

MSCquartets 3.3: Fast Inference of Species Networks and Trees of Blobs under the Multispecies Coalescent

John A. Rhodes,¹ Elizabeth S. Allman,¹ Cécile Ané,² Hector Baños,^{3*} Jonathan D. Mitchell^{4,5}
and Kristina Wicke^{6,7}

¹Department of Mathematics and Statistics, University of Alaska Fairbanks, Fairbanks, AK, USA

²Department of Statistics, University of Wisconsin, Madison, Madison, WI, USA

³Department of Mathematics, California State University San Bernardino, San Bernardino, CA, USA

⁴School of Natural Sciences, University of Tasmania, Hobart, TAS, Australia

⁵Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture, University of Tasmania, Hobart, TAS, Australia

⁶Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ, USA

⁷National Institute for Theory and Mathematics in Biology, Northwestern University and The University of Chicago Chicago, IL, USA

*Corresponding author. hector.banos@csusb.edu

Abstract

The R package `MSCquartets` offers a suite of methods for testing and inferring species relationships from gene trees under coalescent models. `MSCquartets` 3.3 significantly extends the functionality of version 1.0 by implementing the `TINNiK` and `ECToBlob` algorithms for inference of the tree of blobs under the network multispecies coalescent model (NMSC). Additionally, it implements the `NANUQ+` algorithm, a fast method for inference of level-1 networks. Computational speed of all functionalities is greatly improved through use of C++ routines, and a Shiny App is introduced for easier exploration of basic analyses.

Availability and implementation: `MSCquartets` 3.3 is available on the Comprehensive R Archive Network, at <https://CRAN.R-project.org/package=MSCquartets>. The `MSCquartets` 3.3 Shiny App is available at <https://mscquartets.shinyapps.io/MScq-app/>

Keywords phylogenetic network, admixture graph, network multispecies coalescent, hybridization, gene flow, admixture

Abbreviations: MSC (multispecies coalescent model); NMSC (network multispecies coalescent model); qcCF (quartet count concordance factor)

Introduction

Phylogenetic networks are increasingly recognized as necessary to represent evolutionary histories of species and populations. These depict reticulate processes such as hybridization, introgression, admixture, and other forms of lateral gene transfer that a single tree cannot capture. However, inferring such networks from genomic data remains challenging. Incomplete lineage sorting (ILS), a population-genetic process causing gene genealogies to differ from the underlying species history, interferes with the network signal, leading to both statistical and computational difficulties.

The R package `MSCquartets` 1.0 [Rhodes et al., 2020] introduced a suite of methods for inferring species trees and networks under the tree and network multispecies coalescent (MSC and NMSC) using quartet-based summaries of gene trees, that is, unrooted topological gene trees on subsets of 4 taxa. `MSCquartets` 3.3 represents a substantial expansion of this framework, particularly in the context of non-tree-like evolutionary processes.

Notably, `MSCquartets` 3.3 implements the `TINNiK` [Allman et al., 2024] and `ECToBlob` [Rhodes et al., 2026] algorithms for statistically

consistent inference of trees of blobs of networks with quite general form. The tree of blobs of a network is an unrooted topological tree that partially represents species relationships by contracting reticulate substructures into multifurcations while preserving the network's cut edges, which form its tree-like parts. Standard phylogenetic methods cannot recover the trees of blobs or networks directly. For example, `ASTRAL` has been shown to be potentially misleading in the presence of hybridization [Leaché et al., 2014, Solís-Lemus et al., 2016, Long and Kubatko, 2018, Pang and Zhang, 2022, Dinh and Baños, 2025], and practical network inference methods generally require unverifiable structural assumptions on the network.

`MSCquartets` 3.3 also implements the `NANUQ+` [Allman et al., 2025] algorithm for resolving a tree of blobs into a level-1 phylogenetic network. `NANUQ+` refines a tree of blobs by resolving its multifurcations into cycles using a divide-and-conquer strategy. This yields a fast and statistically consistent estimator of level-1 phylogenetic networks under the NMSC. `NANUQ+` directly returns a level-1 phylogenetic network, whereas `NANUQ` produces a splits graph that must be interpreted by the user as a phylogenetic network using additional reconstruction rules.

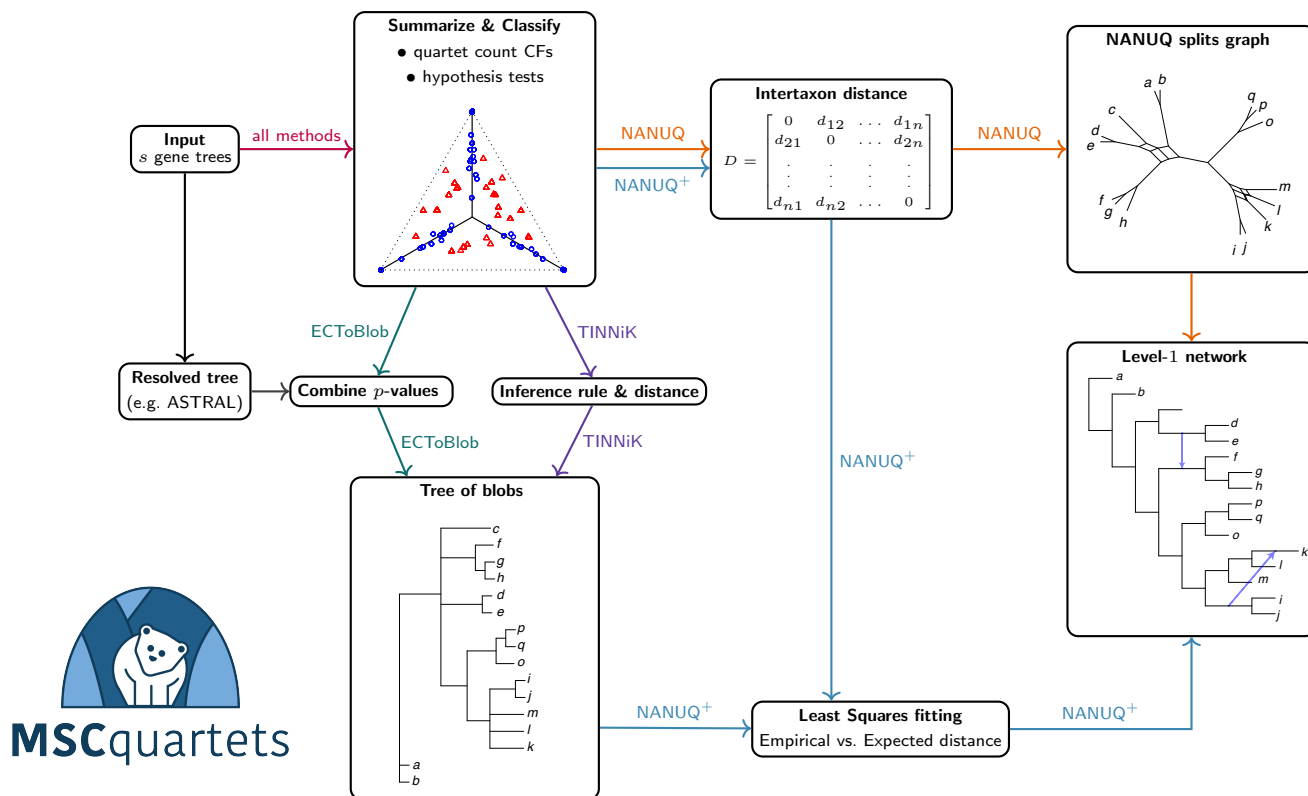


Figure 1 Schematic overview of the core functionalities provided by the MSCquartets 3.3 package.

Figure 1 provides an overview of the main network functionalities implemented in MSCquartets 3.3. These developments broaden the range of evolutionary models that can be explored within a unified quartet-based framework.

Existing functionality has also been significantly accelerated through recoding certain key routines in C++, improving the efficiency of previously implemented methods.

Finally, we developed the MSCquartets Shiny application, providing a web-based interface for key MSCquartets 3.3 functionality which allows users to explore and apply these methods without a local R installation or previous R knowledge.

Materials and Methods

Speed performance improvement

MSCquartets 3.3 uses the Rcpp R-package [Eddelbuettel and François, 2011] for fundamental C++ routines to summarize gene tree data by quartet count Concordance Factors (qcCFs), decreasing runtime substantially. For example, the function `quartetTable()` in MSCquartets v. 1.0 took 15,870 seconds to tabulate all qcCFs for a 23,136 gene tree *Leopardus* dataset [Lescroart et al., 2023] on 16 taxa. This same function in v. 3.3 took only 18.1 seconds, a reduction in computational time by 3 orders of magnitude. (Here and below, all runtimes are reported for an Apple M2 Pro with 16 GB of RAM.)

TINNiK

TINNiK, introduced in Allman et al. [2024], is a statistically consistent algorithm for inferring the tree of blobs of a network

under the NMSC. Input is a collection of gene trees summarized by their qcCFs. A hypothesis test developed in Mitchell et al. [2019] is applied to these to determine the topology of all induced 4-taxon trees of blobs, with a quartet inference rule described in Allman et al. [2023] used to determine larger blobs on the full network. The quartet intertaxon distance of Rhodes [2020] is then computed, and the tree of blobs obtained using a distance method, such as FastME [Desper and Gascuel, 2002] or Neighbor-Joining [Saitou and Nei, 1987] (as implemented in the R-package `ape` [Paradis et al., 2004]).

In contrast to many statistically consistent network inference methods, such as NANUQ [Allman et al., 2019], SNaQ [Solís-Lemus and Ané, 2016], or PhyNEST [Kong et al., 2024], TINNiK does not assume the input data comes from a level-1 network. Instead, the network can be of arbitrary complexity.

The MSCquartets 3.3 implementation of TINNiK took only 19.4 seconds on the *Leopardus* dataset. Subsequent runs on the same dataset take a couple of seconds, since the most expensive step of tallying quartet counts need not be repeated. TINNiK's runtimes scale to allow larger sets of taxa, requiring only 930 seconds for 50 taxa. For a detailed overview of TINNiK's implementation and usage, a vignette is available within the package¹.

ECToBlob

Introduced by Rhodes et al. [2026], ECToBlob is another statistically consistent algorithm for inferring the tree of blobs of a network under the NMSC. Input to this method is a collection of gene trees summarized by their qcCFs and a tree resolving the true tree of blobs.

¹ <https://cran.r-project.org/web/packages/MSQuartets/vignettes/TINNIK.html>.

Such a tree can be inferred via ASTRAL [Mirarab et al., 2014], or TREE-QMC [Han and Molloy, 2023]. ECToBlob contracts edges of the input tree for which, first, there is poor support due to polytomous quartet signal, and then successively, the most extreme non-tree-like signals, correcting for multiple testing.

ECToBlob has stronger statistical grounding than TINNiK, due to its multiple testing correction, and is expected to perform better across a wide range of datasets. Similar to TINNiK, ECToBlob does not assume any specific network class; it can be applied to networks of arbitrary complexity.

The MSCquartets 3.3 implementation of ECToBlob took 21.9 seconds to analyze the *Leopardus* dataset. Similar to other functions described here, subsequent runs on the same dataset take only a few seconds, using qcCFs and hypothesis test results already computed. ECToBlob's implementation and usage is described more thoroughly in a vignette available within the package².

NANUQ⁺

NANUQ⁺ is a collection of routines for statistically consistent inference networks under the NMSC, developed in Allman et al. [2025]. Given a collection of gene trees as input, the method returns a level-1 network (without 2-cycles or 3-cycles). The inferred network contains both cycles and cut edges, corresponding to reticulate and tree-like features, respectively.

Given a tree of blobs (e.g., one inferred by TINNiK or ECToBlob), multifurcations in this tree are resolved into optimal cycles. This procedure uses a least-squares approach, comparing empirical NANUQ distances between taxon groups around a multifurcation (computed from qcCFs) with the expected distances under candidate cycle configurations.

The user may combine cycle resolutions for some or all individual multifurcations, as long as the resulting networks can be rooted. Incompatible cycle resolutions suggest further investigation is needed, and may cast doubt on the fit of the level-1 NMSC model.

The running time of NANUQ⁺ is typically measured in minutes, even for datasets that would take days to analyze using quartet pseudolikelihood methods and that cannot feasibly be analyzed with current Bayesian approaches. The NANUQ⁺ pipeline took 25.4 seconds for the *Leopardus* dataset, including the estimation of the tree of blob using TINNiK.

While the NANUQ function outputs a splits graph, NANUQ⁺ outputs a level-1 network directly. A limitation of NANUQ⁺ arises when resolving a high-degree multifurcation in the tree of blobs. For a multifurcation of degree m , this process involves considering $\frac{m!}{2}$ candidate cycles. For $m \leq 10$, this is computationally efficient (on the order of minutes), but for $m > 10$, a heuristic is employed to speed up the computation (see Allman et al. [2025] for details).

A vignette available within the package³ gives a more detailed overview of the NANUQ⁺ implementation and usage.

MSCquartets 3.3 Shiny App

The MSCquartets 3.3 Shiny App is a web application that offers tools to infer phylogenetic networks and trees from gene tree data, leveraging the full R package. This tool requires no installation of R,

nor any prior knowledge of its use, providing a friendly introduction to explore the package's main inference algorithms.

It contains five of the core functions of MSCquartets 3.3: Hypothesis testing, TINNiK, ECToBlob, NANUQ, and NANUQ⁺. The user can run any of these algorithms on the *Leopardus* dataset or on their own data. It also allows the user to download the plots obtained from their analysis in PDF format.

While this app contains some level of flexibility in the parameter choice of the functions, we recommend directly using the MSCquartets 3.3 R package for more detailed analyses. The App is available through the shinyapps.io service

<https://mscquartets.shinyapps.io/MSQCq-app/>

Conclusion

The R package MSCquartets 3.3, along with its associated Shiny App, provides a suite of fast and statistically consistent algorithms for inference of species trees, trees of blobs (induced by networks of arbitrary complexity), and level-1 networks under the MSC and NMSC models, respectively. Additionally, the package implements hypothesis testing methods to assess the fit of multilocus datasets to the MSC model. All of these features are based on the relationship between tree or network topology and expected quartet concordance factors, providing a common statistical framework for the package.

The TINNiK, ECToBlob, NANUQ, and NANUQ⁺ estimators can be constructed quickly, making them computationally attractive for large-scale analyses. In addition, existing software for optimizing numerical parameters (such as branch lengths and inheritance probabilities) can be used to refine these initial topological estimates and obtain pseudo-likelihood scores efficiently. This strategy substantially reduces computational cost compared to full (pseudo)-likelihood-based network searches. Beyond their direct use for inference, these estimators can also serve as informative starting points for more computationally-intensive searches over network space. In particular, NANUQ and NANUQ⁺ networks can provide useful initial structures for downstream optimization. TINNiK and ECToBlob trees can supply upper bounds on the number of reticulations for methods that require such constraints, such as PhyNEST and SNaQ [Kong et al., 2024, Solís-Lemus and Ané, 2016], and can help divide-and-conquer strategies, such as subsampling taxa to cover a multifurcation in the tree of blobs, for downstream intensive analyses for inferring the reticulate structure of the corresponding blob.

Conflicts of interest

The authors declare that they have no competing interests.

Funding

This work was partially supported by the National Science Foundation [DMS-2051760 to E.S.A. and J.A.R., DMS-2023239 to C.A., DMS-2331660 to H.B.]; and the Australian Research Council Centre of Excellence for Plant Success in Nature and Agriculture [CE200100015 to J.D.M.]. Part of this research was supported by the National Science Foundation grant [DMS-1929284] while the authors were in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the

² <https://cran.r-project.org/web/packages/MSQCquartets/vignettes/ECToBlob.html>.

³ <https://cran.r-project.org/web/packages/MSQCquartets/vignettes/NANUQplus.html>.

“Theory, Methods, and Applications of Quantitative Phylogenomics” program.

References

- E. Allman, H. Baños, J. Mitchell, and J. Rhodes. TINNiK: Inference of the tree of blobs of a species network under the coalescent. *Algorithms for Molecular Biology*, 2024. doi: 10.1101/2024.04.20.590418.
- E. S. Allman, H. Baños, and J. A. Rhodes. NANUQ: a method for inferring species networks from gene trees under the coalescent model. *Algorithms for Molecular Biology*, 14(1), 2019. doi: 10.1186/s13015-019-0159-2.
- E. S. Allman, H. Baños, J. D. Mitchell, and J. A. Rhodes. The tree of blobs of a species network: Identifiability under the coalescent. *Journal of Mathematical Biology*, 86(1):10, 2023. doi: 10.1007/s00285-022-01838-9.
- E. S. Allman, H. Baños, J. A. Rhodes, and K. Wicke. NANUQ⁺: A divide-and-conquer approach to network estimation. *Algorithms for Molecular Biology*, 20(1), 2025. doi: 10.1186/s13015-025-00274-w.
- R. Desper and O. Gascuel. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of Computational Biology*, 9(5):687–705, 2002. doi: 10.1089/106652702761034136. PMID: 12487758.
- V. Dinh and H. Baños. Misspecification strikes: Astral can mislead in the presence of hybridization, even for nonanomalous scenarios. *Molecular Biology and Evolution*, 42(3):msaf049, 03 2025. ISSN 1537-1719. doi: 10.1093/molbev/msaf049.
- D. Eddelbuettel and R. François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011. doi: 10.18637/jss.v040.i08.
- Y. Han and E. K. Molloy. Improving quartet graph construction for scalable and accurate species tree estimation from gene trees. *Genome Research*, 33(7):1042–1052, 2023. doi: 10.1101/gr.277629.122.
- S. Kong, D. L. Swofford, and L. S. Kubatko. Inference of phylogenetic networks from sequence data using composite likelihood. *Systematic Biology*, 74(1):53–69, 10 2024. ISSN 1063-5157. doi: 10.1093/sysbio/syae054.
- A. D. Leaché, R. B. Harris, B. Rannala, and Z. Yang. The influence of gene flow on species tree estimation: a simulation study. *Systematic biology*, 63(1):17–30, 2014.
- J. Lescroart, A. Bonilla-Sánchez, C. Napolitano, D. L. Buitrago-Torres, H. E. Ramírez-Chaves, P. Pulido-Santacruz, W. J. Murphy, H. Svardal, and E. Eizirik. Extensive Phylogenomic Discordance and the Complex Evolutionary History of the Neotropical Cat Genus *Leopardus*. *Molecular Biology and Evolution*, 40(12):msad255, 2023. doi: 10.1093/molbev/msad255.
- C. Long and L. Kubatko. The Effect of Gene Flow on Coalescent-based Species-Tree Inference. *Systematic Biology*, 67(5):770–785, 03 2018. ISSN 1063-5157. doi: 10.1093/sysbio/syy020.
- S. Mirarab, R. Reaz, M. S. Bayzid, T. Zimmermann, M. S. Swenson, and T. Warnow. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, 30(17):i541–i548, 2014. doi: 10.1093/bioinformatics/btu462.
- J. D. Mitchell, E. S. Allman, and J. A. Rhodes. Hypothesis testing near singularities and boundaries. *Electronic Journal of Statistics*, 13(1), 2019. doi: 10.1214/19-ejs1576.
- X.-X. Pang and D.-Y. Zhang. Impact of Ghost Introgression on Coalescent-Based Species Tree Inference and Estimation of Divergence Time. *Systematic Biology*, 72(1):35–49, 07 2022. ISSN 1063-5157. doi: 10.1093/sysbio/syac047.
- E. Paradis, J. Claude, and K. Strimmer. Ape: Analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20(2):289–290, 01 2004. ISSN 1367-4803. doi: 10.1093/bioinformatics/btg412.
- J. A. Rhodes. Topological metrizations of trees, and new quartet methods of tree inference. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6):2107–2118, 2020. doi: 10.1109/tcbb.2019.2917204.
- J. A. Rhodes, H. Baños, J. D. Mitchell, and E. S. Allman. MSCquartets 1.0: quartet methods for species trees and networks under the multispecies coalescent model in R. *Bioinformatics*, 37(12):1766–1768, 10 2020. doi: 10.1093/bioinformatics/btaa868.
- J. A. Rhodes, E. S. Allman, C. Ané, and H. Baños. Statistical inference of the tree of blobs of a phylogenetic network from quartet concordance factors. *bioRxiv*, 2026. doi: 10.64898/2026.05.28.728501.
- N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425, 07 1987. ISSN 0737-4038. doi: 10.1093/oxfordjournals.molbev.a040454.
- C. Solís-Lemus and C. Ané. Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting. *PLoS Genetics*, 12(3):e1005896, 2016. doi: 10.1371/journal.pgen.1005896.
- C. Solís-Lemus, M. Yang, and C. Ané. Inconsistency of species tree methods under gene flow. *Systematic Biology*, 65(5):843–851, 2016.

Data availability

MSCquartets 3.3 is available through the Comprehensive R Archive Network: <https://CRAN.R-project.org/package=MSCquartets>. The MSCquartets 3.3 Shiny App is available through the shinyapps.io service <https://mscquartets.shinyapps.io/MSQCq-app/>

Author contributions statement

All authors contributed to the development of the package’s underlying concepts and manual. J.A.R. is the lead developer and package manager. All authors contributed to the writing, editing, and approval of this manuscript.

Acknowledgments

We thank Jonas Lescroart for input and discussions.