Mukul S. Bansal

Contact Information

Department of Computer Science & Engineering
University of Connecticut
371 Fairfield Way, U-4155
Storrs, CT, USA

Phone: (+1) 860-486-2572
Email: mukul.bansal@uconn.edu
Webpage: https://mukul-bansal.uconn.edu/
Lab Website: https://compbio.engr.uconn.edu/

Research Interests

Computational biology and **bioinformatics**, with a focus on computational phylogenetics and comparative genomics; **Algorithms** and **combinatorial optimization** in general.

Education

Postdoctoral Associate, January 2011 – August 2013
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge, USA
Postdoc Advisors: Manolis Kellis and Eric J. Alm

Postdoctoral Fellow, July 2009 – December 2010
The Blavatnik School of Computer Science
Tel Aviv University, Tel Aviv, Israel
Postdoc Advisor: Ron Shamir

M.S. and Ph.D. in Computer Science, Fall 2004 – Summer 2009
Department of Computer Science
Iowa State University, Ames, Iowa, USA
Advisors: David Fernández-Baca and Oliver Eulenstein

B.Tech. in Computer Science, 2000 – 2004
Department of Computer Science and Engineering
International Institute of Information Technology, Hyderabad, India

Employment History

Associate Professor: *August 2019 – Present*; Department of Computer Science and Engineering, University of Connecticut, USA.

Assistant Professor: *August 2013 – August 2019*; Department of Computer Science and Engineering, University of Connecticut, USA.

Postdoctoral Associate: *January 2011 – August 2013*; Computational Biology Research Group, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA.

Postdoctoral Fellow: *July 2009 – December 2010*; Algorithms for Computational Genomics Group, The Blavatnik School of Computer Science, Tel Aviv University, Israel.

Visiting Graduate Student: *Summer 2007*; Michael Sanderson's Lab, Department of Ecology and Evolutionary Biology, University of Arizona, USA.

Research/Teaching Assistant: *Fall 2004 – Summer 2009*; Department of Computer Science, Iowa State University, USA.

Other Appointments and Affiliations

- **Faculty Member**: The Institute for Systems Genomics (ISG), University of Connecticut.
- **Faculty Affiliate**: The Center for Microbial Systems, Ecology and Evolution (CMSEE), University of Connecticut.

Research Funding

10/2022–9/2025 Co-Principal Investigator (PI: Ion Mandoiu), *Collaborative Research: Algorithms for scalable inference and phylodynamic analysis of tumor haplotypes using low-coverage single cell sequencing data.* $513,341, National Science Foundation, USA.

6/2022–12/2023 Principal Investigator, *Algorithms for Genome-Scale Inference of Microbial Phylogenies.* $47,084, UConn Research Excellence Program.

2/2016–1/2023 Principal Investigator, *CAREER: Algorithms for Domain-Level Analysis of Gene Family Evolution.* $499,576, National Science Foundation, USA.

6/2017–6/2022 Co-Principal Investigator (PI: C.E. Nelson and co-PIs: J.D.Gibson, D.J. Goldhamer, O. Harel, H.Y. Bar, I. Mandoiu, K. Chen, D. Pejril), *A Catalog of Cell Types in the Early Organogenesis Embryo - A Single Cell Lineage Map.* $675,000, Connecticut Innovations Regenerative Medicine Research Fund (RMRF).

8/2016–7/2020 Principal Investigator, *Understanding Horizontal Gene Transfer in Bacteria and Archaea: Units of Transfer and Modes of Integration.* $598,801, National Science Foundation, USA.

8/2016–7/2020 Co-Principal Investigator (PI: Ion Mandoiu), *Collaborative Research: Algorithmic Techniques for Inferring Transmission Networks from Noisy Sequencing Data.* $199,999, National Science Foundation, USA.

9/2016–8/2020 Principal Investigator, *Collaborative Research: Integrating the geological and genomic records: time-calibrating Earth's dynamic biogeochemical history.* $316,714, National Science Foundation, USA.

Other External Funding

10/2018–09/2023 Key Personnel (Project Director: Sanguthevar Rajasekaran), *Graduate Assistance in Areas of National Need (GAANN) Program in Artificial Intelligence.* $951,576, US Department of Education.

Selected Recent Awards and Honors

7/2021 Best paper award at the 32nd Annual Symposium on Combinatorial Pattern Matching (CPM 2021)

9/2018 Best paper award at the Nineth ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2018)

2015, 2016, 2017 Received multiple UConn provost's commendations for teaching excellence

7/2016 Named *UTC professor of engineering innovation* for research excellence (three-year term)

2/2016 Recipient of NSF CAREER award

Journal Publications

(1) S. Weiner and M. S. Bansal, "CNAsim: Improved simulation of single-cell copy number profiles and DNA-seq data from tumors", *Bioinformatics*; in press.

(2) A. Mondal and M. S. Bansal, "Generalizing the Domain-Gene-Species Reconciliation Framework to Microbial Genes and Domains", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*; in press.

(3) A. Mondal, L. T. Rangel, J. G. Payette, G. P. Fournier, and M. S. Bansal, "DaTeR: Error-Correcting Phylogenetic Chronograms Using Relative Time Constraints", *Bioinformatics* 2023, 39(2): btad084.

(4) S. Zaman, S. Sledzieski, B. Berger, Y. Wu, and M. S. Bansal, "virDTL: Viral recombination analysis through phylogenetic reconciliation and its application to sarbecoviruses and SARS-CoV-2", *Journal of Computational Biology* 2023, 30(1): 3–20.

(5) S. Dhar, C. Zhang, I. Mandoiu, and M. S. Bansal, "TNet: Transmission Network Inference Using Within-Host Strain Diversity and its Application to Geographical Tracking of COVID-19 Spread", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2022, 19(1): 230–242.

(6) S. Weiner and M. S. Bansal, "Improved Duplication-Transfer-Loss Reconciliation with Extinct and Unsampled Lineages", *Algorithms* 2021, 14(8): 231.

(7) L. Kloub, S. Gosselin, M. Fullmer, J. Graf, J. P. Gogarten, and M. S. Bansal, "Systematic Detection of Large-Scale Multi-Gene Horizontal Transfer in Prokaryotes", *Molecular Biology and Evolution* 2021, 38(6): 2639–2659.

(8) T. Wade, L. T. Rangel, S. Kundu, G. P. Fournier, and M. S. Bansal, "Assessing the Accuracy of Phylogenetic Rooting Methods on Prokaryotic Gene Families", *PLOS One* 2020, 15(5): e0232950.

(9) M. S. Bansal, "Linear-Time Algorithms for Phylogenetic Tree Completion Under Robinson-Foulds Distance", *Algorithms for Molecular Biology* 2020, 15:6.

(10) S. Kundu and M. S. Bansal, "SaGePhy: An improved phylogenetic simulation framework for gene and subgene evolution ", *Bioinformatics* 2019, 35(18): 3496–3498.

(11) L. Li and M. S. Bansal, "An Integrated Reconciliation Framework for Domain, Gene, and Species Level Evolution", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2019, 16(1): 63–76.

(12) M. Kordi and M. S. Bansal, "Exact Algorithms for Duplication-Transfer-Loss Reconciliation with Non-Binary Gene Trees", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2019, 16(4): 1077–1090.

(13) W. Duchemin, G. Gence, A. Arigon-Chiffoleau, L. Arvestad, M. S. Bansal, V. Berry, B. Boussau, F. Chevenet, N. Comte, A. A. Davin, C. Dessimoz, D. Dylus, D. Hasic, D. Mallo, R. Planel, D. Posada, C. Scornavacca, G. Szollosi, L. Zhang, E. Tannier, and V. Daubin "RecPhyloXML – a format for reconciled gene trees", *Bioinformatics* 2018, 34(21): 3646–3652.

(14) M. S. Bansal, M. Kellis, M. Kordi, and S. Kundu, "RANGER-DTL 2.0: Rigorous Reconstruction of Gene-Family Evolution by Duplication, Transfer, and Loss", *Bioinformatics* 2018, 34(18): 3214–3216.

(15) S. Kundu and M. S. Bansal, "On the Impact of Uncertain Gene Tree Rooting on Duplication-Transfer-Loss Reconciliation", *BMC Bioinformatics* 2018, 19(Suppl 9): 290.

(16) M. Kordi and M. S. Bansal, "On the Complexity of Duplication-Transfer-Loss Reconciliation with Non-Binary Gene Trees", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2017, 14(3): 587–599.

(17) Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungal, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. Farh, S. Feizi, R. Karlic, A. Kim, A. Kulkarni, D. Li, R. Lowdon, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopa, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. Boyer, P. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. Jones, W. Li, M. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson,

T. D. Tlsty, L. Tsai, W. Wang, R. A. Waterland, M. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis, "Integrative analysis of 111 reference human epigenomes", *Nature* 2015, 518: 317–330.

(18) M. S. Bansal, Y. Wu, E. J. Alm, and M. Kellis, "Improved Gene Tree Error Correction in the Presence of Horizontal Gene Transfer", *Bioinformatics* 2015, 31(8): 1211–1218.

(19) F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. Pope, Y. Shen, D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L. H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. L. de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. R. Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutyavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K. H. Chang, A. Skoultchi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X. Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren; Mouse ENCODE Consortium, "A Comparative Encyclopedia of DNA Elements in the Mouse Genome", *Nature* 2014, 515: 355–364.

(20) Y. Wu, M. D. Rasmussen, M. S. Bansal, and M. Kellis, "Most Parsimonious Reconciliation in the Presence of Gene Duplication, Loss, and Deep Coalescence Using Labeled Coalescent Trees", *Genome Research* 2014, 24: 475–486.

(21) M. S. Bansal, E. J. Alm, and M. Kellis, "Reconciliation Revisited: Handling Multiple Optima when Reconciling with Duplication, Transfer, and Loss", *Journal of Computational Biology* 2013, 20(10): 738–754.

(22) M. S. Bansal and O. Eulenstein, "Algorithms for Genome-Scale Phylogenetics Using Gene Tree Parsimony", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2013, 10(4): 939-956.

(23) M. S. Bansal, G. Banay, T. J. Harlow, J. P. Gogarten, and R. Shamir, "Systematic Inference of Highways of Horizontal Gene Transfer in Prokaryotes", *Bioinformatics* 2013, 29(5): 571–579.

(24) Y. Wu, M. D. Rasmussen, M. S. Bansal, and M. Kellis, "TreeFix: statistically informed gene tree error correction using species trees", *Systematic Biology* 2013, 62(1): 110–120.

(25) M. S. Bansal, G. Banay, J. P. Gogarten, and R. Shamir, "Detecting Highways of Horizontal Gene Transfer", *Journal of Computational Biology* 2011, 18(9): 1087–1114.

(26) M. S. Bansal, J. Dong, and D. Fernández-Baca, "Comparing and Aggregating Partially Resolved Trees", *Theoretical Computer Science* 2011, 412: 6634–6652.

(27) M. S. Bansal and R. Shamir, "A Note on the Fixed Parameter Tractability of the Gene-Duplication Problem", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2011, 8(3): 848–850.

(28) J. G. Burleigh, M. S. Bansal, O. Eulenstein, S. Hartmann, A. Wehe, and T. J. Vision, "Genome-Scale Phylogenetics: Inferring the Plant Tree of Life from 18,896 Gene Trees", *Systematic Biology* 2011, 60(2): 117–125.

(29) R. Chaudhary, M. S. Bansal, A. Wehe, D. Fernández-Baca, and O. Eulenstein, "iGTP: A software package for large-scale gene tree parsimony analysis", *BMC Bioinformatics* 2010, 11:574.

(30) M. S. Bansal, J. G. Burleigh, O. Eulenstein, and D. Fernández-Baca, "Robinson-Foulds Supertrees", *Algorithms for Molecular Biology* 2010, 5:18.

(31) M. S. Bansal, O. Eulenstein, and A. Wehe, "The Gene-Duplication Problem: Near-Linear Time Algorithms for NNI Based Local Searches", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2009, 6(2): 221–231.

(32) J. G. Burleigh, M. S. Bansal, A. Wehe, and O. Eulenstein, "Locating Large-Scale Gene Duplication Events Through Reconciled Trees: Implications For Identifying Ancient Polyploidy Events In Plants", *Journal of Computational Biology* 2009, 16(8): 1071–1083.

(33) M. S. Bansal and D. Fernández-Baca, "Computing Distances Between Partial Rankings", *Information Processing Letters* 2009, 109(4): 238–241.

(34) M. S. Bansal and O. Eulenstein, "An $\Omega(n^2/\log n)$ Speed-Up of TBR Heuristics for the Gene-Duplication Problem", *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2008, 5(4): 514–524.

(35) A. Wehe, M. S. Bansal, J. G. Burleigh, and O. Eulenstein, "DupTree: A program for large-scale phylogenetic analyses using gene tree parsimony", *Bioinformatics* 2008, 24(13): 1540–1541.

(36) D. Chen, J. G. Burleigh, M. S. Bansal, and D. Fernández-Baca, "PhyloFinder: An Intelligent Search Engine for Phylogenetic Tree Databases", *BMC Evolutionary Biology* 2008, 8:90.

REFEREED CONFERENCE PUBLICATIONS
   (Including conference proceedings published in journal special issues, not listed above.)

(37) S. Zaman and M. S. Bansal "On Partial Gene Transfer and its impact on Gene Tree Reconstruction", *Nineteenth Annual RECOMB Comparative Genomics Conference (***RECOMB-CG 2022***), LNBI 13234*: 168–186.

(38) K. Yao and M. S. Bansal, "Optimal Completion and Comparison of Incomplete Phylogenetic Trees Under Robinson-Foulds Distance", *Combinatorial Pattern Matching (***CPM 2021***), LIPIcs, Vol. 191*: Article No. 25, 1–23. This paper won the **Best Student Paper Award**.

(39) S. Sledzieski, C. Zhang, I. Mandoiu, and M. S. Bansal, "TreeFix-TP: Phylogenetic Error-Correction for Infectious Disease Transmission Network Inference", *Pacific Symposium on Biocomputing (***PSB 2021***); Proceedings:* 119–130.

(40) A. Mondal, M. Kordi, and M. S. Bansal, "A Supervised Machine Learning Approach for Distinguishing Between Additive and Replacing Horizontal Gene Transfers", *Eleventh ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (***ACM-BCB 2020***); Proceedings:* Article No. 16, 1–11.

(41) S. Dhar, C. Zhang, I. Mandoiu, and M. S. Bansal, "TNet: Phylogeny-Based Inference of Disease Transmission Networks Using Within-Host Strain Diversity", *Sixteenth International Symposium on Bioinformatics Research and Applications (***ISBRA 2020***), LNCS 12304*: 203–216.

(42) M. Kordi and M. S. Bansal, "TreeSolve: Rapid Error-Correction of Microbial Gene Trees", *7th International Conference on Algorithms for Computational Biology (***AlCoB 2020***), LNCS 12099*: 125–139.

(43) M. Kordi, S. Kundu, and M. S. Bansal, "On Inferring Additive and Replacing Horizontal Gene Transfers Through Phylogenetic Reconciliation", *Tenth ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (***ACM-BCB 2019***); Proceedings:* 514–523.

(44) L. Li and M. S. Bansal, "Simultaneous Multi-Domain-Multi-Gene Reconciliation under the Domain-Gene-Species Reconciliation Model", *Fifteenth International Symposium on Bioinformatics Research and Applications (***ISBRA 2019***), LNCS 11490*: 73–86.

(45) M. S. Bansal, "Linear-Time Algorithms for Some Phylogenetic Tree Completion Problems under Robinson-Foulds Distance", *Sixteenth Annual RECOMB Comparative Genomics Conference (***RECOMB-CG 2018***), LNCS 11183*: 209–226.

(46) L. Li and M. S. Bansal "An Integer Linear Programming Solution for the Domain-Gene-Species Reconciliation Problem", *Nineth ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (**ACM-BCB 2018**); Proceedings:* 386–397. This paper won the **Best Student Paper Award**.

(47) S. Kundu and M. S. Bansal, "On the Impact of Uncertain Gene Tree Rooting on Duplication-Transfer-Loss Reconciliation", *Thirteenth International Symposium on Bioinformatics Research and Applications (**ISBRA 2017**), LNCS 10330:* L–LI. (short paper)

(48) M. Kordi and M. S. Bansal, "Exact Algorithms for Duplication-Transfer-Loss Reconciliation with Non-Binary Gene Trees", *Seventh ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (**ACM-BCB 2016**); Proceedings:* 297–306.

(49) Misagh Kordi and Mukul S. Bansal, "On the Complexity of Duplication-Transfer-Loss Reconciliation with Non-Binary Gene Trees", *Eleventh International Symposium on Bioinformatics Research and Applications (**ISBRA 2015**), LNCS 9096:* 187–198.

(50) R. Libeskind-Hadas, Y. Wu, M. S. Bansal, and M. Kellis, "Pareto-Optimal Phylogenetic Tree Reconciliation", *22ndAnnual International Conference on Intelligent Systems for Molecular Biology (**ISMB 2014**); Bioinformatics* 2014, 30: i87–i95.

(51) M. S. Bansal, E. J. Alm, and M. Kellis, "Reconciliation Revisited: Handling Multiple Optima when Reconciling with Duplication, Transfer, and Loss", *Seventeenth Annual International Conference on Research in Computational Molecular Biology (**RECOMB 2013**), LNCS 7821:* 1–13.

(52) M. S. Bansal, E. J. Alm, and M. Kellis, "Efficient Algorithms for the Reconciliation Problem with Gene Duplication, Horizontal Transfer, and Loss", *Twentieth Annual International Conference on Intelligent Systems for Molecular Biology (**ISMB 2012**); Bioinformatics* 2012, 28: i283–i291.

(53) M. S. Bansal, J. P. Gogarten, and R. Shamir, "Detecting Highways of Horizontal Gene Transfer", *Eighth Annual RECOMB Comparative Genomics Workshop (**RECOMB-CG 2010**), LNCS 6398:* 109–120.

(54) H. Narayanappa, M. S. Bansal, and H. Rajan, "Property-Aware Program Sampling", *Ninth ACM SIGPLAN-SIGSOFT Workshop on Program Analysis for Software Tools and Engineering (**PASTE 2010**):* 45–52.

(55) J. G. Burleigh, M. S. Bansal, O. Eulenstein, and T. J. Vision, "Inferring Species Trees From Gene Duplication Episodes". *First ACM International Conference on Bioinformatics and Computational Biology (**ACM-BCB 2010**):* 198–203.

(56) M. S. Bansal, J. G. Burleigh, and O. Eulenstein, "Efficient genome-scale phylogenetic analysis under the duplication-Loss and deep coalescence cost models", *Eighth Asia-Pacific Bioinformatics Conference (**APBC 2010**); BMC Bioinformatics* 2010, 11(Suppl 1): S42.

(57) M. S. Bansal, W. Chang, O. Eulenstein, and D. Fernández-Baca, "Generalized Binary Tanglegrams: Algorithms and Applications". *First International Conference on Bioinformatics and Computational Biology (**BICoB 2009**), LNCS 5462:* 114–125. This paper won the **Best Paper Award**.

(58) M. S. Bansal and O. Eulenstein, "The Multiple Gene Duplication Problem Revisited", *Sixteenth Annual International Conference on Intelligent Systems for Molecular Biology (**ISMB 2008**); Bioinformatics* 2008, 24(13): i132–i138.

(59) M. S. Bansal and O. Eulenstein, "The Gene-Duplication Problem: Near-Linear Time Algorithms for NNI Based Local Searches", *Fourth International Symposium on Bioinformatics Research and Applications (**ISBRA 2008**), LNCS 4983:* 14–25.

(60) J. G. Burleigh, M. S. Bansal, A. Wehe, and O. Eulenstein, "Locating Multiple Gene Duplications Through Reconciled Trees", *Twelfth Annual International Conference on Research in Computational Molecular Biology (**RECOMB 2008**), LNCS 4955:* 273–284.

(61) M. S. Bansal, J. Dong, and D. Fernández-Baca, "Comparing and Aggregating Partially Resolved Trees", *Eighth Latin American Theoretical Informatics Symposium (**LATIN 2008**), LNCS 4957:* 72–83.

(62) M. S. Bansal and O. Eulenstein, "An $\Omega(n^2/\log n)$ Speed-Up of TBR Heuristics for the Gene-Duplication Problem", *Seventh Workshop on Algorithms in Bioinformatics (**WABI 2007**), LNCS 4645*: 124–135.

(63) M. S. Bansal, J. G. Burleigh, O. Eulenstein, and A. Wehe, "Heuristics for the Gene-Duplication Problem: A $\Theta(n)$ Speed-Up for the Local Search", *Eleventh Annual International Conference on Research in Computational Molecular Biology (**RECOMB 2007**), LNCS 4453*: 238–252.

Proceedings Editor

(64) M. S. Bansal, Z. Cai, S. Manghul (eds), "Bioinformatics Research and Applications", *Lecture Notes in Bioinformatics, volume 13760, Springer, Berlin, 2023.*

(65) M. S. Bansal, I. Mandoiu, M. Moussa, M. Patterson, S. Rajasekaran, P. Skums, A. Zelikovsky (eds), "Computational Advances in Bio and Medical Sciences", *Lecture Notes in Bioinformatics, volume 13254, Springer, Berlin, 2022.*

Book Chapters

(66) M. S. Bansal, "Deciphering Microbial Gene Family Evolution Using Duplication-Transfer-Loss Reconciliation and RANGER-DTL", *In: H. Luo (eds), Environmental Microbial Evolution. Methods in Molecular Biology, vol 2569, 2022. Humana, New York, NY.*

Technical Reports

(67) M. Kordi, S. Kundu, and M. S. Bansal, "On Inferring Additive and Replacing Horizontal Gene Transfers Through Phylogenetic Reconciliation", *Technical Report 2020, bioRxiv.* (This is the full version of a paper published in abridged form in the proceedings of ACM-BCB 2019.)

(68) Y. Wu, M. S. Bansal, M. D. Rasmussen, J. Herrero, and M. Kellis, "Phylogenetic Identification and Functional Characterization of Orthologs and Paralogs across Human, Mouse, Fly, and Worm", *Technical Report 2014, bioRxiv.*

(69) M. S. Bansal and V. Ch. Venkaiah, "Improved Fully Polynomial Time Approximation Scheme for the 0-1 Multiple-Choice Knapsack Problem", *Technical Report 2004, IIIT-H, India.*

Manuscripts Under Review

(70) S. Zaman and M. S. Bansal, "Reducing the impact of domain rearrangement on sequence alignment and phylogeny reconstruction".

(71) L. Kloub, S. Gosselin, J. Graf, J. P. Gogarten, and M. S. Bansal, "Investigating Additive and Replacing Horizontal Gene Transfers using Phylogenies and Whole Genomes".

Software Packages

- **RANGER-DTL**: (https://compbio.engr.uconn.edu/software/RANGER-DTL/) Short for Rapid ANalysis of Gene family Evolution using Reconciliation-DTL, this is a software package for inferring gene family evolution by speciation, gene duplication, horizontal gene transfer, and gene loss. RANGER-DTL v2.0 is described in ref. (14) and implements the algorithms presented in refs. (52), (21), (15), and (12), and makes it possible to perform rigorous evolutionary analyses of even large gene families with thousands of taxa while accounting for confounding factors such as gene tree uncertainty, uncertain rooting, and multiple optima.

- **SaGePhy**: (https://compbio.engr.uconn.edu/software/sagephy/) Short for "Simulation framework for Subgene and Gene Phylogenies", this is an easy-to-use, open-source, and platform independent software package for simulating gene family evolution within species trees as well as subgene or protein-domain evolution within one or more gene trees. SaGePhy implements a number important features not found in other phylogenetic simulation software and is described in ref. (10). SaGePhy was programmed by Soumya Kundu.

- **HoMer**: (https://compbio.engr.uconn.edu/software/homer/) Short for "Horizontal Multi-gene transfer inference", HoMer is a software package for inferring instances of horizontal multi-gene transfer (HMGT) during the evolutionary history of a collection of microbial species/strains. An HMGT occurs when multiple genes are horizontally transferred in single horizontal transfer event, and the software implements the HMGT inference framework introduced in ref. (7). HoMer was implemented by Lina Kloub.

- **CNAsim**: (https://compbio.engr.uconn.edu/software/cnasim/) CNAsim is a software package for improved simulation of single-cell copy number alteration (CNA) data from tumors. It can be used to generate copy number profiles with noise patterns that mimic those of single-cell CNA detection algorithms, and to generate DNA-seq data for sampled cells. CNAsim offers significantly improved scalability, a high degree of customizability, and improved biological realism of simulated data, and is described in ref. (1). CNAsim was implemented by Samson Weiner.

- **TNet**: (https://compbio.engr.uconn.edu/software/tnet/) This is a phylogeny-based method for reconstructing transmission networks for infectious diseases. It takes as input a phylogeny of the strain (pathogen) sequences sampled from infected hosts and analyzes it to estimate the underlying transmission network. TNet relies on the availability of multiple strain sequences from each sampled host to infer transmissions and is simpler and more accurate than existing approaches. The method is parameter-free and highly scalable and can be easily applied within seconds to datasets with hundreds of strain sequences and hosts. This software implements the algorithms described in refs. (41) and (5). TNet was implemented by Saurav Dhar.

- **TNet-Geo**: (https://compbio.engr.uconn.edu/software/tnet-geo/) This is a phylogeny-based method for reconstructing geographical transmission networks for infectious diseases. It is an extended version of the TNet software above and is specifically designed for inferring the extent of infection spread between different geographical regions (e.g., countries) within a specified time frame. As with TNet, TNet-Geo is parameter-free and highly scalable and can be easily applied within seconds to datasets with thousands of strain sequences from hundreds of geographical regions. This software implements the approach described in ref. (5). TNet-Geo was implemented by Saurav Dhar.

- **RF+**: (https://compbio.engr.uconn.edu/software/rf_plus/) This is a program for computing RF(+) distances between phylogenetic trees. RF(+) distance is designed to more meaningfully compute the Robinson-Foulds distance between two trees that only have a partially overlapping leaf set. The RF(+) distance between two arbitrary trees is computed by first optimally completing each tree on the union of the leaf sets of both trees so as to minimize the Robinson-Foulds distance between them, and then reporting the Robinson-Foulds distance between the two completed trees. This software implements the algorithms described in refs. (9), (45), and (38). RF+ was implemented by Ashim Ranjeet and Keegan Yao.

- **DaTeR**: (https://compbio.engr.uconn.edu/software/dater/) This is a program for improved dating of microbial species phylogenies using relative time constraints (e.g., obtained from high-confidence horizontal gene transfer events). DaTeR can minimally error-correct input chronograms estimated using traditional approaches to ensure compatibility with all available relative time constraints. This software implements the algorithms described in ref. (3). DaTeR was implemented by Abhijit Mondal.

- **TreeFix-TP**: (https://compbio.engr.uconn.edu/software/treefix-tp/) This is a program for reconstructing highly accurate transmission phylogenies, i.e., phylogenies depicting the evolutionary relationships between infectious disease strains (viral or bacterial) transmitted between different hosts, and is described in ref. (39). TreeFix-TP is designed for scenarios where multiple strain sequences have been sampled from each infected host, and it uses the host assignment of each sequence sample to error-correct a given maximum likelihood phylogeny of the strain sequences. TreeFix-TP was

programmed by Samuel Sledzieski.

- **SEADOG**: (https://compbio.engr.uconn.edu/software/seadog/) Short for "Simultaneous Evolutionary Analysis of DOmains and Genes through phylogenetic reconciliation", this is a software package for simultaneous inference of domain-level and gene-level evolution through a joint phylogenetic reconciliation of domain, gene, and species trees. The software implements the Domain-Gene-Species (DGS) reconciliation model described in ref. (11) and the algorithms described in refs. (11), (44), and (46). SEADOG was programmed by Lei Li.

- **SEADOG-Gen**: (https://compbio.engr.uconn.edu/software/seadog-gen/) SEADOG-Gen is a software package for joint phylogenetic reconciliation of domain, gene, and species trees for microbial species. SEADOG-Gen is similar to the SEADOG software above but assumes that both domain transfer and gene transfer can occur easily in the species being analyzed. The software implements the Generalized Domain-Gene-Species (Gen-DGS) reconciliation model and algorithms described in ref. (2). SEADOG-Gen was implemented by Abhijit Mondal.

- **ARTra**: (https://compbio.engr.uconn.edu/software/ARTra/) This is a program for inferring and distinguishing between additive and replacing horizontal gene transfer events. ARTra uses Duplication-Transfer-Loss (DTL) reconciliation to infer transfer events and then uses a trained machine learning classifier to classify the inferred transfers as additive or replacing. The machine learning framework and rule-based heuristics used by ARTra are described in ref. (40). ARTra was implemented by Abhijit Mondal.

- **trippd**: (https://github.com/suz11001/Tripartition) trippd is a prototype implementation of a simple proof-of-concept approach for detecting the presence of partial gene transfer (i.e., horizontal transfer of a fragment of a gene) in a given gene family. trippd can be used to easily identify gene families whose gene trees may have been impacted by the presence of significant partial gene transfer. trippd implements the approach described in ref. (37) and was implemented by Sumaira Zaman.

- **virDTL**: (https://github.com/suz11001/virDTL) virDTL is a computational protocol for the inference of both extant and ancestral strain recombination in viral genomes using phylogenetic reconciliation. virDTL leverages Duplication-Transfer-Loss reconciliation to analyze incongruencies between the strain evolutionary tree and the evolutionary trees of each gene family (or genomic regions) to infer possible horizontal gene transfers, which correspond to possible recombination events in the context of viruses. virDTL was implemented by Sumaira Zaman and Samuel Sledzieski and further details appear in ref. (4).

- **TreeSolve**: (https://compbio.engr.uconn.edu/software/treesolve/) This is a program for gene tree error-correction. TreeSolve is designed for the error-correction of microbial gene trees (with horizontal gene transfer) but can be easily applied to non-microbial gene trees as well. TreeSolve serves a similar purpose as the TreeFix-DTL program described below, but is far more scalable and yields multiple candidate error-corrected gene trees. TreeSolve implements the algorithm described in ref. (42) and was implemented by Misagh Kordi.

- **RF-Supertrees**: (https://genome.cs.iastate.edu/rfsupertrees) This is a fast and accurate supertree program for rooted phylogenetic trees. RF-supertrees implements efficient search algorithms described in ref. (30) and constructs a supertree that minimizes the total Robinson-Foulds distance between the supertree and the input trees.

- **DupTree**: (https://genome.cs.iastate.edu/DupTree) This is a tool box for constructing species phylogenies from genome-scale multi-locus data using gene tree parsimony. DupTree implements the fast local search algorithm from ref. (63) and was programmed jointly with André Wehe.

- **TreeFix**: (http://compbio.mit.edu/treefix/) This is a program for very accurate reconstruction of eukaryotic gene trees. It takes as input a maximum likelihood gene tree topology, a known species tree, and a multiple sequence alignment for the gene family and computes a more accurate gene tree topology using the approach described in ref. (24). TreeFix was programmed by Yi-Chieh Wu.

- **TreeFix-DTL**: (http://compbio.mit.edu/treefix-dtl/) This is a program for very accurate reconstruction of prokaryotic gene trees. Like the program TreeFix above, it takes as input a maximum likelihood gene tree topology, a known species tree, and a multiple sequence alignment for the gene family and

computes a more accurate gene tree topology while accounting for horizontal gene transfer, gene duplication, and gene loss as described in ref. (18). TreeFix-DTL was programmed by Yi-Chieh Wu.

- **DupLoss** and **DeepC**: (https://genome.cs.iastate.edu/igtp/home) These programs extend on the program DupTree and make it possible to construct species phylogenies, from genome-scale multi-locus data, under the duplication-loss and deep coalescence cost models respectively. They implement the local search algorithms described in ref. (56) and are available as part of the iGTP (29) package.

- **HiDe**: (http://acgt.cs.tau.ac.il/hide/) Short for Highway Detection, HiDe is a software package for inferring highways of horizontal gene transfer (representing large-scale horizontal transfer of genes) in the evolutionary history of a set of species. HiDe implements the highway detection method described in ref. (23) and was programmed by undergraduate summer student Guy Banay under my supervision.

TEACHING

Department of Computer Science & Engineering, University of Connecticut, USA.

1. CSE 2500 Introduction to Discrete Systems: *Fall 2015, Spring 2016*
2. CSE 3500 Algorithms and Complexity: *Spring 2014, Spring 2018, Fall 2023*
3. CSE 3502 Theory of Computation: *Fall 2016, Fall 2017, Spring 2019*
4. CSE 3800/CSE 5800/BME 4800 Bioinformatics: *Fall 2014*
5. CSE 4939/4940 Senior Design Project I and II: *Fall 2020/Spring 2021, Fall 2022/Spring 2023*
6. CSE 5050 Algorithms and Complexity (graduate level): *Spring 2022, Fall 2022, Fall 2023*
7. CSE 5095 Approximation, Randomized, and Fixed Parameter Algorithms: *Fall 2013*
8. CSE 5095 Algorithmic Graph Theory: *Fall 2020*
9. CSE 5500 Algorithms (graduate level): *Spring 2015, Fall 2015, Fall 2016, Fall 2017, Spring 2021, Fall 2021, Spring 2023*
10. CSE 5860 Computational Problems in Evolutionary Genomics: *Fall 2018, Spring 2020*

STUDENTS

Current lab members:

1. Sumaira Zaman (PhD student; since Fall 2018)
2. Samson Weiner (PhD student; since Fall 2020. Former undergraduate researcher)
3. Shreya Seshadri (undergraduate researcher; since Spring 2023)
4. Vihaan Shah (undergraduate researcher; since Summer 2023)
5. Sunwang Luo (undergraduate researcher; since Summer 2023)
6. Abigail Horning (undergraduate researcher; since Summer 2023)

Previous lab members:

1. Rachel Parsons (undergraduate researcher; Fall 2021 - Summer 2023)
2. Abhijit Mondal (PhD student; Fall 2018 - Fall 2022)
3. Sai Chimbili (high-school student researcher; Spring 2022 - Fall 2022)
4. Rishik Gowrishetti (high-school student researcher; Summer 2022 - Fall 2022)
5. Dylan Cassidy (undergraduate researcher; Summer 2022)
6. Andrew Abi-Karam (undergraduate researcher; Fall 2021)
7. Zack Eisbach (high-school student researcher; Summer 2021 - Fall 2021)
8. Aryamaan Dhomne (high-school student researcher; Summer 2021- Fall 2021)
9. Lina Kloub (PhD student; Fall 2017 - Summer 2021)

10. Keegan Yao (undergraduate researcher; Spring 2019 - Summer 2021)
11. Saurav Dhar (MS student jointly advised with Ion Mandoiu; Fall 2018 - Spring 2021)
12. Taylor Wade (undergraduate researcher; Fall 2017 - Spring 2020)
13. Emily Maciejewski (undergraduate researcher; Fall 2018 - Spring 2020)
14. Ming Luo (undergraduate researcher; Summer 2019 - Fall 2019)
15. Misagh Kordi (PhD student; Spring 2014 - Spring 2019)
16. Lei Li (PhD student; Fall 2013 - Spring 2019)
17. Samuel Sledzieski (undergraduate researcher; Spring 2017 - Spring 2019)
18. Ashim Ranjeet (undergraduate researcher; Spring 2018 - Spring 2019)
19. Brian Jaworowski (undergraduate researcher; Spring 2018 - Fall 2018)
20. Soumya Kundu (undergraduate research assistant and MS student; Spring 2016 - Summer 2018)
21. Hannah Reed (undergraduate research assistant; Summer 2017 - Spring 2018)
22. Alex Masi (undergraduate research assistant; Summer 2017 - Fall 2017)
23. Eric van Heel (undergraduate research assistant; Summer 2017 - Fall 2017)
24. Chengchen Zhang (undergraduate research assistant; Spring 2017 - Summer 2017)
25. Kevin Grabowski (undergraduate research assistant; Summer 2016 - Spring 2017)
26. Maryzyrene Adao (undergraduate research assistant; Summer 2016)
27. James Boivie (undergraduate research assistant; Summer 2015 - Fall 2015)

PROFESSIONAL SERVICE

Conference program committees:

- International Symposium on Bioinformatics Research and Applications (ISBRA) 2023
- International Conference on Intelligent Systems for Molecular Biology (ISMB) 2023
- RECOMB Comparative Genomics (RECOMB-CG) 2023
- International Conference on Research in Computational Molecular Biology (RECOMB) 2023
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2022; (*PC co-chair*)
- International Conference on Intelligent Systems for Molecular Biology (ISMB) 2022
- RECOMB Comparative Genomics (RECOMB-CG) 2022
- International Conference on Research in Computational Molecular Biology (RECOMB) 2022
- RECOMB Comparative Genomics (RECOMB-CG) 2021
- IEEE International Workshop on High Performance Computational Biology (HiCOMB) 2021
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2021
- International Conference on Research in Computational Molecular Biology (RECOMB) 2021
- IEEE International Workshop on High Performance Computational Biology (HiCOMB) 2020
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2020
- International Conference on Research in Computational Molecular Biology (RECOMB) 2020
- International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM) 2020
- IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2019
- ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) 2019
- IEEE International Workshop on High Performance Computational Biology (HiCOMB) 2019
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2019

- RECOMB Satellite Workshop on Massively Parallel Sequencing (RECOMB-Seq) 2019
- International Conference on Bioinformatics and Computational Biology (BICoB) 2019
- ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) 2018
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2018
- International Conference on Bioinformatics and Computational Biology (BICoB) 2018
- ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) 2017
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2017
- International Conference on Bioinformatics and Computational Biology (BICoB) 2017
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2016
- International Conference on Intelligent Systems for Molecular Biology (ISMB) 2016
- International Conference on Bioinformatics and Computational Biology (BICoB) 2016
- International Conference on Intelligent Systems for Molecular Biology and European Conference on Computational Biology (ISMB/ECCB) 2015
- International Symposium on Bioinformatics Research and Applications (ISBRA) 2015
- International Conference on Bioinformatics and Computational Biology (BICoB) 2015
- International Conference on Research in Computational Molecular Biology (RECOMB) 2014
- International Conference on Applied Algorithms (ICAA) 2014
- International Conference on Intelligent Systems for Molecular Biology (ISMB) 2013
- International Conference on Bioinformatics and Computational Biology (BICoB) 2013
- ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) 2012
- International Workshop on Data Mining in Bioinformatics (BIOKDD) 2012
- International Conference on Bioinformatics and Computational Biology (BICoB) 2012

Conference/workshop organizing:
- Program committee co-chair for International Symposium on Bioinformatics Research and Applications (ISBRA) 2022.
- Workshop co-chair for Computational Advances in Molecular Epidemiology (CAME) 2021.
- Tutorials co-chair in the organizing committee of ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) 2020.
- Co-chair for "Inferring Phylogenies and Haplotypes" track at the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) 2017.
- Co-organizer and host for "Biogeochemical Dating in Deep Time" workshop. May 2017
- Finance chair for IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS). 2014, 2015, 2016

Selected journal and conference reviewing:
- Algorithms for Molecular Biology (AMB)
- Bioinformatics
- BMC Bioinformatics
- BMC Evolutionary Biology
- Cladistics
- Genome Biology
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)
- Information Processing Letters (IPL)

- Journal of Computational Biology (JCB)
- Journal of Mathematical Biology (JOMB)
- Journal of Theoretical Biology (JTB)
- Mathematical Biosciences
- Methods in Ecology and Evolution
- Molecular Biology and Evolution (MBE)
- Molecular Phylogenetics and Evolution (MPE)
- Nature Ecology & Evolution
- PLoS Computational Biology
- Proceedings of the National Academy of Sciences (PNAS)
- SIAM Journal on Discrete Mathematics (SIDMA)
- Systematic Biology
- Theoretical Computer Science (TCS)
- Theoretical Population Biology (TPB)
- Combinatorial Pattern Matching (CPM)
- International Symposium on Bioinformatics research and Applications (ISBRA)
- Latin American Theoretical Informatics (LATIN)
- Pacific Symposium on Biocomputing (PSB)
- International Conference on Research in Computational Molecular Biology (RECOMB)
- RECOMB Comparative Genomics (RECOMB-CG)
- ACM-SIAM Symposium on Discrete Algorithms (SODA)
- Workshop on Algorithms in Bioinformatics (WABI)

Professional Memberships

- Association for Computing Machinery (ACM)
- ACM Special Interest Group on Bioinformatics, Computational Biology, and Biomedical Informatics (SIGBio)
- International Society for Computational Biology (ISCB)
- Society for Molecular Biology and Evolution (SMBE)
- Society of Systematic Biologists (SSB)