References Cited

[1] "The Target Malaria Project." `targetmalaria.org`.

[2] W. Liao, C. T. Atkinson, D. A. LaPointe, and M. D. Samuel, "Mitigating Future Avian Malaria Threats to Hawaiian Forest Birds from Climate Change," *PLoS ONE*, vol. 12, no. 1, p. e0168880, 2017.

[3] B. Owens, "Behind New Zealand's wild plan to purge all pests," *Nature*, vol. 541, pp. 148–150, 01 2017.

[4] K. A. Oye, K. Esvelt, E. Appleton, F. Catteruccia, G. Church, T. Kuiken, S. B. Lightfoot, J. McNamara, A. Smidler, and J. P. Collins, "Biotechnology. Regulating gene drives," *Science*, vol. 345, pp. 626–628, Aug 2014.

[5] K. C. Beng, K. W. Tomlinson, X. H. Shen, Y. Surget-Groba, A. C. Hughes, R. T. Corlett, and J. W. Slik, "The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics," *Sci Rep*, vol. 6, p. 24965, 04 2016.

[6] P. D. N. Hebert, A. Cywinska, S. L. Ball, and J. R. deWaard, "Biological identifications through DNA barcodes," *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, no. 1512, pp. 313–321, 2003.

[7] V. Savolainen, R. S. Cowan, A. P. Vogler, G. K. Roderick, and R. Lane, "Towards writing the encyclopaedia of life: an introduction to DNA barcoding," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1462, pp. 1805–1811, 2005.

[8] P. TABERLET, E. COISSAC, F. POMPANON, C. BROCHMANN, and E. WILLERSLEV, "Towards next-generation biodiversity assessment using DNA metabarcoding," *Molecular Ecology*, vol. 21, pp. 2045–2050, apr 2012.

[9] K. A. Seifert, R. A. Samson, J. R. deWaard, J. Houbraken, C. A. Levesque, J.-M. Moncalvo, G. Louis-Seize, and P. D. N. Hebert, "Prospects for fungus identification using CO1 DNA barcodes, with Penicillium as a test case," *Proceedings of the National Academy of Sciences*, vol. 104, no. 10, pp. 3901–3906, 2007.

[10] M. Vences, M. Thomas, A. van der Meijden, Y. Chiari, and D. R. Vieites, "Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians," *Frontiers in zoology*, vol. 2, p. 5, 2005.

[11] P. M. Hollingsworth, L. L. Forrest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M. W. Chase, R. S. Cowan, D. L. Erickson, A. J. Fazekas, S. W. Graham, K. E. James, K.-J. Kim, W. J. Kress, H. Schneider, J. van AlphenStahl, S. C. Barrett, C. van den Berg, D. Bogarin, K. S. Burgess, K. M. Cameron, M. Carine, J. Chacon, A. Clark, J. J.

Clarkson, F. Conrad, D. S. Devey, C. S. Ford, T. A. Hedderson, M. L. Hollingsworth, B. C. Husband, L. J. Kelly, P. R. Kesanakurti, J. S. Kim, Y.-D. Kim, R. Lahaye, H.-L. Lee, D. G. Long, S. Madrinan, O. Maurin, I. Meusnier, S. G. Newmaster, C.-W. Park, D. M. Percy, G. Petersen, J. E. Richardson, G. A. Salazar, V. Savolainen, O. Seberg, M. J. Wilkinson, D.-K. Yi, and D. P. Little, "A DNA barcode for land plants," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 12794–12797, 8 2009.

[12] C. L. Schoch, K. A. Seifert, S. Huhndorf, V. Robert, J. L. Spouge, C. A. Levesque, W. Chen, E. Bolchacova, K. Voigt, P. W. Crous, A. N. Miller, M. J. Wingfield, M. C. Aime, K.-D. An, F.-Y. Bai, R. W. Barreto, D. Begerow, M.-J. Bergeron, M. Blackwell, T. Boekhout, M. Bogale, N. Boonyuen, A. R. Burgaz, B. Buyck, L. Cai, Q. Cai, G. Cardinali, P. Chaverri, B. J. Coppins, A. Crespo, P. Cubas, C. Cummings, U. Damm, Z. W. de Beer, G. S. de Hoog, R. Del-Prado, B. Dentinger, J. Dieguez-Uribeondo, P. K. Divakar, B. Douglas, M. Duenas, T. A. Duong, U. Eberhardt, J. E. Edwards, M. S. Elshahed, K. Fliegerova, M. Furtado, M. A. Garcia, Z.-W. Ge, G. W. Griffith, K. Griffiths, J. Z. Groenewald, M. Groenewald, M. Grube, M. Gryzenhout, L.-D. Guo, F. Hagen, S. Hambleton, R. C. Hamelin, K. Hansen, P. Harrold, G. Heller, C. Herrera, K. Hirayama, Y. Hirooka, H.-M. Ho, K. Hoffmann, V. Hofstetter, F. Hognabba, P. M. Hollingsworth, S.-B. Hong, K. Hosaka, J. Houbraken, K. Hughes, S. Huhtinen, K. D. Hyde, T. James, E. M. Johnson, J. E. Johnson, P. R. Johnston, E. B. G. Jones, L. J. Kelly, P. M. Kirk, D. G. Knapp, U. Koljalg, G. M. Kovacs, C. P. Kurtzman, S. Landvik, S. D. Leavitt, A. S. Liggenstoffer, K. Liimatainen, L. Lombard, J. J. Luangsaard, H. T. Lumbsch, H. Maganti, S. S. N. Maharachchikumbura, M. P. Martin, T. W. May, A. R. McTaggart, A. S. Methven, W. Meyer, J.-M. Moncalvo, S. Mongkolsamrit, L. G. Nagy, R. H. Nilsson, T. Niskanen, I. Nyilasi, G. Okada, I. Okane, I. Olariaga, J. Otte, T. Papp, D. Park, T. Petkovits, R. Pino-Bodas, W. Quaedvlieg, H. A. Raja, D. Redecker, T. L. Rintoul, C. Ruibal, J. M. Sarmiento-Ramirez, I. Schmitt, A. Schussler, C. Shearer, K. Sotome, F. O. P. Stefani, S. Stenroos, B. Stielow, H. Stockinger, S. Suetrong, S.-O. Suh, G.-H. Sung, M. Suzuki, K. Tanaka, L. Tedersoo, M. T. Telleria, E. Tretter, W. A. Untereiner, H. Urbina, C. Vagvolgyi, A. Vialle, T. D. Vu, G. Walther, Q.-M. Wang, Y. Wang, B. S. Weir, M. Weiss, M. M. White, J. Xu, R. Yahr, Z. L. Yang, A. Yurkov, J.-C. Zamora, N. Zhang, W.-Y. Zhuang, and D. Schindel, "Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 6241–6246, apr 2012.

[13] D. Steinke, M. Vences, W. Salzburger, and A. Meyer, "TaxI: a software tool for DNA barcoding using distance methods," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1462, pp. 1975–1980, 2005.

[14] S. A. Berger, D.K., A. Stamatakis, and D. Krompass, "Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood," *Systematic Biology*, vol. 60, pp. 291–302, 5 2011.

[15] F. A. Matsen, R. B. Kodner, and E. V. Armbrust, "pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree.," *BMC bioinformatics*, vol. 11, p. 538, 1 2010.

[16] S. Ratnasingham and P. D. N. Hebert, "BOLD : The Barcode of Life Data System (www.barcodinglife.org)," *Molecular Ecology Notes*, vol. 7, no. April 2016, pp. 355–364, 2007.

[17] M. J. Hickerson, C. P. Meyer, C. Moritz, and M. Hedin, "DNA Barcoding Will Often Fail to Discover New Animal Species over Broad Parameter Space," *Systematic Biology*, vol. 55, pp. 729–739, 10 2006.

[18] D. L. J. Quicke, M. Alex Smith, D. H. Janzen, W. Hallwachs, J. Fernandez-Triana, N. M. Laurenne, A. Zald??var-River??n, M. R. Shaw, G. R. Broad, S. Klopfstein, S. R. Shaw, J. Hrcek, P. D. N. Hebert, S. E. Miller, J. J. Rodriguez, J. B. Whitfield, M. J. Sharkey, B. J. Sharanowski, R. Jussila, I. D. Gauld, D. Chesters, and A. P. Vogler, "Utility of the DNA barcoding gene fragment for parasitic wasp phylogeny (Hymenoptera: Ichneumonoidea): Data release and new measure of taxonomic congruence," *Molecular Ecology Resources*, vol. 12, pp. 676–685, 7 2012.

[19] E. Coissac, P. M. Hollingsworth, S. Lavergne, and P. Taberlet, "From barcodes to genomes: Extending the concept of DNA barcoding," 2016.

[20] S. Liu, Y. Li, J. Lu, X. Su, M. Tang, R. Zhang, L. Zhou, C. Zhou, Q. Yang, Y. Ji, D. W. Yu, and X. Zhou, "SOAP Barcode: revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons," *Methods in Ecology and Evolution*, vol. 4, pp. 1142–1150, dec 2013.

[21] "France Génomique - Mutualisation des compétences et des équipements français pour lanalyse génomique et la bio-informatique." `https://www.france-genomique.org/`.

[22] "Norwegian Barcode of Life (NorBOL)." `http://www.norbol.org/en/`.

[23] "DNAmark." `http://dnamark.ku.dk/english/`.

[24] B. E. Blaisdell, "A measure of the similarity of sets of sequences not requiring sequence alignment.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 14, pp. 5155–5159, 1986.

[25] S. Vinga and J. Almeida, "Alignment-free sequence comparison–a review," *Bioinformatics*, vol. 19, pp. 513–523, 3 2003.

[26] G. Reinert, D. Chew, F. Sun, and M. S. Waterman, "Alignment-free sequence comparison (I): statistics and power.," *Journal of computational biology : a journal of computational molecular cell biology*, vol. 16, pp. 1615–34, 12 2009.

[27] K. Song, J. Ren, Z. Zhai, X. Liu, M. Deng, and F. Sun, "Alignment-Free Sequence Comparison Based on Next-Generation Sequencing Reads," *Journal of Computational Biology*, vol. 20, pp. 64–79, 2 2013.

[28] G. Marçais and C. Kingsford, "A fast, lock-free approach for efficient parallel counting of occurrences of k-mers," *Bioinformatics*, vol. 27, pp. 764–770, 3 2011.

[29] D. E. Wood and S. L. Salzberg, "Kraken: ultrafast metagenomic sequence classification using exact alignments," *Genome Biology*, vol. 15, no. 3, p. R46, 2014.

[30] B. D. Ondov, T. J. Treangen, P. Melsted, A. B. Mallonee, N. H. Bergman, S. Koren, and A. M. Phillippy, "Mash: fast genome and metagenome distance estimation using MinHash," *Genome Biology*, vol. 17, p. 132, 12 2016.

[31] E. D. Jarvis, S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. W. Ho, B. C. Faircloth, B. Nabholz, and J. T. Howard, "Phylogenomic analyses data of the avian phylogenomics project," *GigaScience*, vol. 4, no. 1, p. 4, 2015.

[32] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[33] R. C. Edgar, "Search and clustering orders of magnitude faster than BLAST," *Bioinformatics*, vol. 26, no. 19, pp. 2460–2461, 2010.

[34] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," 2012.

[35] D. A. Petrov, "Mutational Equilibrium Model of Genome Size Evolution," *Theoretical Population Biology*, vol. 61, no. 4, pp. 531–544, 2002.

[36] H. Fan, A. R. Ives, Y. Surget-Groba, and C. H. Cannon, "An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data," *BMC Genomics*, vol. 16, p. 522, 7 2015.

[37] P. Buneman, "A note on the metric properties of trees," *Journal of Combinatorial Theory, Series B*, vol. 17, no. 1, pp. 48–50, 1974.

[38] C. E. Hinchliff, S. a. Smith, J. F. Allman, J. G. Burleigh, R. Chaudhary, L. M. Coghill, K. A. Crandall, J. Deng, B. T. Drew, R. Gazis, K. Gude, D. S. Hibbett, L. a. Katz, H. D. Laughinghouse, E. J. McTavish, P. E. Midford, C. L. Owen, R. H. Ree, J. a. Rees, D. E. Soltis, T. L. Williams, and K. a. Cranston, "Synthesis of phylogeny and taxonomy into a comprehensive tree of life," *Proceedings of the National Academy of Sciences*, 2015.

[39] T. H. Jukes and C. R. Cantor, "Evolution of protein molecules," in *In Mammalian protein metabolism, Vol. III (1969), pp. 21-132*, vol. III, pp. 21–132, 1969.

[40] M. Kimura, "A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences," *Journal of Molecular Evolution*, vol. 16, no. 2, pp. 111–120, 1980.

[41] S. Tavaré, "Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences," *Lectures on Mathematics in the Life Sciences*, vol. 17, pp. 57–86, 1986.

[42] F. Rodríguez, J. Oliver, A. Marín, and J. Medina, "The general stochastic model of nucleotide substitution," *Journal of Theoretical Biology*, vol. 142, pp. 485–501, 2 1990.

[43] P. Erdos, M. Steel, L. Szekely, and T. Warnow, "A few logs suffice to build (almost) all trees: Part II," *Theoretical Computer Science*, vol. 221, no. 1-2, pp. 77–118, 1999.

[44] J. A. Lake, "Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, pp. 1455–9, 2 1994.

[45] R. Desper and O. Gascuel, "Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle," *Journal of computational biology*, vol. 9, no. 5, pp. 687–705, 2002.

[46] N. Saitou and M. Nei, "The neighbour-joining method: a new method for reconstructing phylogenetic trees," *Molecular biology and evolution*, vol. 4, no. 4, pp. 406–425, 1987.

[47] D. H. Huson, S. M. Nettles, and T. J. Warnow, "Disk-covering, a fast-converging method for phylogenetic tree reconstruction.," *Journal of computational biology*, vol. 6, no. 3-4, pp. 369–86, 1999.

[48] V. Lefort, R. Desper, and O. Gascuel, "FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program," *Molecular Biology and Evolution*, vol. 32, no. 10, pp. 2798–2800, 2015.

[49] A. Criscuolo and O. Gascuel, "Fast NJ-like algorithms to deal with incomplete distance matrices.," *BMC bioinformatics*, vol. 9, no. 1, p. 166, 2008.

[50] T. J. Wheeler, "Large-scale neighbor-joining with NINJA," in *Algorithms in Bioinformatics*, pp. 375–389, Springer, 2009.

[51] S. Roch, "Toward Extracting All Phylogenetic Information from Matrices of Evolutionary Distances," *Science*, vol. 327, no. 5971, pp. 1376–1379, 2010.

[52] J. L. Thorne and H. Kishino, "Freeing phylogenies from artifacts of alignment.," *Molecular biology and evolution*, vol. 9, pp. 1148–62, 11 1992.

[53] M. Höhl and M. A. Ragan, "Is multiple-sequence alignment required for accurate inference of phylogeny?," *Systematic Biology*, vol. 56, no. 2, pp. 206–221, 2007.

[54] C. Daskalakis and S. Roch, "Alignment-free phylogenetic reconstruction: Sample complexity via a branching process analysis," *Annals of Applied Probability*, vol. 23, no. 2, pp. 693–721, 2013.

[55] B. E. Blaisdell, "A measure of the similarity of sets of sequences not requiring sequence alignment.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, pp. 5155–9, 7 1986.

[56] Q. Dai, Y. Yang, and T. Wang, "Markov model plus k-word distributions: a synergy that produces novel statistical measures for sequence comparison," *Bioinformatics*, vol. 24, pp. 2296–2302, 10 2008.

[57] K. Yang and L. Zhang, "Performance comparison between k-tuple distance and four model-based distances in phylogenetic tree reconstruction," *Nucleic Acids Research*, vol. 36, pp. e33–e33, 1 2008.

[58] J. Qi, H. Luo, and B. Hao, "CVTree: a phylogenetic tree reconstruction tool based on whole genomes," *Nucleic Acids Research*, vol. 32, pp. W45–W47, 7 2004.

[59] I. Ulitsky, D. Burstein, T. Tuller, and B. Chor, "The Average Common Substring Approach to Phylogenomic Reconstruction," *Journal of Computational Biology*, vol. 13, pp. 336–350, 3 2006.

[60] H. Yi and L. Jin, "Co-phylog: an assembly-free phylogenomic approach for closely related organisms," *Nucleic Acids Research*, vol. 41, pp. e75–e75, 4 2013.

[61] T. Roychowdhury, A. Vishnoi, and A. Bhattacharya, "Next-Generation Anchor Based Phylogeny (NexABP): Constructing phylogeny from Next-generation sequencing data," *Scientific Reports*, vol. 3, p. 2634, 12 2013.

[62] M. Domazet-Lošo and B. Haubold, "Alignment-free detection of local similarity among viral and bacterial genomes," *Bioinformatics*, vol. 27, pp. 1466–1472, 6 2011.

[63] B. Haubold, "Alignment-free phylogenetics and population genetics," *Briefings in Bioinformatics*, vol. 15, pp. 407–418, 5 2014.

[64] S.-R. Jun, G. E. Sims, G. A. Wu, and S.-H. Kim, "Whole-proteome phylogeny of prokaryotes by feature frequency profiles: An alignment-free method with optimal feature resolution," *Proceedings of the National Academy of Sciences*, vol. 107, no. 1, pp. 133–138, 2010.

[65] G. A. Wu, S.-R. Jun, G. E. Sims, and S.-H. Kim, "Whole-proteome phylogeny of large dsDNA virus families by an alignment-free method," *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 12826–12831, 2009.

[66] E. S. Allman, J. A. Rhodes, and S. Sullivant, "Statistically Consistent k -mer Methods for Phylogenetic Tree Reconstruction," *Journal of Computational Biology*, vol. 24, pp. 153–171, 2 2017.

[67] H. Fan, A. R. Ives, Y. Surget-Groba, and C. H. Cannon, "An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data," *BMC Genomics*, vol. 16, p. 522, 12 2015.

[68] L. Nakhleh, U. Roshan, K. St John, J. Sun, and T. Warnow, "Designing fast converging phylogenetic methods.," *Bioinformatics (Oxford, England)*, vol. 17 Suppl 1, no. 1, pp. S190–S198, 2001.

[69] R. R. Hudson, "Generating samples under a Wright-Fisher neutral model of genetic variation," *Bioinformatics*, vol. 18, no. 2, pp. 337–338, 2002.

[70] R. C. Edgar, G. Asimenos, S. Batzoglou, and A. Sidow, "Evolver," 2017.

[71] D. Earl, K. Bradnam, J. St. John, A. Darling, D. Lin, J. Fass, H. O. K. Yu, V. Buffalo, D. R. Zerbino, M. Diekhans, N. Nguyen, P. N. Ariyaratne, W. K. Sung, Z. Ning, M. Haimel, J. T. Simpson, N. A. Fonseca, I. Birol, T. R. Docking, I. Y. Ho, D. S. Rokhsar, R. Chikhi, D. Lavenier, G. Chapuis, D. Naquin, N. Maillet, M. C. Schatz, D. R. Kelley, A. M. Phillippy, S. Koren, S. P. Yang, W. Wu, W. C. Chou, A. Srivastava, T. I. Shaw, J. G. Ruby, P. Skewes-Cox, M. Betegon, M. T. Dimon, V. Solovyev, I. Seledtsov, P. Kosarev, D. Vorobyev, R. Ramirez-Gonzalez, R. Leggett, D. MacLean, F. Xia, R. Luo, Z. Li, Y. Xie, B. Liu, S. Gnerre, I. MacCallum, D. Przybylski, F. J. Ribeiro, T. Sharpe, G. Hall, P. J. Kersey, R. Durbin, S. D. Jackman, J. A. Chapman, X. Huang, J. L. DeRisi, M. Caccamo, Y. Li, D. B. Jaffe, R. E. Green, D. Haussler, I. Korf, and B. Paten, "Assemblathon 1: A competitive assessment of de novo short read assembly methods," 2011.

[72] E. Sayyari, J. B. Whitfield, and S. Mirarab, "Fragmentary Gene Sequences Negatively Impact Gene Tree and Species Tree Reconstruction," *Molecular Biology and Evolution*, vol. advanced p, pp. msx261–msx261, 2017.

[73] S. Mirarab, M. S. Bayzid, B. Boussau, and T. Warnow, "Statistical binning enables an accurate coalescent-based estimation of the avian tree," *Science*, vol. 346, pp. 1250463–1250463, 12 2014.

[74] C. Zhang, E. Sayyari, and S. Mirarab, "ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (J. Meidanis and L. Nakhleh, eds.), vol. 10562 LNBI, pp. 53–75, Cham: Springer International Publishing, 2017.

[75] S. Mirarab, M. S. Bayzid, and T. Warnow, "Evaluating Summary Methods for Multilocus Species Tree Estimation in the Presence of Incomplete Lineage Sorting," *Systematic Biology*, vol. 65, pp. 366–380, 5 2016.

[76] W. P. Maddison, "Gene Trees in Species Trees," *Systematic Biology*, vol. 46, pp. 523–536, 9 1997.

[77] J. H. Degnan and N. A. Rosenberg, "Discordance of Species Trees with Their Most Likely Gene Trees," *PLoS Genetics*, vol. 2, 5 2006.

[78] L. Nakhleh, "Evolutionary Phylogenetic Networks: Models and Issues," *Networks*, pp. 125–158, 2011.

[79] G. J. Szöllsi, E. Tannier, V. Daubin, and B. Boussau, "The inference of gene trees with species trees," *Systematic Biology*, vol. 64, pp. e42–e62, 7 2014.

[80] S. Mirarab, R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow, "ASTRAL: genome-scale coalescent-based species tree estimation," *Bioinformatics*, vol. 30, pp. i541–i548, 9 2014.

[81] S. Mirarab and T. Warnow, "ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes," *Bioinformatics*, vol. 31, pp. i44–i52, 6 2015.

[82] Y. Wu, "Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood," *Evolution*, vol. 66, no. 3, pp. 763–775, 2012.

[83] L. Liu and L. Yu, "Estimating species trees from unrooted gene trees," *Systematic Biology*, vol. 60, pp. 661–667, 2011.

[84] 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*, vol. 11, p. 574, 1 2010.

[85] L. Liu, L. Yu, and S. V. Edwards, "A maximum pseudo-likelihood approach for estimating species trees under the coalescent model," *BMC Evolutionary Biology*, vol. 10, no. 1, p. 302, 2010.

[86] D. Bryant, R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. Roychoudhury, "Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis," *Molecular Biology and Evolution*, vol. 29, no. 8, pp. 1917–1932, 2012.

[87] J. Chifman and L. S. Kubatko, "Quartet Inference from SNP Data Under the Coalescent Model," *Bioinformatics*, vol. 30, pp. 3317–3324, 8 2014.

[88] J. Heled and A. J. Drummond, "Bayesian inference of species trees from multilocus data.," *Molecular Biology and Evolution*, vol. 27, pp. 570–580, 3 2010.

[89] B. Boussau, G. J. Szöllősi, and L. Duret, "Genome-scale coestimation of species and gene trees," *Genome Research*, vol. 23, pp. 323–330, 12 2013.

[90] Y.-C. Wu, M. D. Rasmussen, M. S. Bansal, and M. Kellis, "TreeFix: Statistically Informed Gene Tree Error Correction Using Species Trees," *Systematic Biology*, vol. 62, no. 1, pp. 110–120, 2013.

[91] G. Dasarathy, E. Mossel, R. Nowak, and S. Roch, "Coalescent-based species tree estimation: a stochastic Farris transform," vol. 1149312, 2017.

[92] G. Dasarathy, R. Nowak, and S. Roch, "Data requirement for phylogenetic inference from multiple loci: a new distance method," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 12, no. 2, pp. 422–432, 2015.

[93] C. Durden and S. Sullivant, "Identifiability of phylogenetic parameters from k-mer data under the coalescent," 5 2017.

[94] P. Pamilo and M. Nei, "Relationships between gene trees and species trees," *Molecular biology and evolution*, vol. 5, no. 5, pp. 568–583, 1988.

[95] S. Maere, S. De Bodt, J. Raes, T. Casneuf, M. Van Montagu, M. Kuiper, and Y. Van de Peer, "Modeling gene and genome duplications in eukaryotes," *Proceedings of the National Academy of Sciences*, vol. 102, no. 15, pp. 5454–5459, 2005.

[96] H. Innan and F. Kondrashov, "The evolution of gene duplications: classifying and distinguishing between models," *Nature Reviews Genetics*, vol. 11, no. 4, p. 4, 2010.

[97] D. Sankoff, C. Zheng, and Q. Zhu, "Polyploids, genome halving and phylogeny," in *Bioinformatics*, vol. 23, 2007.

[98] C. Chauve, N. El-Mabrouk, L. Guéguen, M. Semeria, and E. Tannier, "Duplication, Rearrangement and Reconciliation: A Follow-Up 13 Years Later," in *Models and Algorithms for Genome Evolution* (C. Chauve, N. El-Mabrouk, and E. Tannier, eds.), vol. 19 of *Computational Biology*, pp. 47–62, London: Springer London, 2013.

[99] S. Mirarab, N. Nguyen, and T. Warnow, "SEPP: SATé-Enabled Phylogenetic Placement.," *Pacific Symposium On Biocomputing*, pp. 247–58, 2012.

[100] F. A. Matsen, "Phylogenetics and the human microbiome," *Systematic Biology*, vol. 64, no. 1, pp. e26–e41, 2014.

[101] E. Sayyari and S. Mirarab, "Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies," *Molecular Biology and Evolution*, vol. 33, pp. 1654–1668, 7 2016.

[102] L. Salichos and A. Rokas, "Inferring ancient divergences requires genes with strong phylogenetic signals," *Nature*, vol. 497, no. 7449, pp. 327–31, 2013.

[103] S. Guindon, F. Delsuc, J.-F. Dufayard, and O. Gascuel, "Estimating maximum likelihood phylogenies with PhyML," *Methods in Molecular Biology*, vol. 537, pp. 113–137, 2009.

[104] B. Liu, T. Gibbons, M. Ghodsi, and M. Pop, "MetaPhyler: Taxonomic profiling for metagenomic sequences," in *Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on*, p. 95100, IEEE, 2011.

[105] N.-p. Nguyen, S. Mirarab, B. Liu, M. Pop, and T. Warnow, "TIPP: taxonomic identification and phylogenetic profiling," *Bioinformatics*, vol. 30, pp. 3548–3555, 12 2014.

[106] N. Segata, L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson, and C. Huttenhower, "Metagenomic microbial community profiling using unique clade-specific marker genes," *Nature Methods*, vol. 9, no. 8, pp. 811–814, 2012.

[107] A. C. McHardy, H. G. Martín, A. Tsirigos, P. Hugenholtz, and I. Rigoutsos, "Accurate phylogenetic classification of variable-length DNA fragments.," *Nature methods*, vol. 4, pp. 63–72, 1 2007.

[108] K. R. Patil, P. Haider, P. B. Pope, P. J. Turnbaugh, M. Morrison, T. Scheffer, and A. C. McHardy, "Taxonomic metagenome sequence assignment with structured output models," *Nature methods*, vol. 8, no. 3, pp. 191–192, 2011.

[109] A. Brady and S. L. Salzberg, "Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models.," *Nature methods*, vol. 6, pp. 673–6, 9 2009.

[110] G. L. Rosen, E. R. Reichenberger, and A. M. Rosenfeld, "NBC: The naive Bayes classification tool webserver for taxonomic classification of metagenomic reads," *Bioinformatics*, vol. 27, no. 1, pp. 127–129, 2011.

[111] E. D. Jarvis, S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. W. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, A. Suh, C. C. Weber, R. R. da Fonseca, J. Li, F. Zhang, H. Li, L. Zhou, N. Narula, L. Liu, G. Ganapathy, B. Boussau, M. S. Bayzid, V. Zavidovych, S. Subramanian, T. Gabaldón, S. Capella-Gutiérrez, J. Huerta-Cepas, B. Rekepalli, K. Munch, M. H. Schierup, B. Lindow, W. C. Warren, D. Ray, R. E. Green, M. W. Bruford, X. Zhan, A. Dixon, S. Li, N. Li, Y. Huang, E. P. Derryberry, M. F. Bertelsen, F. H. Sheldon, R. T. Brumfield, C. V. Mello, P. V. Lovell, M. Wirthlin, M. P. C. Schneider, F. Prosdocimi, J. A. Samaniego, A. M. V. Velazquez, A. Alfaro-Núñez, P. F. Campos, B. Petersen, T. Sicheritz-Ponten, A. Pas, T. Bailey, P. Scofield, M. Bunce, D. M. Lambert, Q. Zhou, P. Perelman, A. C. Driskell, B. Shapiro, Z. Xiong, Y. Zeng, S. Liu, Z. Li, B. Liu, K. Wu, J. Xiao, X. Yinqi, Q. Zheng, Y. Zhang, H. Yang, J. Wang, L. Smeds, F. E. Rheindt, M. J. Braun, J. Fjeldså, L. Orlando, F. K. Barker, K. A. Jønsson, W. Johnson, K.-P. Koepfli, S. OBrien, D. Haussler, O. A. Ryder, C. Rahbek, E. Willerslev, G. R. Graves, T. C. Glenn, J. E. McCormack, D. W. Burt, H. Ellegren, P. Alström, S. V. Edwards, A. Stamatakis, D. P. Mindell, J. Cracraft,

E. L. Braun, T. Warnow, W. Jun, M. T. P. Gilbert, and G. Zhang, "Whole-genome analyses resolve early branches in the tree of life of modern birds," *Science*, vol. 346, pp. 1320–1331, 12 2014.

[112] N. J. Wickett, S. Mirarab, N. Nguyen, T. Warnow, E. J. Carpenter, N. Matasci, S. Ayyampalayam, M. S. Barker, J. G. Burleigh, M. A. Gitzendanner, B. R. Ruhfel, E. Wafula, J. P. Der, S. W. Graham, S. Mathews, M. Melkonian, D. E. Soltis, P. S. Soltis, N. W. Miles, C. J. Rothfels, L. Pokorny, A. J. Shaw, L. DeGironimo, D. W. Stevenson, B. Surek, J. C. Villarreal, B. Roure, H. Philippe, C. W. DePamphilis, T. Chen, M. K. Deyholos, R. S. Baucom, T. M. Kutchan, M. M. Augustin, J. Wang, Y. Zhang, Z. Tian, Z. Yan, X. Wu, X. Sun, G. K.-S. Wong, and J. Leebens-Mack, "Phylotranscriptomic analysis of the origin and early diversification of land plants," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 4859–4868, 10 2014.

[113] K. M. Turner, V. Deshpande, D. Beyter, T. Koga, J. Rusert, C. Lee, B. Li, K. Arden, B. Ren, D. A. Nathanson, H. I. Kornblum, M. D. Taylor, S. Kaushal, W. K. Cavenee, R. Wechsler-Reya, F. B. Furnari, S. R. Vandenberg, P. N. Rao, G. M. Wahl, V. Bafna, and P. S. Mischel, "Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity," *Nature*, vol. 543, pp. 122–125, 03 2017.

[114] W. K. Chu, P. Edge, H. S. Lee, V. Bansal, V. Bafna, X. Huang, and K. Zhang, "Ultraaccurate genome sequencing and haplotyping of single human cells," *Proc. Natl. Acad. Sci. U.S.A.*, Oct 2017.

[115] P. Edge, V. Bafna, and V. Bansal, "HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies," *Genome Res.*, vol. 27, pp. 801–812, May 2017.

[116] A. Patel, R. Schwab, Y. T. Liu, and V. Bafna, "Amplification and thrifty single-molecule sequencing of recurrent somatic structural variations," *Genome Res.*, Jan 2014.

[117] R. Ronen, D. Zhou, V. Bafna, and G. G. Haddad, "The genetic basis of chronic mountain sickness," *Physiology (Bethesda)*, vol. 29, pp. 403–412, Nov 2014.

[118] R. Ronen, G. Tesler, A. Akbari, S. Zakov, N. A. Rosenberg, and V. Bafna, "Predicting Carriers of Ongoing Selective Sweeps without Knowledge of the Favored Allele," *PLoS Genet.*, vol. 11, p. e1005527, Sep 2015.

[119] C. Lo, S. Kim, S. Zakov, and V. Bafna, "Evaluating genome architecture of a complex region via generalized bipartite matching," *BMC Bioinformatics*, vol. 14 Suppl 5, p. S13, 2013. PMID:23734567. Free article.

[120] R. Ronen, N. Udpa, E. Halperin, and V. Bafna, "Learning natural selection from the site frequency spectrum," *Genetics*, vol. 195, pp. 181–193, Sep 2013.

[121] D. Zhou, N. Udpa, R. Ronen, T. Stobdan, J. Liang, O. Appenzeller, H. W. Zhao, Y. Yin, Y. Du, L. Guo, R. Cao, Y. Wang, X. Jin, C. Huang, W. Jia, D. Cao, G. Guo, J. L. Gamboa, F. Villafuerte, D. Callacondo, J. Xue, S. Liu, K. A. Frazer, Y. Li, V. Bafna, and G. G. Haddad, "Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders," *Am. J. Hum. Genet.*, vol. 93, pp. 452–462, Sep 2013.

[122] C. Lo, R. Liu, J. Lee, K. Robasky, S. Byrne, C. Lucchesi, J. Aach, G. Church, V. Bafna, and K. Zhang, "On the design of clone-based haplotyping," *Genome Biol.*, vol. 14, p. R100, Sep 2013.

[123] N. Udpa, R. Ronen, D. Zhou, J. Liang, T. Stobdan, O. Appenzeller, Y. Yin, Y. Du, L. Guo, R. Cao, Y. Wang, X. Jin, C. Huang, W. Jia, D. Cao, G. Guo, V. E. Claydon, R. Hainsworth, J. L. Gamboa, M. Zibenigus, G. Zenebe, J. Xue, S. Liu, K. A. Frazer, Y. Li, V. Bafna, and G. G. Haddad, "Whole genome sequencing of Ethiopian highlanders reveals conserved hypoxia tolerance genes," *Genome Biol.*, vol. 15, p. R36, Feb 2014.

[124] M. Kramer, J. Dutkowski, M. Yu, V. Bafna, and T. Ideker, "Inferring gene ontologies from pairwise similarity data," *Bioinformatics*, vol. 30, pp. i34–i42, Jun 2014.

[125] M. Kinsella, A. Patel, and V. Bafna, "The elusive evidence for chromothripsis," *Nucleic Acids Res.*, Jun 2014.

[126] H. S. Chen, C. M. Hutter, L. E. Mechanic, C. I. Amos, V. Bafna, E. R. Hauser, R. D. Hernandez, C. Li, D. A. Liberles, K. McAllister, J. H. Moore, D. N. Paltoo, G. J. Papanicolaou, B. Peng, M. D. Ritchie, G. Rosenfeld, J. S. Witte, E. M. Gillanders, and E. J. Feuer, "Genetic simulation tools for post-genome wide association studies of complex diseases," *Genet. Epidemiol.*, vol. 39, pp. 11–19, Jan 2015.

[127] S. Zakov and V. Bafna, "Reconstructing breakage fusion bridge architectures using noisy copy numbers," *J. Comput. Biol.*, vol. 22, pp. 577–594, Jun 2015.

[128] E. L. Flannery, T. Wang, A. Akbari, V. C. Corey, F. Gunawan, A. T. Bright, M. Abraham, J. F. Sanchez, M. L. Santolalla, G. C. Baldeviano, K. A. Edgel, L. A. Rosales, A. G. Lescano, V. Bafna, J. M. Vinetz, and E. A. Winzeler, "Next-Generation Sequencing of Plasmodium vivax Patient Samples Shows Evidence of Direct Evolution in Drug-Resistance Genes," *ACS Infect Dis*, vol. 1, pp. 367–379, Aug 2015.

[129] D. Beyter, P. Z. Tang, S. Becker, T. Hoang, D. Bilgin, Y. W. Lim, T. C. Peterson, S. Mayfield, F. Haerizadeh, J. B. Shurin, V. Bafna, and R. McBride, "Diversity, Productivity, and Stability of an Industrial Microbial Ecosystem," *Appl. Environ. Microbiol.*, vol. 82, pp. 2494–2505, Apr 2016.

[130] A. Patel, P. Edge, S. Selvaraj, V. Bansal, and V. Bafna, "InPhaDel: integrative shotgun and proximity-ligation sequencing to phase deletions with single nucleotide polymorphisms," *Nucleic Acids Res.*, Apr 2016.

[131] T. Stobdan, D. Zhou, E. Ao-Ieong, D. Ortiz, R. Ronen, I. Hartley, Z. Gan, A. D. McCulloch, V. Bafna, P. Cabrales, and G. G. Haddad, "Endothelin receptor B, a candidate gene from human studies at high altitude, improves cardiac tolerance to hypoxia in genetically engineered heterozygote mice," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, pp. 10425–10430, Aug 2015.

[132] P. Azad, H. W. Zhao, P. J. Cabrales, R. Ronen, D. Zhou, O. Poulsen, O. Appenzeller, Y. H. Hsiao, V. Bafna, and G. G. Haddad, "Senp1 drives hypoxia-induced polycythemia via GATA1 and Bcl-xL in subjects with Monge's disease," *J. Exp. Med.*, vol. 213, pp. 2729–2744, Nov 2016.

[133] R. Ronen, G. Tesler, A. Akbari, S. Zakov, N. A. Rosenberg, and V. Bafna, "Correction: Predicting Carriers of Ongoing Selective Sweeps without Knowledge of the Favored Allele," *PLoS Genet.*, vol. 12, p. e1006472, Nov 2016.

[134] A. Iranmehr, A. Akbari, C. Schlotterer, and V. Bafna, "Clear: Composition of Likelihoods for Evolve and Resequence Experiments," *Genetics*, vol. 206, pp. 1011–1023, Jun 2017.

[135] P. Azad, T. Stobdan, D. Zhou, I. Hartley, A. Akbari, V. Bafna, and G. G. Haddad, "High-altitude adaptation in humans: from genomics to integrative physiology," *J. Mol. Med.*, Sep 2017.

[136] T. Stobdan, A. Akbari, P. Azad, D. Zhou, O. Poulsen, O. Appenzeller, G. F. Gonzales, A. Telenti, E. H. M. Wong, S. Saini, E. F. Kirkness, J. C. Venter, V. Bafna, and G. G. Haddad, "New insights into the genetic basis of Monge's disease and adaptation to high-altitude," *Mol. Biol. Evol.*, Sep 2017.

[137] E. Sayyari and S. Mirarab, "Anchoring quartet-based phylogenetic distances and applications to species tree reconstruction," *BMC Genomics*, vol. 17, pp. 101–113, 11 2016.

[138] U. Mai, E. Sayyari, and S. Mirarab, "Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction," *PLOS ONE*, vol. 12, p. e0182238, 8 2017.

[139] S. Shekhar, S. Roch, and S. Mirarab, "Species tree estimation using ASTRAL: how many genes are enough?," *Proceedings of International Conference on Research in Computational Molecular Biology (RECOMB)*, no. to appear, 2017.

[140] U. Mai and S. Mirarab, "TreeShrink: Efficient Detection of Outlier Tree Leaves," in *Comparative Genomics: 15th International Workshop, RECOMB CG 2017, Barcelona, Spain, October 4-6, 2017, Proceedings* (J. Meidanis and L. Nakhleh, eds.), pp. 116–140, Cham: Springer International Publishing, 2017.

[141] N. Moshiri and S. Mirarab, "A Two-State Model of Tree Evolution and its Applications to Alu Retrotransposition," *Systematic Biology*, p. in press, 2017.