. *Journal of Computational Biology* **19**(6), 577–590.
10. Halldórsson^{*}, B. V., **D. Aguiar**^{*}, and S. Istrail (2011). "Haplotype phasing by multi-assembly of shared haplotypes: phase-dependent interactions between rare variants". In: *Pacific Symposium on Biocomputing*, pp.88–99.
11. Halldórsson^{*}, B. V., **D. Aguiar**^{*}, R. Tarpine, and S. Istrail (2011). The Clark Phaseable sample size problem: long-range phasing and loss of heterozygosity in GWAS. *Journal of Computational Biology* **18**(3). *also in proceedings of RECOMB 2010*, 323–333.

Book Chapters

12. Istrail, S., R. Tarpine, K. Schutter, and **D. Aguiar** (2010). "Practical Computational Methods for Regulatory Genomics: A cisGRN-Lexicon and cisGRN-Browser for Gene Regulatory Networks". In: *Computational Biology of Transcription Factor Binding*. Totowa, NJ: Humana Press, pp. 369–399.

^{*} denotes co-first authorship when ambiguous

Invited Talks

1. New York Area Population Genomics Workshop, "Efficient construction of haplotype cluster graphs and estimating the age of mutations." January 2017
2. American Society of Human Genetics (ASHG) Annual Meeting, "Haplotype phasing using cluster graphs." October 2015
3. NIPS Workshop on Machine Learning in Computational Biology (MLCB) "A Bayesian nonparametric statistical framework for haplotype phasing." December 2014
4. 18th Annual International Conference on Research in Computational Molecular Biology "Tractatus: an exact and subquadratic algorithm for inferring identical-by-descent multi-shared haplotype tracts." April 2014
5. 19th Pacific Symposium on Biocomputing "Tumor Haplotype Assembly Algorithms for Cancer Genomics." January 2014
6. Brown University IPP Symposium: Putting Big Data to Work. "Ome sweet ome: the genome as a model for big data." April 25, 2013
7. 21st Annual International Conference on Intelligent Systems for Molecular Biology "Haplotype assembly in polyploid genomes and identical by descent shared tracts." July 2013
8. HiT-Seq workshop in the 21st Annual International Conference on Intelligent Systems for Molecular Biology "Haplotype assembly in the presence of hemizyosity, haplotype sharing, polyploidy, and viral quasispecies." July 2013
9. 20th Annual International Conference on Intelligent Systems for Molecular Biology "DELISHUS: An efficient and exact algorithm for Genome-Wide detection of deletion polymorphism in autism." July 2012
10. Brown Computational Biology Open House "SNPs and haplotypes and GWAS oh my!" Feb. 27, 2012
11. Second Annual IEEE ICCABS CANGS Workshop "Robust algorithms for inferring haplotype phase and deletion polymorphism from high throughput whole genome sequence data." Feb. 24, 2012
12. Brown Computer Science - Research Exchange Seminars with Tea (REST) "Computational Challenges in Genome-wide Association Studies." Nov. 30, 2010
13. Fourteenth International Conference on Research in Computational Molecular Biology "The Clark Phase-able Sample Size Problem Long-range Phasing and Loss of Heterozygosity in GWAS." August 12, 2010

Extended Abstracts and Posters

1. Derek Aguiar, Li-Fang Cheng, Bianca Dumitrascu, Fantine Mordelet, Barbara E. Engelhardt, "BIISQ: Bayesian nonparametric discovery of Isoforms and Individual Specific Quantification from RNA-seq data." RECOMB 2016, Santa Monica, CA.
2. Derek Aguiar, Li-Fang Cheng, Bianca Dumitrascu, Fantine Mordelet, Barbara E. Engelhardt, "BIISQ: Bayesian nonparametric discovery of Isoforms and Individual Specific Quantification from RNA-seq data." Probabilistic Modeling in Genomics 2016, Oxford, UK.
3. Derek Aguiar, Li-Fang Cheng, Bianca Dumitrascu, Fantine Mordelet, Athma Pai, Barbara E. Engelhardt, "Bayesian nonparametric isoform discovery and quantification of many RNA-seq samples." American Society of Human Genetics Annual Meeting 2016, Baltimore, MD.

4. Derek Aguiar, Lloyd T. Elliott, Yee Whye Teh, and Barbara E. Engelhardt, "Haplotype phasing using cluster graphs." American Society of Human Genetics Annual Meeting 2015, Baltimore, MD.
5. Derek Aguiar, Lloyd T. Elliott, Yee Whye Teh, and Barbara E. Engelhardt, "Tractable haplotype phasing and imputation with nonparametric fragmentation-coagulation processes." Probabilistic Modeling in Genomics 2015, Cold Spring Harbor Laboratory, NY.
6. Derek Aguiar, Lloyd T. Elliott, Yee Whye Teh, and Barbara E. Engelhardt, "A statistical framework for modeling genetic data as haplotype cluster graphs with application to haplotype phasing, association mapping, and whole-genome compression." The Biology of Genomes 2015, Cold Spring Harbor Laboratory, NY.
7. Derek Aguiar, Lloyd T. Elliott, Yee Whye Teh, and Barbara E. Engelhardt, "A Bayesian nonparametric statistical framework for haplotype phasing." Workshop: Machine Learning in Computational Biology in NIPS 2014, Montreal, Canada. (selected for oral presentation)
8. Derek Aguiar, Austin Huang, Rami Kantor, Eric Morrow and Sorin Istrail, "Haplotype assembly in the presence of hemizyosity, haplotype sharing, polyploidy, and viral quasispecies." HiT-Seq workshop in the 21st Annual International Conference on Intelligent Systems for Molecular Biology, July 2013, Berlin, Germany. (selected for oral presentation)
9. Sarah Tulin, Derek Aguiar, Sorin Istrail, and Joel Smith, "Nematostella reference transcriptome and high throughput gene regulatory network construction." SDB 71st Annual Meeting, July 2012, Montreal, Canada.
10. Derek Aguiar and Sorin Istrail, "HAPCOMPASS: A fast cycle basis algorithm for accurate haplotype assembly of next-generation sequence data." 20th Annual Intelligent Systems for Molecular Biology, July 2012, Long Beach, CA.
11. Derek Aguiar, Ryan Tarpine, Fumei Lam, Bjarni V. Halldorsson, Eric Morrow, and Sorin Istrail, "Long-Range Haplotype Phasing by Multi-Assembly of Shared Haplotypes: Phase-Dependent Interactions Between Rare Variants." Genomics of Common Disease, Wellcome Trust Sanger Genome Center, Cambridge UK, December 2011
12. Derek Aguiar, Ryan Tarpine, Fumei Lam, Bjarni V. Halldorsson, Eric Morrow, and Sorin Istrail, "Long-Range Haplotype Phasing by Multi-Assembly of Shared Haplotypes: Phase-Dependent Interactions Between Rare Variants." The Gordon Research Conference on Human Genetics and Genomics, July 17-22, 2011, Salve Regina University, Newport, RI.
13. Ryan Tarpine, James Hart, Timothy Johnstone, Derek Aguiar, Sorin Istrail, "Report on the Cyrene Project: A cis-Lexicon Containing the Regulatory Architecture of 557 Regulatory Genes Experimentally Validated Using the *Davidson Criteria*." The Developmental Biology of the Sea Urchin Meeting, April 27-30, 2011, Woods Hole, MA.
14. Derek Aguiar, Ryan Tarpine, Eric Ruggieri, Jessica Nadel, David Moskowitz, Sorin Istrail, "Beyond GWAS: Robust Computational Analysis of the Multiple Sclerosis Genetic Consortium Data." Fourth Annual Center for Computational Biology Poster Session, April 28, 2010, Brown University, RI.

Scholarships & Honors

Charles J. Epstein Trainee Awards for Excellence in Human Genetics Research semifinalist, 2015
Research in Computational Molecular Biology (RECOMB) Student Travel Fellowship, 2010, 2014
Intelligent Systems for Molecular Biology (ISMB) Student Travel Fellowship, 2012, 2013
National Science Foundation (NSF) Established Program to Stimulate Competitive Research (EPSCoR) Academy Travel Award, 2013

IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) Computational Advances for Next Generation Sequencing (CANGS) Workshop Student Travel Award, 2012
 Outstanding Graduating Senior in Computer Engineering, URI, 2007
 Outstanding Junior in Computer Engineering, URI, 2006

Software

1. **Aguiar, D.**, L.-F. Cheng, B. Dumitrascu, and B. Engelhardt (2017). *Bayesian nonparametric discovery of Isoforms and Individual Specific Quantification with BIIISQ*. Version 1.0.0. [\[code\]](#).
2. **Aguiar, D.** and S. Istrail (2016). *HapCompass: fast cycle basis algorithms for accurate diploid and polyploid haplotype assembly*. Version 0.8.2. [\[code\]](#).
3. **Aguiar, D.** and S. Istrail (2013). *Tractatus: an exact and subquadratic algorithm for inferring identity-by-descent multi-shared haplotype tracts*. Version 0.0.1. [\[code\]](#).
4. **Aguiar, D.**, S. Tulin, S. Istrail, and J. Smith (2013). *A quantitative reference transcriptome for Nematostella vectensis earlyembryonic development: a pipeline for de novo assembly in emergingmodel systems*. Version 1.0.0. [\[code\]](#).
5. **Aguiar, D.** and S. Istrail (2012). *DELISHUS: an efficient and exact algorithm for genome-wide detection of deletion polymorphism in autism*. Version 0.1.0. [\[code\]](#).
6. Tarpine, R., **D. Aguiar**, and S. Istrail (2012). *CYRENE cis-Regulatory Genome Browser*. Version 1.0.0. [\[code\]](#).

Community Outreach

Explore Engineering Young Engineering Science Scholars Program, School of Engineering, UNIVERSITY OF CONNECTICUT, Storrs, CT 2019

- [Explore Engineering](#) is a one-week STEM summer camp for high school juniors and seniors where they are introduced to engineering fields.
- Students in my CS section were introduced to Programming, Data Science, Machine Learning and Artificial Intelligence through a series of lectures, activities, Jupyter notebooks, and programming exercises.
- On a scale of 1 (Unacceptable) to 5 (Outstanding), students rated the CS section an average score of 4.77.

Senior Design Project Judge, School of Engineering, UNIVERSITY OF CONNECTICUT, Storrs, CT 2019

- [Senior design projects](#) are the culmination of a two-semester course where students team up with faculty and industry sponsors to work collaboratively to solve a real-world problem.

Invention Convention Judge, Connecticut Invention Convention, Storrs, CT 2019

- The [Connecticut Invention Convention](#) is an internationally recognized nonprofit educational program that encourages interest in STEM fields and develops critical thinking skills in children in grades K-12 through invention and innovation.

AI4All Postdoctoral Mentor, Fragile Families group, PRINCETON UNIVERSITY, Princeton, NJ 2018

- [AI4All](#) is a high school summer camp aimed at popularizing AI among groups underrepresented in the field.

- Responsibilities include mentoring the Fragile Families project team; development of curriculum and Jupyter notebooks detailing machine learning applications to the Fragile Families data.

Volunteer for Kids Run Durham, Bull City Running Company, Durham, NC 2015-2016

- [Kids Run Durham](#) is a program meant to encourage building healthy lifestyle habits, self-esteem, and character development, goal setting, and management in young children.
- Responsible for registration of children into their age-appropriate race-series and organizing the children into their respective groups at the race starting line.

Academic Service

Advising

Undergraduate Advisor, Matthew Jane, Computational Neuroscience;

Committees

Program Committee, ISMB (2018), BNP Workshop at NIPS (2018), RECOMB (2019), BIBM (2019)

PhD Thesis Committee, Kingsley Udeh, Computer Science;

Departmental, Faculty Search Open 2018-2019, Systems and ML 2018-2019;

Peer-Review

Nature Communications, 2017; *Genome Research*, 2016; *Bioinformatics*, 2012-2017; *Proceedings of the National Academy of Sciences (PNAS)*, 2011; *Journal of Computational Biology (JCB)*, 2010-2017

Academic Web Administration and Development

- Built and maintained the [Princeton BEEHIVE](#), 2014-2018
- Built and maintained the [Istrail Lab Web](#); in 2012 rebuilt in PHP, 2009-2014
- Managed Brown University's Center for Computational Molecular Biology (CCMB) website, 2009-2011

Organizational Support

- Produced presentations and brochures for Brown University's CCMB Symposium, 2010
- Developed presentations and movies using Google Earth, Google SketchUp, and KML to support creation of the Ph.D. program for the Brown University's CCMB; Ph.D. program was approved in 2009

Other professional activities

- North Carolina Student Academy of Science State Competition Judge, 2016-2017

Industry Experience

RAYTHEON INTEGRATED DEFENSE SYSTEMS (IDS), Portsmouth, RI 2006-2008

Software Engineer, 2007-2008

- Developed and tested data control and management software for the Zumwalt Total Ship Computing Environment Infrastructure.

Software Engineer, Mission Innovation (MI) Group, 2006-2007

- Applied company technology and resources to address world-threatening issues.
- Used Google Earth, KML, SQL, and .NET to create web-based collaborative disaster planning tool (Joint Rapid Integrated Planning Service); listed as co-inventor on Raytheon patent.

Memberships & Activities

International Society for Computational Biology (ISCB), 2012-
Institute of Electrical and Electronics Engineers (IEEE), 2006-
Association for Computing Machinery (ACM), 2006-
American Association for the Advancement of Science (AAAS), 2019-
Six Sigma Specialist, 2007
Phi Eta Sigma Honor Society, 2006
Tau Beta Pi Honor Society, 2006