

List of publications

Siavash Mirarab

December 15, 2023

1. Chao Zhang, Rasmus Nielsen and **Siavash Mirarab**. CASTER: Direct species tree inference from whole-genome alignments. *bioRxiv*, [prepress](#), 2023.
2. Shayesteh Arasti, Puoya Tabaghi, Yasamin Tabatabaee and Siavash Mirarab. Optimal Tree Metric Matching Enables Phylogenomic Branch Length Reconciliation. *bioRxiv*, [2023.11.13.566962](#) 2023.
3. Metin Balaban, Yueyu Jiang, Qiyun Zhu, Daniel McDonald, Rob Knight and **Siavash Mirarab**. Generation of accurate, expandable phylogenomic trees with uDance. *Nature Biotechnology*, [Online](#) 2023.
4. Daniel McDonald, Yueyu Jiang, Metin Balaban et al. Greengenes2 unifies microbial data in a single reference tree. *Nature Biotechnology*, [Online](#) 2023.
5. Yasamin Tabatabaee, Chao Zhang, Tandy Warnow and **Siavash Mirarab**. Phylogenomic branch length estimation using quartets. *Bioinformatics*, [39\(Supplement_1\):i185–i193](#) 2023.
6. Ali Osman Berk Şapcı, Eleonora Rachtman and **Siavash Mirarab**. CONSULT-II: Taxonomic Identification Using Locality Sensitive Hashing. In *Comparative Genomics*, 196–214, Cham, 2023. Springer Nature Switzerland.
7. Shayesteh Arasti and **Siavash Mirarab**. Optimal Subtree Prune and Regraft for Quartet Score in Sub-Quadratic Time. In *23rd International Workshop on Algorithms in Bioinformatics (WABI 2023)*, volume 273 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 4:1–4:20, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISSN: 1868-8969.

8. Uyen Mai, Eduardo Charvel and **Siavash Mirarab**. Expectation-Maximization enables Phylogenetic Dating under a Categorical Rate Model. In *RECOMB*, 2023.
9. Afif Elghraoui, **Siavash Mirarab**, Krister M Swenson and Faramarz Valafar. Evaluating impacts of syntenic block detection strategies on rearrangement phylogeny using *Mycobacterium tuberculosis* isolates. *Bioinformatics*, [39\(1\):btad024](#) 2023.
10. Chao Zhang, Andrey V. Bzikadze, Yana Safonova and **Siavash Mirarab**. A scalable model for simulating multi-round antibody evolution and benchmarking of clonal tree reconstruction methods. *Frontiers in Immunology*, [13:1014439](#), 2022.
11. Chao Zhang and **Siavash Mirarab**. Weighting by Gene Tree Uncertainty Improves Accuracy of Quartet-based Species Trees. *MBE*, [39\(12\):msac215](#) 2022.
12. Chao Zhang and **Siavash Mirarab**. ASTRAL-Pro 2: ultrafast species tree reconstruction from multi-copy gene family trees. *Bioinformatics*, [38\(21\):4949–4950](#) 2022.
13. Metin Balaban, Nishat A. Bristy, Ahnaf Faisal, Md. Shamsuzzoha Bayzid and **Siavash Mirarab**. Genome-wide alignment-free phylogenetic distance estimation under a no strand-bias model. *Bioinformatics Advances*, [2\(1\):2021.11.10.468111](#) 2022.
14. Maryam Rabiee and **Siavash Mirarab**. QuCo: quartet-based co-estimation of species trees and gene trees. *Bioinformatics*, [38\(Supplement_1\):i413–i421](#) 2022.
15. Yueyu Jiang, Puoya Tabaghi and **Siavash Mirarab**. Learning Hyperbolic Embedding for Phylogenetic Tree Placement and Updates. *Biology*, [11\(9\):1256](#), 2022. ISBN: 2079-7737.
16. Navid Bin Hasan, Metin Balaban, Avijit Biswas, Md. Shamsuzzoha Bayzid and **Siavash Mirarab**. Distance-Based Phylogenetic Placement with Statistical Support. *Biology*, [11\(8\):1212](#) 2022. ISBN: 2079-7737.
17. Eleonora Rachtman, Shahab Sarmashghi, Vineet Bafna and **Siavash Mirarab**. Quantifying the uncertainty of assembly-free genome-wide distance estimates and phylogenetic relationships using subsampling. *Cell Systems*, [13\(10\):817–829.e3](#) 2022.
18. Uyen Mai and **Siavash Mirarab**. Completing gene trees without species trees in sub-quadratic time. *Bioinformatics*, [38\(6\):1532–1541](#) 2022.

19. Metin Balaban, Yueyu Jiang, Daniel Roush, Qiyun Zhu and **Siavash Mirarab**. Fast and accurate distance-based phylogenetic placement using divide and conquer. *Molecular Ecology Resources*, [22\(3\):1213–1227](#) 2022.
20. Shahab Sarmashghi, Metin Balaban, Eleonora Rachtman, Behrouz Touri, **Siavash Mirarab** and Vineet Bafna. Estimating repeat spectra and genome length from low-coverage genome skims with RESPECT. *PLOS Comp. Bio.*, [17\(11\):e1009449](#) 2021.
21. **Siavash Mirarab**, Luay Nakhleh and Tandy Warnow. Multispecies Coalescent: Theory and Applications in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, [52\(1\):247–268](#) 2021.
22. Eleonora Rachtman, Vineet Bafna and **Siavash Mirarab**. CONSULT: accurate contamination removal using locality-sensitive hashing. *NAR Genomics and Bioinformatics*, [3\(3\):10.1101/2021.03.18.436035](#) 2021.
23. Chao Zhang, Yiming Zhao, Edward Louis Braun and **Siavash Mirarab**. TAPER: Pinpointing errors in multiple sequence alignments despite varying rates of evolution. *Methods in Ecology and Evolution*, [12\(11\):2145–2158](#) 2021.
24. Niema Moshiri, Davey M Smith and **Siavash Mirarab**. HIV Care Prioritization Using Phylogenetic Branch Length. *JAIDS*, [86\(5\):626–637](#) 2021.
25. Maryam Rabiee and **Siavash Mirarab**. SODA: multi-locus species delimitation using quartet frequencies. *Bioinformatics*, [36\(24\):5623–5631](#) 2021.
26. Shaohong Feng, Josefin Stiller, Yuan Deng et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature*, [587\(7833\):252–257](#) 2020.
27. Uyen Mai and **Siavash Mirarab**. Log Transformation Improves Dating of Phylogenies. *MBE*, [38\(3\):1151–1167](#) 2021.
28. Chao Zhang, Celine Scornavacca, Erin K Molloy and **Siavash Mirarab**. ASTRAL-Pro: Quartet-Based Species-Tree Inference despite Paralogy. *MBE*, [37\(11\):3292–3307](#) 2020.
29. Kristine Bohmann, **Siavash Mirarab**, Vineet Bafna and M. Thomas P. Gilbert. Beyond DNA barcoding: The unrealized potential of genome skim data in sample identification. *Molecular Ecology*, [mec.15507](#), 2020.

30. Metin Balaban and **Siavash Mirarab**. Phylogenetic double placement of mixed samples. *Bioinformatics*, [36\(Supplement_1\):i335–i343](#) 2020.
31. Ekin Tilic, Erfan Sayyari, Josefin Stiller, **Siavash Mirarab** and Greg W Rouse. More is needed—Thousands of loci are required to elucidate the relationships of the ‘flowers of the sea’ (Sabellida, Annelida). *Molecular Phylogenetics and Evolution*, [151:106892](#) 2020.
32. Francesco Asnicar, Andrew Maltez Thomas, Francesco Beghini et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications*, [11\(1\):2500](#) 2020. Publisher: Springer US ISBN: 0348556977.
33. Eleonora Rachtman, Metin Balaban, Vineet Bafna and **Siavash Mirarab**. The impact of contaminants on the accuracy of genome skimming and the effectiveness of exclusion read filters. *Molecular Ecology Resources*, [20\(3\):1755–0998.13135](#) 2020.
34. Metin Balaban, Shahab Sarmashghi and **Siavash Mirarab**. APPLES: Scalable Distance-Based Phylogenetic Placement with or without Alignments. *Systematic Bio.*, [69\(3\):566–78](#) 2020.
35. Maryam Rabiee and **Siavash Mirarab**. INSTRAL: Discordance-Aware Phylogenetic Placement Using Quartet Scores. *Systematic Bio.*, [69\(2\):384–391](#) 2020.
36. Maryam Rabiee and **Siavash Mirarab**. Forcing external constraints on tree inference using ASTRAL. *BMC Genomics*, [21\(S2\):218](#) 2020. Publisher: BMC Genomics.
37. Qiyun Zhu, Uyen Mai, Wayne Pfeiffer et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nature Communications*, [10\(1\):5477](#) 2019.
38. James H Leebens-Mack, Michael S Barker, Eric J Carpenter et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, [574\(7780\):679–685](#) 2019.
39. Erfan Sayyari, Ban Kawas and **Siavash Mirarab**. TADA: phylogenetic augmentation of microbiome samples enhances phenotype classification. *Bioinformatics*, [35\(14\):i31–i40](#) 2019.

40. Lei Chen, Qiang Qiu, Yu Jiang et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science*, [364\(6446\):eaav6202](#) 2019.
41. Shahab Sarmashghi, Kristine Bohmann, M. Thomas P. Gilbert, Vineet Bafna and **Siavash Mirarab**. Skmer: assembly-free and alignment-free sample identification using genome skims. *Genome Biology*, [20\(1\):34](#) 2019.
42. John Yin, Chao Zhang and **Siavash Mirarab**. ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. *Bioinformatics*, [35\(20\):3961–3969](#) 2019.
43. Metin Balaban, Niema Moshiri, Uyen Mai, Xingfan Jia and **Siavash Mirarab**. TreeCluster: Clustering biological sequences using phylogenetic trees. *PLOS ONE*, [14\(8\):e0221068](#), 2019.
44. Niema Moshiri, Manon Ragonnet-Cronin, Joel O Wertheim and **Siavash Mirarab**. FAVITES: simultaneous simulation of transmission networks, phylogenetic trees and sequences. *Bioinformatics*, [35\(11\):1852–1861](#) 2019.
45. Peter Houde, Edward L Braun, Nitish Narula, Uriel Minjares and **Siavash Mirarab**. Phylogenetic Signal of Indels and the Neoavian Radiation. *Diversity*, [11\(7\):108](#) 2019.
46. Maryam Rabiee, Erfan Sayyari and **Siavash Mirarab**. Multi-allele species reconstruction using ASTRAL. *Molecular Phylogenetics and Evolution*, [130:286–296](#) 2019. Publisher: Elsevier.
47. Shyam Gopalakrishnan, Mikkel-Holger S. Sinding, Jazmín Ramos-Madrigal et al. Interspecific Gene Flow Shaped the Evolution of the Genus *Canis*. *Current Biology*, [28\(21\):3441–3449.e5](#) 2018.
48. Daniel McDonald, Embriette Hyde, Justine W. Debelius et al. American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems*, [3\(3\)](#) 2018.
49. Wolf L. Eiserhardt, Alexandre Antonelli, Dominic J. Bennett et al. A roadmap for global synthesis of the plant tree of life. *American Journal of Botany*, [105\(3\):614–622](#) 2018.

50. Stefan Janssen, Daniel McDonald, Antonio Gonzalez et al. Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *mSystems*, [3\(3\):e00021–18](#) 2018.
51. Chao Zhang, Maryam Rabiee, Erfan Sayyari and **Siavash Mirarab**. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, [19\(S6\):153](#) 2018.
52. Uyen Mai and **Siavash Mirarab**. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, [19\(S5\):272](#) 2018. ISBN: 9783319679785.
53. Erfan Sayyari and **Siavash Mirarab**. Testing for Polytomies in Phylogenetic Species Trees Using Quartet Frequencies. *Genes*, [9\(3\):132](#) 2018. Publisher: Multidisciplinary Digital Publishing Institute.
54. Erfan Sayyari, James B. Whitfield and **Siavash Mirarab**. DiscoVista: Interpretable visualizations of gene tree discordance. *Molecular Phylogenetics and Evolution*, [122:110–115](#) 2018. arXiv: 1709.09305.
55. Ali Akbari, Joseph J. Vitti, Arya Iranmehr et al. Identifying the favored mutation in a positive selective sweep. *Nature Methods*, [15\(4\):279–282](#) 2018. Publisher: Nature Publishing Group.
56. Shubhanshu Shekhar, Sebastien Roch and **Siavash Mirarab**. Species Tree Estimation Using ASTRAL: How Many Genes Are Enough? *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, [15\(5\):1738–1747](#), 2018.
57. Niema Moshiri and **Siavash Mirarab**. A Two-State Model of Tree Evolution and Its Applications to Alu Retrotransposition. *Systematic Bio.*, [67\(3\):475–489](#) 2018.
58. Erfan Sayyari, James Whitfield and **Siavash Mirarab**. Fragmentary Gene Sequences Negatively Impact Gene Tree and Species Tree Reconstruction. *MBE*, [34\(12\):3279–91](#) 2017.
59. Luke R. Thompson, Jon G. Sanders, Daniel McDonald et al. A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature*, [551\(7681\):457–463](#), 2017.

60. Uyen Mai and **Siavash Mirarab**. TreeShrink: Efficient Detection of Outlier Tree Leaves. In *Comparative Genomics. RECOMB-CG 2017. Lecture Notes in Computer Science, vol 10562*, 116–140. Springer International Publishing, Cham, 2017.
61. Dong-sheng Zhang, Ya-dong Zhou, Chun-sheng Wang and Greg Rouse. A new species of Ophryotrocha (Annelida, Eunicida, Dorvilleidae) from hydrothermal vents on the Southwest Indian Ridge. *ZooKeys*, [687:1–9](#) 2017.
62. Uyen Mai, Erfan Sayyari and **Siavash Mirarab**. Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction. *PLOS ONE*, [12\(8\):e0182238](#) 2017. Publisher: Public Library of Science ISBN: 1111111111.
63. **Siavash Mirarab**. Phylogenomics: Constrained gene tree inference. *Nature Ecology & Evolution*, [1\(2\):0056](#) 2017. Publisher: Nature Publishing Group.
64. Erfan Sayyari and **Siavash Mirarab**. Anchoring quartet-based phylogenetic distances and applications to species tree reconstruction. *BMC Genomics*, [17\(S10\):101–113](#) 2016.
65. 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.
66. Nam-phuong N.-P. Nguyen, Michael Nute, **Siavash Mirarab** and Tandy Warnow. HIPPI: highly accurate protein family classification with ensembles of HMMs. *BMC Genomics*, [17\(S10\):765](#) 2016.
67. James E Tarver, Mario dos Reis, **Siavash Mirarab** et al. The Interrelationships of Placental Mammals and the Limits of Phylogenetic Inference. *Genome Biology and Evolution*, [8\(2\):330–344](#) 2016.
68. **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.
69. Joel Cracraft, Peter Houde, Simon Y W Ho et al. Response to Comment on “Whole-genome analyses resolve early branches in the tree of life of modern birds”. *Science*, [349\(6255\):1460](#) 2015.

70. 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.
71. Jed Chou, Ashu Gupta, Shashank Yaduvanshi et al. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, [16\(Suppl 10\):S2](#), 2015.
72. **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.
73. Nam-phuong Nguyen, **Siavash Mirarab**, Keerthana Kumar and Tandy Warnow. Ultra-large alignments using phylogeny-aware profiles. *Genome Biology*, [16\(1\):124](#) 2015.
74. 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. Publisher: Public Library of Science.
75. Erich D Jarvis, **Siavash Mirarab**, Andre J Aberer et al. Phylogenomic analyses data of the avian phylogenomics project. *GigaScience*, [4\(1\):1–9](#) 2015. Publisher: BioMed Central.
76. **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\(5\):377–86](#) 2015.
77. E. D. Jarvis, **Siavash Mirarab**, A. J. Aberer et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, [346\(6215\):1320–1331](#) 2014.
78. **Siavash Mirarab**, M. S. Bayzid, Bastien Boussau and Tandy Warnow. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science*, [346\(6215\):1250463–1250463](#) 2014.
79. Naim Matasci, Ling-Hong L.-H. Hung, Zhixiang Yan et al. Data access for the 1,000 Plants (1KP) project. *GigaScience*, [3\(1\):17](#), 2014. Publisher: BioMed Central.

80. Norman Wickett, **Siavash Mirarab**, Nam Nguyen et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *PNAS*, [111\(45\):E4859–68](#) 2014. arXiv: 1301.5277v2 ISBN: 0027-8424.
81. **Siavash Mirarab**, Shamsuzzoha Bayzid and Tandy Warnow. Evaluating Summary Methods for Multilocus Species Tree Estimation in the Presence of Incomplete Lineage Sorting. *Systematic Bio.*, [65\(3\):366–380](#) 2016. Publisher: Oxford University Press ISBN: 1063-5157.
82. **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)*, 177–191, 2014.
83. **Siavash Mirarab**, Rezwana Reaz, Md. Shamsuzzoha Bayzid, Théo Zimmermann, M. S. Swenson and Tandy Warnow. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, [30\(17\):i541–i548](#) 2014.
84. Nam-phuong Nguyen, **Siavash Mirarab**, Bo Liu, Mihai Pop and Tandy Warnow. TIPP: taxonomic identification and phylogenetic profiling. *Bioinformatics*, [30\(24\):3548–3555](#) 2014.
85. Théo Zimmermann, **Siavash Mirarab** and Tandy Warnow. BBICA: Improving the scalability of *BEAST using random binning. *BMC Genomics*, [15\(Suppl 6\):S11](#), 2014.
86. Arlin Stoltzfus, Hilmar Lapp, Naim Matasci et al. Phylotastic! Making tree-of-life knowledge accessible, reusable and convenient. *BMC bioinformatics*, [14\(1\):158](#), 2013. Publisher: BioMed Central Ltd ISBN: 1471-2105 (Electronic)\r1471-2105 (Linking).
87. 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)*, 250–61, 2013.
88. Nam Nguyen, **Siavash Mirarab** and Tandy Warnow. MRL and SuperFine+ MRL: new supertree methods. *Algorithms for Molecular Biology*, [7\(1\):3](#) 2012.
89. **Siavash Mirarab**, Nam Nguyen and Tandy Warnow. SEPP: SATé-Enabled Phylogenetic Placement. In *PSB*, 247–58 2012.

90. **Siavash Mirarab**, Soroush Akhlaghi and Ladan Tahvildari. Size-constrained regression test case selection using multicriteria optimization. *Software Engineering, IEEE Transactions on*, **38(4)**:936–956, 2012. Publisher: IEEE.
91. **Siavash Mirarab** and Tandy Warnow. FastSP: linear time calculation of alignment accuracy. *Bioinformatics*, **27(23)**:3250–8, 2011. ISBN: 1367-4811; 1367-4803.
92. Hyunsook Do, **Siavash Mirarab**, Ladan Tahvildari and Gregg Rothermel. The effects of time constraints on test case prioritization: A series of controlled experiments. *Software Engineering, IEEE Transactions on*, **36(5)**:593–617, 2010. Publisher: IEEE.
93. **Siavash Mirarab** and Ladan Tahvildari. An empirical study on bayesian network-based approach for test case prioritization. In *Software Testing, Verification, and Validation, 2008 1st International Conference on*, 278–287. IEEE, 2008.
94. **Siavash Mirarab**, Afshar Ganjali, Ladan Tahvildari, Shimin Li, Weining Liu and Mike Morrissey. A requirement-based software testing framework: An industrial practice. In *Software Maintenance, 2008. ICSM 2008. IEEE International Conference on*, 452–455. IEEE, 2008.
95. 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*, 71–82. ACM, 2008.
96. Mehdi Amoui, Mazeiar Salehie, **Siavash Mirarab** and Ladan Tahvildari. Adaptive action selection in autonomic software using reinforcement learning. In *Autonomic and Autonomous Systems, 2008. ICAS 2008. Fourth International Conference on*, 175–181. IEEE, 2008.
97. **Siavash Mirarab** and Ladan Tahvildari. A prioritization approach for software test cases based on bayesian networks. In *Fundamental Approaches to Software Engineering*, volume 4422 LNCS, 276–290. Springer Berlin Heidelberg, 2007. ISSN: 03029743.
98. **Siavash Mirarab**, Alaa Hassouna and Ladan Tahvildari. Using bayesian belief networks to predict change propagation in software systems. In *Program Comprehension, 2007. ICPC’07. 15th IEEE International Conference on*, 177–188. IEEE, 2007.

99. Mehdi Amoui, **Siavash Mirarab**, Sepand Ansari and Caro Lucas. A genetic algorithm approach to design evolution using design pattern transformation. *International Journal of Information Technology and Intelligent Computing*, 1(2):235–244, 2006.