

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

- 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.

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 Ninth 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, "Improved Duplication-Transfer-Loss Reconciliation with Extinct and Unsampled Lineages", *Algorithms* 2021, 14(8): 231.
- (2) 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* 2021 (*in press*).
- (3) 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.
- (4) 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.
- (5) M. S. Bansal, "Linear-Time Algorithms for Phylogenetic Tree Completion Under Robinson-Foulds Distance", *Algorithms for Molecular Biology* 2020, 15:6.

- (6) S. Kundu and M. S. Bansal, “SaGePhy: An improved phylogenetic simulation framework for gene and subgene evolution”, *Bioinformatics* 2019, 35(18): 3496–3498.
- (7) 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.
- (8) 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.
- (9) W. Duchemin, G. Gence, A. Arigon-Chiffolleau, 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.
- (10) 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.
- (11) 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.
- (12) 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.
- (13) 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. Shores, 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. Siebenthal, 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.
- (14) 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.
- (15) 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. Kutuyavin, 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. Skoultschi, 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.
- (16) 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.
- (17) 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.
- (18) 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.
- (19) 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.
- (20) 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.
- (21) 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.
- (22) M. S. Bansal, J. Dong, and D. Fernández-Baca, “Comparing and Aggregating Partially Resolved Trees”, *Theoretical Computer Science* 2011, 412: 6634–6652.
- (23) 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.
- (24) 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.
- (25) 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.
- (26) M. S. Bansal, J. G. Burleigh, O. Eulenstein, and D. Fernández-Baca, “Robinson-Foulds Supertrees”, *Algorithms for Molecular Biology* 2010, 5:18.
- (27) 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.
- (28) 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.
- (29) M. S. Bansal and D. Fernández-Baca, “Computing Distances Between Partial Rankings”, *Information Processing Letters* 2009, 109(4): 238–241.
- (30) 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.
- (31) 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.
- (32) 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.)

- (33) 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**.
- (34) 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.
- (35) 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.
- (36) 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.
- (37) M. Kordi and M. S. Bansal, “TreeSolve: Rapid Error-Correction of Microbial Gene Trees”, *7th International Conference on Algorithms for Computational Biology (AICoB 2020)*, *LNCS 12099*: 125–139.
- (38) 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.
- (39) 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.
- (40) 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.
- (41) L. Li and M. S. Bansal, “An Integer Linear Programming Solution for the Domain-Gene-Species Reconciliation Problem”, *Ninth ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2018)*; *Proceedings*: 386–397. This paper won the **Best Student Paper Award**.
- (42) 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)
- (43) 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.
- (44) 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.
- (45) R. Libeskind-Hadas, Y. Wu, M. S. Bansal, and M. Kellis, “Pareto-Optimal Phylogenetic Tree Reconciliation”, *22nd Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2014)*; *Bioinformatics* 2014, 30: i87–i95.
- (46) 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.
- (47) 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.

- (48) 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.
- (49) 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.
- (50) 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.
- (51) 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.
- (52) 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**.
- (53) 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.
- (54) 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.
- (55) 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.
- (56) 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.
- (57) 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.
- (58) 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.

TECHNICAL REPORTS

- (59) 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.)
- (60) 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*.
- (61) M. S. Bansal and V. Ch. Venkaiah, “Improved Fully Polynomial Time Approximation Scheme for the 0-1 Multiple-Choice Knapsack Problem”, *Technical Report 2004*, *IIT-H, India*.

MANUSCRIPTS UNDER REVIEW

- (62) S. Zaman, S. Sledzieski, B. Berger, Y. Wu, and M. S. Bansal “Phylogenetic reconciliation reveals extensive ancestral recombination in sarbecoviruses and the SARS-CoV-2 lineage”.
- (63) 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. (10) and implements the algorithms presented in refs. (47), (17), (11), and (8), 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. (6). 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. (3). HoMer was implemented by Lina Kloub.
- **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. (36) and (2). 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. (2). 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. (5), (40), and (33). RF+ was implemented by Ashim Ranjeet and Keegan Yao.

- **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. (34). 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. (7) and the algorithms described in refs. (7), (39), and (41). SEADOG was programmed by Lei Li.
- **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 classifier uses the error-prone classifications generated by several simple rule-based classification heuristics, along with some additional features, to generate an improved ensemble classification. The machine learning framework and rule-based heuristics used by ARTra are described in ref. (35). ARTra was implemented by Abhijit Mondal.
- **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 works by collapsing weakly supported branches in the input gene tree and then optimally resolving it based on a given sample of bootstrap replicate or posterior gene trees and a given species tree, while accounting for horizontal gene transfer, gene duplication, and gene loss. 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. (37) 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. (26) 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. (58) 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. (20). 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. (14). 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. (51) and are available as part of the iGTP (25) 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. (19) and was programmed by undergraduate summer student Guy Banay under my supervision.

SELECTED ORAL PRESENTATIONS AND INVITED TALKS

1. “TreeSolve: Rapid Error-Correction of Microbial Gene Trees”; *International Conference on Algorithms for Computational Biology (AlCoB)*, Virtual, November 2021.
2. “Understanding Horizontal Gene Transfer: Units of Transfer and Modes of Integration”; *Institute for Systems Genomics Symposium*, Virtual, May 2021.
3. “On Inferring Additive and Replacing Horizontal Gene Transfers Through Phylogenetic Reconciliation”; *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*, Niagara Falls, NY, USA. September 2019.
4. “TNet: Phylogeny-Based Inference of Disease Transmission Networks Using Within-Host Strain Diversity”; *Workshop on Computational Advances in Molecular Epidemiology (CAME)*, Niagara Falls, NY, USA. September 2019.
5. “Automated Inference of Horizontal Gene Transfers for Phylogenetic Dating”; *Dating in Deep Time II workshop*, Cambridge, MA, USA. August 2019
6. “Systematic Analysis of Genome-Wide Horizontal Gene Transfer Patterns Reveals Extensive Multi-Gene Transfer in Prokaryotes”; *Evolution meeting*, Providence, RI, USA. June 2019
7. “Genome-Scale Phylogenetics Through Gene Tree Parsimony”, *Trees in the Desert Workshop*, Tucson, USA. April 2019. *Invited*.
8. “Reconstructing Highly Accurate Gene Trees using TreeFix and TreeFix-DTL: Tutorial and Software Demonstration”, *Trees in the Desert Workshop*, Tucson, USA. April 2019. *Invited*.
9. “Linear-Time Algorithms for Some Phylogenetic Tree Completion Problems under Robinson-Foulds Distance”, *RECOMB Comparative Genomics Conference (RECOMB-CG)*, Sherbrooke, Canada. October 2018.
10. “TreeFix-VP: Error-correcting Viral Transmission Phylogenies”, *Workshop on Computational Advances in Molecular Epidemiology (CAME)*, Washington, D.C., USA. August 2018.
11. “An Integrated Reconciliation Framework for Domain, Gene, and Species Level Evolution”, *Asia-Pacific Bioinformatics Conference (APBC)*, Yokohama, Japan. January 2018.
12. “Phylogenetic Error-Correction for Viral Transmission Network Inference”; *Workshop on Computational Advances in Molecular Epidemiology (CAME)*, Cambridge, USA. August 2017.
13. “Inferring Microbial Gene Family Evolution Despite Phylogenetic Uncertainty”; *Evolution meeting*, Portland, OR, USA. June 2017
14. “Phylogenetic Uncertainty and Transmission Network Inference: Lessons from Phylogenetic Reconciliation”; *Workshop on Computational Advances in Molecular Epidemiology (CAME)*, Atlanta, USA. October 2016.
15. “Deciphering microbial evolution using phylogenetic reconciliation”; *IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, Atlanta, USA. October 2016. *Invited*.
16. “Exact Algorithms for Duplication-Transfer-Loss Reconciliation with Non-Binary Gene Trees”; *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*, Seattle, USA. October 2016.
17. “Using Phylogenetic Reconciliation to Decipher Microbial Evolution: Theory and Practice”; *Evolution meeting*, Austin, USA. June 2016
18. “Deciphering Reticulate Evolution Using Phylogenetic Reconciliation”; *Networks in Biological Sciences*, Singapore. July 2015. *Invited*.
19. “Deciphering Reticulate Evolution Using Phylogenetic Reconciliation”; *Utilizing Genealogical Phylogenetic Networks in Evolutionary Biology: Touching the Data*, Leiden, Netherlands. July 2014. *Invited*.

20. “Deciphering Microbial Evolution Using Phylogenetic Reconciliation”; *Computational Biology and Bioinformatics Organization meeting, Broad Institute of MIT and Harvard, Cambridge, USA. March 2014. Invited.*
21. “Deciphering Microbial Evolution Using Phylogenetic Reconciliation”; *Biomedical Engineering Seminar, University of Connecticut, Storrs, USA. March 2014.*
22. “Reconciliation Revisited: Handling Multiple Optima when Reconciling with Duplication, Transfer, and Loss”; *International Conference on Research in Computational Molecular Biology (RECOMB), April 2013, Beijing, China.*
23. “Accurate gene tree reconstruction using TreeFix and TreeFix-DTL”; *Symposium and Workshop on New Methods for Phylogenomics and Metagenomics, University of Texas, Austin, USA. February 2013. Invited.*
24. “Reconciliation revisited: Towards faster and more accurate inference of gene family evolution by duplication, transfer, and loss”; *Computational and Statistical Phylogenomics Meeting, Vigo, Spain. September 2012. Invited.*
25. “Efficient Algorithms for the Reconciliation Problem with Gene Duplication, Horizontal Transfer, and Loss”; *International Conference on Intelligent Systems for Molecular Biology (ISMB), July 2012, Long Beach, USA.*
26. “Reconciliation revisited: Towards more accurate inference of gene family evolution”; *Mathematical and Computational Evolutionary Biology (MCEB), June 2012, Montpellier, France.*
27. “Reconstructing gene family evolution by speciation, duplication, horizontal transfer & loss”; *Boston Evolutionary Genomics Supergroup Meeting, August 2011, Cambridge, USA.*
28. “Systematic Detection of Large-Scale Horizontal Gene Transfer Events on the Tree of Life”; *Computer Science Colloquium, Iowa State University, Ames, USA. April 2011. Invited.*
29. “Detecting Highways of Horizontal Gene Transfer”; *RECOMB Comparative Genomics Workshop (RECOMB-CG), October 2010, Ottawa, Canada.*
30. “Inferring Species Trees and Large-Scale Horizontal Gene Transfer from Discordant Gene Trees”; *Computational Biology Group and Laboratory for Microbiology, Massachusetts Institute of Technology, Cambridge, USA. October 2010. Invited.*
31. “Inferring Species Trees and Large-Scale Horizontal Gene Transfer from Discordant Gene Trees”; *Evolutionary Genomics Research Group, NCBI, National Institutes of Health, Bethesda, USA. July 2010. Invited.*
32. “Detecting Highways of Horizontal Gene Transfer”; *Bertinoro Computational Biology Meeting, May 2010, Bertinoro, Italy.*
33. “Detecting Highways of Horizontal Gene Transfer”; *Stringology Research Workshop, May 2010, Tel-Aviv, Israel.*
34. “Genome-Scale Phylogenetic Analyses using Gene Tree Parsimony”; *Asia-Pacific Bioinformatics Conference (APBC), January 2010, Bangalore, India.*
35. “Generalized Binary Tanglegrams”; *International Conference on Bioinformatics and Computational Biology (BICoB), April 2009, New Orleans, USA.*
36. “The Gene-Duplication Problem”; *Computational Genomics Research Group, Tel Aviv University, Tel Aviv, Israel. February 2009. Invited.*
37. “The multiple gene duplication problem revisited”; *International Conference on Intelligent Systems for Molecular Biology (ISMB), July 2008, Toronto, Canada.*
38. “The Gene-Duplication Problem: Near-Linear Time Algorithms for NNI Based Local Searches”; *International Symposium on Bioinformatics Research and Applications (ISBRA), May 2008, Atlanta, USA.*
39. “Heuristics for the Gene-Duplication Problem: A $\Theta(n)$ Speed-up of the Local Search”; *International Conference on Research in Computational Molecular Biology (RECOMB), April 2007, San Francisco, USA.*

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*
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*
6. CSE 5050 Algorithms and Complexity (graduate level): *Spring 2022*
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*
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. Abhijit Mondal (PhD student; since Fall 2018)
3. Samson Weiner (PhD student; since Fall 2020. Former undergraduate researcher)
4. Andrew Abi-Karam (undergraduate researcher; since Fall 2021)
5. Rachel Parsons (undergraduate researcher; since Fall 2021)

Previous lab members:

1. Zack Eisbach (high-school student researcher; Summer 2021 - Fall 2021)
2. Aryamaan Dhomne (high-school student researcher; Summer 2021- Fall 2021)
3. Lina Kloub (PhD student; Fall 2017 - Summer 2021)
4. Keegan Yao (undergraduate researcher; Spring 2019 - Summer 2021)
5. Saurav Dhar (MS student jointly advised with Ion Mandoiu; Fall 2018 - Spring 2021)
6. Taylor Wade (undergraduate researcher; Fall 2017 - Spring 2020)
7. Emily Maciejewski (undergraduate researcher; Fall 2018 - Spring 2020)
8. Ming Luo (undergraduate researcher; Summer 2019 - Fall 2019)
9. Misagh Kordi (PhD student; Spring 2014 - Spring 2019)
10. Lei Li (PhD student; Fall 2013 - Spring 2019)
11. Samuel Sledzieski (undergraduate researcher; Spring 2017 - Spring 2019)
12. Ashim Ranjeet (undergraduate researcher; Spring 2018 - Spring 2019)
13. Brian Jaworowski (undergraduate researcher; Spring 2018 - Fall 2018)
14. Soumya Kundu (undergraduate research assistant and MS student; Spring 2016 - Summer 2018)
15. Hannah Reed (undergraduate research assistant; Summer 2017 - Spring 2018)
16. Alex Masi (undergraduate research assistant; Summer 2017 - Fall 2017)
17. Eric van Heel (undergraduate research assistant; Summer 2017 - Fall 2017)
18. Chengchen Zhang (undergraduate research assistant; Spring 2017 - Summer 2017)
19. Kevin Grabowski (undergraduate research assistant; Summer 2016 - Spring 2017)
20. Maryzyrene Adao (undergraduate research assistant; Summer 2016)

21. James Boivie (undergraduate research assistant; Summer 2015 - Fall 2015)

PROFESSIONAL SERVICE

Conference program committees:

- 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:

- 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 (ICCBS). 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)