

Inference of species networks, which go beyond tree-like evolution to allow hybridization and other forms of lateral gene transfer, poses many problems. Histories of individual genes are still described by trees which, under the simplest models, are those displayed on the network. However, moving backward in time, ancestral lineages may not meet at a common ancestor until much earlier than their species merge. This can cause gene trees to differ substantially from displayed trees, an effect called *incomplete lineage sorting* (ILS) (See Figure 1 for an example of a gene tree forming within a species tree). Thus gene trees may only indirectly reflect the species network structure.

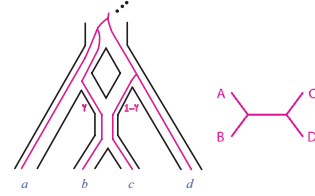


Fig. 1:

Incomplete lineage sorting on a network allows a gene tree, $AB|CD$, to differ substantially from the network structure for the taxa a, b, c, d .

The Network Multispecies Coalescent model (NMSC) describes gene tree formation when both hybridization and ILS may occur [6]. Since it leads to probabilities of gene tree topologies that are polynomials in (transformed) network parameters, both the study of the model and development of inference schemes are amenable to algebraic methods.

One way the NMSC has been used for computational-tractable inference is that after individual gene trees are inferred by standard methods, they are summarized through *quartet concordance factors* (CFs), the frequencies of each of the 3 gene quartet topologies on subsets of 4 taxa. Several schemes treat these as data for tree and level-1 network inference [9, 8, 1]. But to justify the use of CFs for inference, one must understand whether they retain enough information to *identify* the tree or network.

For species trees and level-1 networks identifiability from quartet CFs is well understood [3, 8, 4, 2], with some issues only resolved through study of the varieties they parameterize for 6-taxon networks. Here “level-1” refers to a network in which all cycles caused by hybridization are disjoint. When cycles are allowed to overlap, almost nothing is known about identifiability from CFs, although a few non-identifiable network substructures have been found [7]. Although many classes of networks have been defined [5] (e.g., level- k , tree-child, galled) it is unclear which if any of these might be identifiable from quartet CFs.

Working under the NMSC on a binary semi-directed network, this working group will:

- (1) Understand how the NMSC on an arbitrary network leads to a polynomial form for quartet CFs, and work these out for small networks (e.g., 6-taxon) with specific non-level-1 structures (e.g., level-2 blobs),
- (2) Determine the resulting varieties for these networks, with a goal of understanding their intersections and containments,
- (3) Use these results to characterize a class of networks for which identifiability by quartet CFs holds for generic parameters.

REFERENCES

- [1] E.S. Allman, H. Baños, and J.A. Rhodes. NANUQ: A method for inferring species networks from gene trees under the coalescent model. *Algorithms Mol. Biol.*, 14(24):1–25, 2019.
- [2] E.S. Allman, H. Baños, M. Garrote-Lopez, and J.A. Rhodes. Identifiability of level-1 species networks from gene tree quartets. *arXiv:2401.06290*, 2024.
- [3] E.S. Allman, J.H. Degnan, and J.A. Rhodes. Identifying the rooted species tree from the distribution of unrooted gene trees under the coalescent. *Journal of Mathematical Biology*, 62(6):833–862, 2011.
- [4] H. Baños. Identifying species network features from gene tree quartets. *Bulletin of Mathematical Biology*, 81:494–534, 2019.
- [5] Sungsik Kong, Joan Carles Pons, Laura Kubatko, and Kristina Wicke. Classes of explicit phylogenetic networks and their biological and mathematical significance. 84(6):Paper No. 47, 44, 2022.
- [6] C. Meng and L.S. Kubatko. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical Population Biology*, 75(1):35–45, 2009.
- [7] John A. Rhodes, Hector Banos, Jingcheng Xu, and Cécile Ané. Identifying circular orders for blobs in phylogenetic networks. *arXiv:2402.11693*, 2024.
- [8] Claudia Solís-Lemus and Cécile Ané. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLOS Genetics*, 12(3):1–21, 03 2016.
- [9] C. Zhang, M. Rabiee, E. Sayyari, and S. Mirarab. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(6):153, 2018.