Siavash Mirarab

Contact Information

Department of Electrical and Computer Engineering
UC San Diego
9500 Gilman Drive, Mail code 0407
La Jolla, CA 92093-0407

phone: 858 822 6245
e-mail: `smirarab@ucsd.edu`
`http://eceweb.ucsd.edu/~smirarab/`
google scholar: `http://goo.gl/gEswQ5`

Academic Positions

UC San Diego, San Diego, CA, United States
Department of Electrical and Computer Engineering

Assistant Professor **July 2015 – present**

Education

Ph.D **Jan. 2011 – July 2015**
University of Texas at Austin, Austin, Texas, Department of Computer Science

- Advisors: Prof. Tandy Warnow, Prof. Keshav Pingali
- Novel scalable approaches for multiple sequence alignment and phylogenetic reconstruction

Master of Science **Sept. 2006 – Sept. 2008**
University of Waterloo, Waterloo, Ontario, Canada,
Department of Electrical and Computer Engineering

- Advisors: Prof. Ladan Tahvildari
- Thesis: A Bayesian Framework for Software Regression Testing

University of Tehran, Tehran, Iran
Department of Electrical and Computer Engineering

Bachelor of Science **Sept. 2001 – Dec. 2005**

- Computer engineering, specialized in software engineering.

Major Awards

Alfred P. Sloan Research Fellow. **2017**

Honorable Mention for the 2015 ACM Doctoral Dissertation Award. **2016**

Howard Hughes Medical Institute International Student Fellowship. **2012 - 2015**

National Science and Engineering Research Council of Canada (NSERC), Canada Graduate Scholarship (CGS-D) converted to PGS-D outside Canada. **2011- 2012**

Other Awards

University of Texas at Austin, Computer Science, Bert Kay Dissertation Award, 2015

University of Texas at Austin, College of Natural Sci., Dean's Excellence Award, 2011

University of Texas at Austin, Graduate Dean's Prestigious Fellowship Supplement Award, 2011, 2012, 2013, 2014

University of Texas at Austin, MCD fellowship (I declined to take CGS instead)

Research Funding

NSF-1565862: "CRII: III: Using genomic context to understand evolutionary ..." 2016
NIH-CFAR: "Accuracy of HIV transmission network reconstruction methods." 2016

PUBLICATIONS

1. Uyen Mai, Erfan Sayyari, and **Siavash Mirarab**. Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction. *GLBIO*, (accepted), 2017.
2. **Siavash Mirarab**. Phylogenomics: Constrained gene tree inference. *Nature Ecology & Evolution*, 1:0056, 2017.
3. Shubhanshu Shekhar, Sebastien Roch, and **Siavash Mirarab**. Species tree estimation using ASTRAL: how many genes are enough? *Proceedings of International Conference on Research in Computational Molecular Biology (RECOMB)*, (to appear), 2017.
4. Erfan Sayyari and **Siavash Mirarab**. Anchoring quartet-based phylogenetic distances and applications to species tree reconstruction. *BMC Genomics*, 17(S10):101–113, 2016.
5. Erfan Sayyari and **Siavash Mirarab**. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular Biology and Evolution*, 33(7):1654–1668, 2016.
6. Nam Nguyen, Michael Nute, **Siavash Mirarab**, and Tandy Warnow. HIPPI: Highly Accurate Protein Family Classification with Ensembles of HMMs. *BMC Genomics*, in press, 2016.
7. James E Tarver, Mario Dos Reis, **Siavash Mirarab**, Raymond J Moran, Sean Parker, Joseph E O'Reilly, Benjamin L King, Mary J O'Connell, Robert J Asher, Tandy Warnow, Kevin J Peterson, Philip C J Donoghue, and Davide Pisani. The Interrelationships of Placental Mammals and the Limits of Phylogenetic Inference. *Genome biology and evolution*, page evv261, 2016.
8. **Siavash Mirarab**, Md. Shamsuzzoha Bayzid, Bastien Boussau, and Tandy Warnow. Response to Comment on "Statistical binning enables an accurate coalescent-based estimation of the avian tree". *Science*, 350(6257):171, 2015.
9. Joel Cracraft, Peter Houde, Simon Y W Ho, David P Mindell, Jon Fjeldså, Bent Lindow, Scott V Edwards, Carsten Rahbek, **Siavash Mirarab**, Tandy Warnow, M Thomas P Gilbert, Guojie Zhang, Edward L Braun, and Erich D Jarvis. Response to Comment on "Whole-genome analyses resolve early branches in the tree of life of modern birds". *Science*, 349(6255):1460, 2015.
10. Ruth Davidson, Pranjal Vachaspati, **Siavash Mirarab**, and Tandy Warnow. Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer. *BMC Genomics*, 16(Suppl 10):S1, 2015.
11. Jed Chou, Ashu Gupta, Shashank Yaduvanshi, Ruth Davidson, Mike Nute, **Siavash Mirarab**, and Tandy Warnow. A comparative study of SVDquartets

* co-first authorship

and other coalescent-based species tree estimation methods. *BMC Genomics*, 16(Suppl 10):S2, 2015.

12. **Siavash Mirarab** and Tandy Warnow. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31(12):i44–i52, 2015.

13. Nam Nguyen, **Siavash Mirarab**, Keerthana Kumar, and Tandy Warnow. Ultra-large alignments using phylogeny-aware profiles. *Genome Biology*, 16(1):124, 2015.

14. Md Shamsuzzoha Bayzid, **Siavash Mirarab**, Bastien Boussau, and Tandy Warnow. Weighted Statistical Binning: Enabling Statistically Consistent Genome-Scale Phylogenetic Analyses. *PLoS ONE*, 10(6):e0129183, 2015.

15. Erich D Jarvis, **Siavash Mirarab**, Andre J Aberer, Bo Li, Peter Houde, Cai Li, Simon Y W Ho, Brant C Faircloth, Benoit Nabholz, and Jason T Howard. Phylogenomic analyses data of the avian phylogenomics project. *GigaScience*, 4(1):4, 2015.

16. Erich D Jarvis, Siavash Mirarab*, 100 other authors, Edward L Braun, Tandy Warnow, Wang Jun, M Thomas P Gilbert, and Guojie Zhang. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346(6215):1320–1331, 2014.

17. **Siavash Mirarab**, Md. Shamsuzzoha Bayzid, Bastien Boussau, and Tandy Warnow. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science*, 346(6215):1250463, 2014.

18. Naim Matasci, Ling-Hong Hung, Zhixiang Yan, Eric J Carpenter, Norman J Wickett, **Siavash Mirarab**, Nam Nguyen, Tandy Warnow, Saravanaraj Ayyampalayam, and Michael Barker. Data access for the 1,000 Plants (1KP) project. *GigaScience*, 3(1):17, 2014.

19. Norman J. Wickett, Siavash Mirarab*, Nam Nguyen, Tandy Warnow, Eric Carpenter, Naim Matasci, Saravanaraj Ayyampalayam, Michael Barker, J. Gordon Burleigh, Matthew A. Gitzendanner, Brad R. Ruhfel, Eric Wafula, Joshua P. Der, Sean W. Graham, Sarah Mathews, Michael Melkonian, Douglas E. Soltish, Pamela S. Soltish, Nicholas W. Miles, Carl J. Rothfels, Lisa Pokorny, A. Jonathan Shaw, Lisa DeGironimo, Dennis W. Stevenson, Barbara Surek, Juan Carlos Villarreal, Béatrice Rourev, Hervé Philippe, Claude W. dePamphilis, Tao Chen, Michael K. Deyholos, Regina S. Baucom, Toni M. Kutchan, Megan M. Augustin, Jun Wang, Yong Zhang, Zhijian Tian, Zhixiang Yan, Xiaolei Wu, Xiao Sun, Gane Ka-Shu Wong, and James Leebens-Mack. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences (PNAS)*, 111(45):E4859–4868, 2014.

20. **Siavash Mirarab**, Md. Shamsuzzoha Bayzid, and Tandy Warnow. Evaluating Summary Methods for Multilocus Species Tree Estimation in the Presence of Incomplete Lineage Sorting. *Systematic Biology*, 65(3):366–380, 2014.
21. **Siavash Mirarab**, Nam Nguyen, Sheng Guo, Li-San Wang, Junhyong Kim, and Tandy Warnow. PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences. *Journal of Computational Biology*, 22(05):377–386, 2015.
22. **Siavash Mirarab**, Nam Nguyen, and Tandy Warnow. PASTA: ultra-large multiple sequence alignment. In *Proceedings of the 18th International Conference on Research in Computational Molecular Biology (RECOMB)*, pages 177–191, 2014.
23. **Siavash Mirarab**, Rezwana Reaz, Md. Shamsuzzoha Bayzid, Théo Zimmermann, M Shel Swenson, and Tandy Warnow. ASTRAL: Genome-Scale Coalescent-Based Species Tree. *Bioinformatics (ECCB)*, 30(17):i541–i548, 2014.
24. Nam Nguyen, **Siavash Mirarab**, Bo Liu, Mihai Pop, and Tandy Warnow. TIPP: Taxonomic Identification and Phylogenetic Profiling. *Bioinformatics*, 30(24):3548–3555, 2014.
25. Théo Zimmermann, **Siavash Mirarab**, and Tandy Warnow. BBCA: improving the scalability of *BEAST using random binning. *BMC Genomics*, 15(Suppl 6):S11, 2014.
26. Arlin Stoltzfus, Hilmar Lapp, Naim Matasci, Helena Deus, Brian Sidlauskas, Christian M Zmasek, Gaurav Vaidya, Enrico Pontelli, Karen Cranston, Rutger Vos, Campbell O Webb, Luke J Harmon, Megan Pirrung, Brian O'Meara, Matthew W Pennell, **Siavash Mirarab**, Michael S Rosenberg, James P Balhoff, Holly M Bik, Tracy A Heath, Peter E Midford, Joseph W Brown, Emily Jane McTavish, Jeet Sukumaran, Mark Westneat, Michael E Alfaro, Aaron Steele, and Greg Jordan. Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient. *BMC bioinformatics*, 14(1):158, 2013.
27. Md. Shamsuzzoha Bayzid, **Siavash Mirarab**, and Tandy Warnow. Inferring optimal species trees under gene duplication and loss. In *Proceedings of the Pacific Symposium on Biocomputing (PSB)*, pages 250–61, 2013.
28. Nam Nguyen, **Siavash Mirarab**, and Tandy Warnow. MRL and SuperFine+ MRL: new supertree methods. *Algorithms for Molecular Biology*, 7(1):3, 2012.
29. **Siavash Mirarab**, Nam Nguyen, and Tandy Warnow. SEPP: SATé-Enabled Phylogenetic Placement. In *Proceedings of the Pacific Symposium on Biocomputing (PSB)*, pages 247–58, 2012.

30. **Siavash Mirarab**, Soroush Akhlaghi, and Ladan Tahvildari. Size-constrained regression test case selection using multicriteria optimization. *IEEE Transactions on Software Engineering*, 38(4):936–956, 2012.
31. **Siavash Mirarab** and Tandy Warnow. FastSP: linear time calculation of alignment accuracy. *Bioinformatics*, 27:3250–8, 2011.
32. Hyunsook Do, **Siavash Mirarab**, Ladan Tahvildari, and Gregg Rothermel. The effects of time constraints on test case prioritization: A series of controlled experiments. *IEEE Transactions on Software Engineering*, 36(5):593–617, 2010.
33. **Siavash Mirarab** and Ladan Tahvildari. An empirical study on bayesian network-based approach for test case prioritization. In *Proceedings of the 1st International Conference on Software Testing, Verification, and Validation (ICST)*, pages 278–287. IEEE, 2008.
34. **Siavash Mirarab**, Afshar Ganjali, Ladan Tahvildari, Shimin Li, Weining Liu, and Mike Morrissey. A requirement-based software testing framework: An industrial practice. In *Proceedings of the 24th International Conference on Software Maintenance (ICSM)*, pages 452–455. IEEE, 2008.
35. Hyunsook Do, **Siavash Mirarab**, Ladan Tahvildari, and Gregg Rothermel. An empirical study of the effect of time constraints on the cost-benefits of regression testing. In *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE)*, pages 71–82. ACM, 2008.
36. Mehdi Amoui, Mazeiar Salehie, **Siavash Mirarab**, and Ladan Tahvildari. Adaptive action selection in autonomic software using reinforcement learning. In *Proceedings of the Fourth International Conference on Autonomic and Autonomous Systems (ICAS)*, pages 175–181. IEEE, 2008.
37. **Siavash Mirarab** and Ladan Tahvildari. A prioritization approach for software test cases based on bayesian networks. In *Fundamental Approaches to Software Engineering (FASE)*, pages 276–290. Springer, 2007.
38. **Siavash Mirarab**, Alaa Hassouna, and Ladan Tahvildari. Using bayesian belief networks to predict change propagation in software systems. In *Proceedings of the 15th International Conference on Program Comprehension (ICPC)*, pages 177–188. IEEE, 2007.
39. Mehdi Amoui, **Siavash Mirarab**, Sepand Ansari, and Caro Lucas. A genetic algorithm approach to design evolution using design pattern transformation. *Journal of Information Technology and Intelligent Computing*, 1(2):235–244, 2006.

Software (Publicly available)

ASTRAL (`https://github.com/smirarab/ASTRAL`)
Coalescent-based species tree estimation from gene trees

PASTA (`https://github.com/smirarab/pasta`)

Multiple sequence alignment of ultra-large datasets (up to a million sequences)

SEPP,TIPP (`https://github.com/smirarab/sepp`)
Phylogenetic placement and taxonomic identification of metagenomic reads

UPP (`https://github.com/smirarab/sepp/blob/master/README.UPP.md`)
Multiple sequences alignment using families of HMMs for ultra-large datasets

FastSP (`https://github.com/smirarab/FastSP`)
Fast comparison of very large multiple sequence alignments

See `https://github.com/smirarab/?tab=repositories` for more.

Professional Experience

International Business Machines (IBM), Vancouver, British Columbia, Canada

IT Specialist **Sept. 2008 – Dec. 2010**

Worked on a software platform for implementing implantable wireless pace-maker devices for a large US-based biomedical company.

Teaching

Instructor (University of California, San Diego):

Statistical learning in bioinformatics. **Winter 2016**

Algorithms for biological data analysis. **Winter 2016**

Introduction to Computer Engineering **Winter, Spring 2016, Winter 2017**

Teaching Assistance:

Software Testing and Quality Assurance, University of Waterloo. **Winter 2007**

Scientific & Technical Communication, University of Tehran. **Fall 2005**

Workshops, Tutorials, and Short Courses:

Ann Arbor, Michigan, Phylogenomics Symposium and Software School **2015**

UT Austin, Summer school on Phylogenomics and Metagenomics **Summer 2014**

Evolution meetings (NC), Software School on ASTRAL and PASTA **Summer 2014**

UT Austin, Workshop on New Methods for Phylogeny and Alignment Estimation, Tutorials on SEPP and SATe **Winter 2013**

Highlights and Invited Talks

Broad Inst., MA, "Taxonomic Profiling using Scalable Phylogenetic Placement" **2017**

RECOMB, Warsaw, Poland, "Statistical binning enables an accurate ..." **2015**

IPAM, UCLA, "Ultra-large multiple sequence alignments". **2015**

INFORMS, Philadelphia, "Reconstruction of species trees using genomic data". **2015**

MolPhy-3, Moscow, Russia, "SEPP and TIPP: phylogenetic placement and taxon identification methods for metagenomic". **2012**

Professional Services

Program Committee: ISMB/ECCB **2015, 2017**

Paper Review: Nature Ecology and Evolution, GigaScience, Journal of ACM, IEEE Transaction on Computational Biology and Bioinformatics, Theoretical Population Genetics, Bioinformatics, BMC Bioinformatics, Journal of ACM, ISMB, RECOMB-CG, Transactions on Software Engineering, and ICPC **2007-2016**