

Developing Novel Algorithms for Evaluating Reticulation Events

Anupam Kumar, Sameer Siddiqi, Ronald M. Clouse, Lavanya Kannan



Abstract

Various phylogenetic network-building software are available today. Despite being able to identify possible cases of horizontal gene transfer, these software lack an objective methodology for finding the most optimal phylogenetic network. This research utilizes nuclear and mitochondrial DNA sequences from the ant genus *Camponotus* were used to construct bifurcating trees in the program POY (*Phylogenetic Analysis Using Dynamic Homology*). These trees were subsequently used to test the behavior of two phylogenetic network-building programs, SplitsTree and Phylonet, in building and predicting reticulation events respectively. This research paves the way for a method of finding the cost of phylogenetic networks, based on the parsimony optimality criterion, which will allow us to find the most optimal phylogenetic hypothesis, whether it be a tree or a network.



Data

DNA samples extracted from Pacific ants from the genus *Camponotus Maculatus* were sequenced for analysis. These specimens were from Papua New Guinea, Federal States of Micronesia, Indonesia, Australia, Palau, and Fiji.

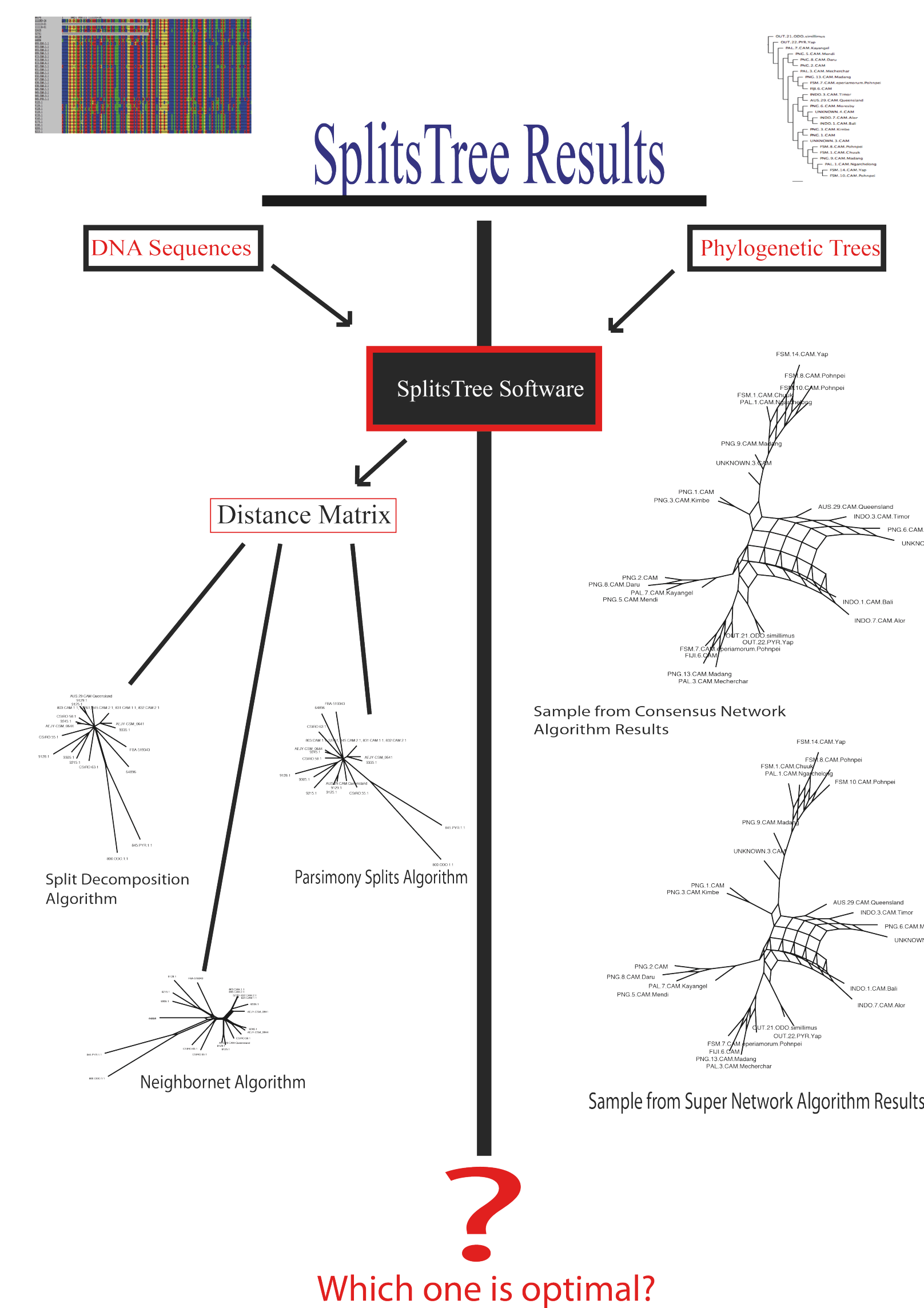
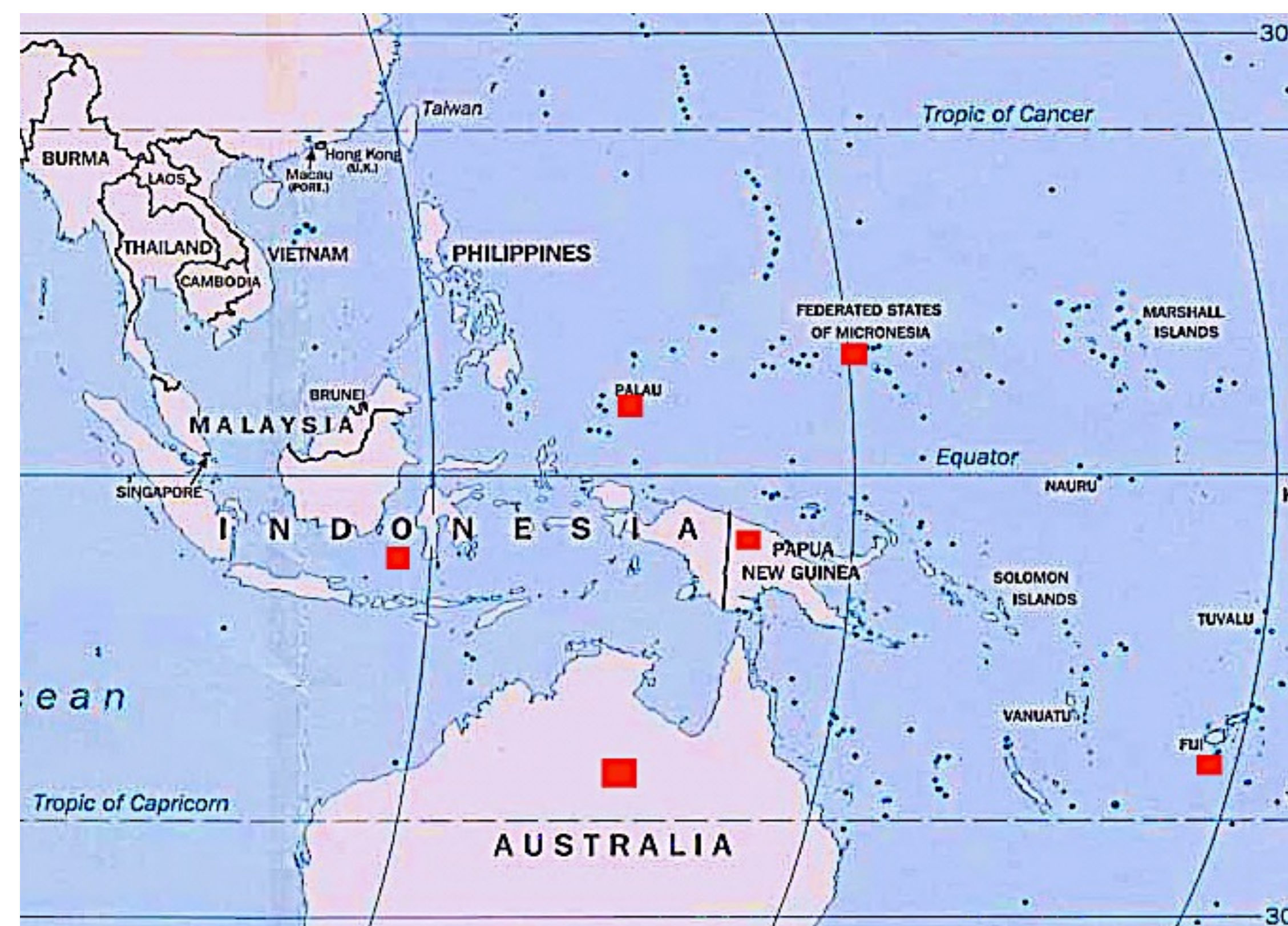


Figure 1: SplitsTree networks, constructed using raw sequence data (left) and pre-existing trees (right).

Phylonet Results in SplitsTree Consensus Network

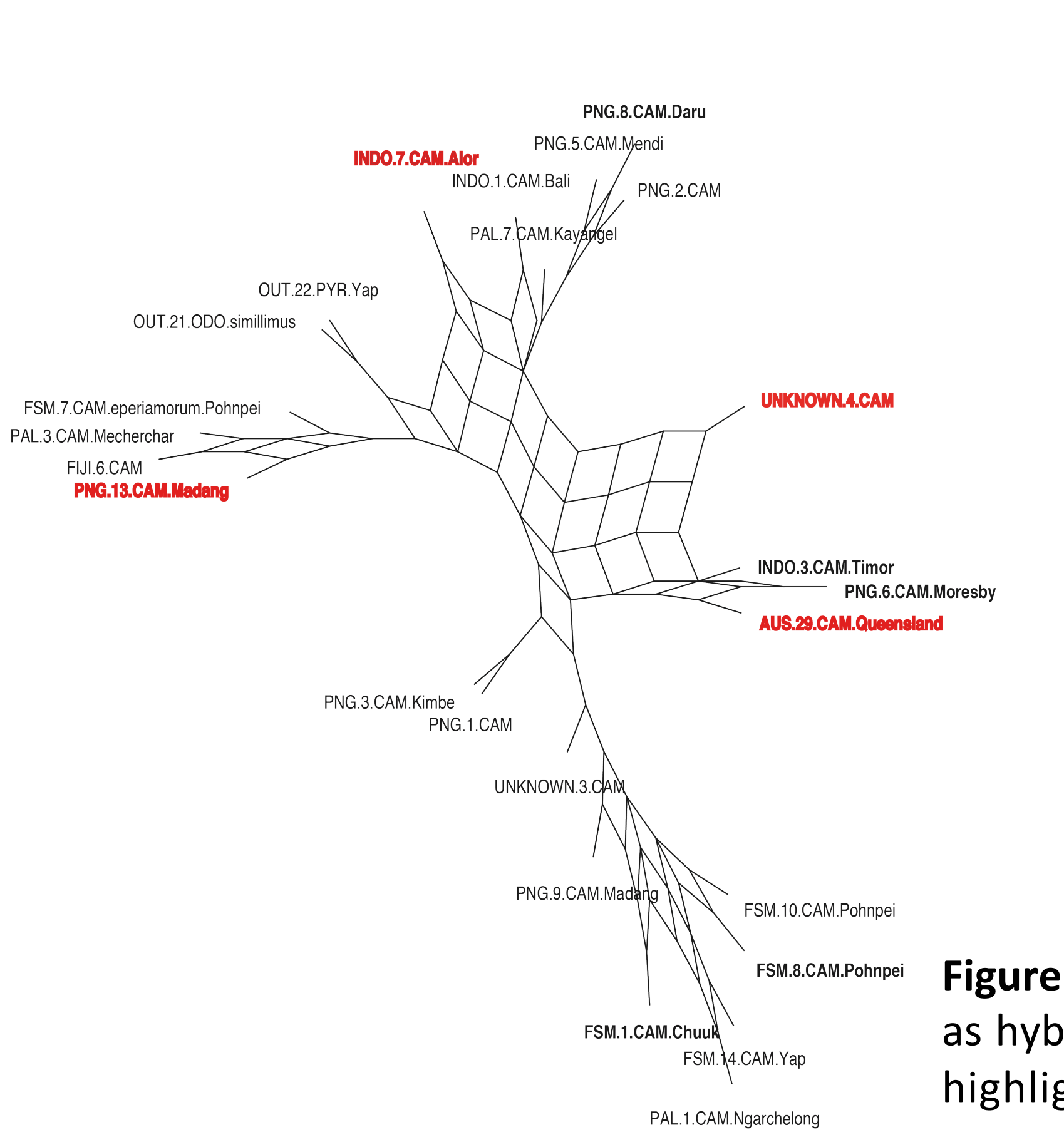


Figure 3: Taxa consistently identified as hybrids from the MAST analysis are highlighted red on the SplitsTree consensus network.

Acknowledgements

Thank you to Hileary Osheroff and Samara Rubenstein for giving us with the opportunity to do research and to our Mentors, Dr. Ronald Clouse and Dr. Lavanya Kannan, for their guidance and support. This project was taken up as part of the Wheeler Lab in the Invertebrate Zoology department of AMNH.

Methods

At various cost matrices, trees constructed from the mitochondrial markers COI and CytB were tested for conflict with counterparts constructed from nuclear markers Ef-1a and LR.

Taxa that did not conflict between the MtDNA and NucDNA trees were found to be redundant, and most were thus deleted. The reduced data set was tested again to confirm conflict. Only data from LR and COI were used.

Networks were built in the program SplitsTree from both sequence data and trees.

In the program Phylonet, conflicting taxa found between mtDNA and nucDNA trees at certain costs were deleted to create a Maximum Agreement Subtree (MAST). The higher the number of deleted species, the greater the probability of hybridization.

A novel algorithm was derived with the purpose of assigning costs to phylogenetic trees and networks from the reduced data set above.

Results

Networks generated from SplitsTree were shown to vary significantly with alterations in inputs (Fig 1).

Through the Phylonet MAST analysis (Fig 2), a total of 11 species were deleted and suspected of hybridization.

4 taxa were consistently recovered as suspect hybrids in various cost matrices (Fig 3).

A preliminary novel algorithm was tested in which trees and networks were scored, and the best reconstruction was recovered and is shown as the "Final Network Output" (Fig 4).

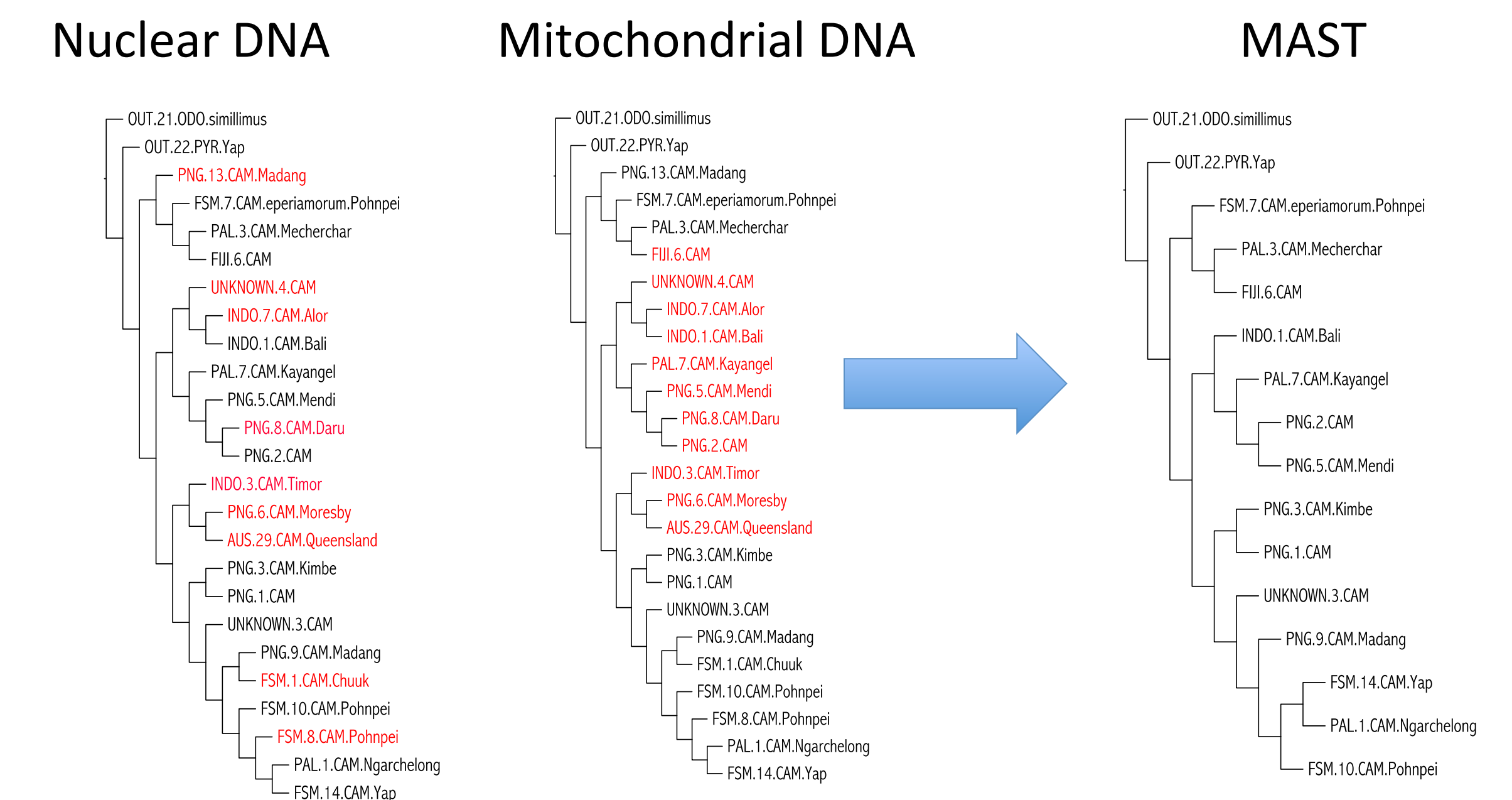


Figure 2: Conflicting taxa in Nuclear and Mitochondrial DNA are highlighted red. MAST tree includes all non-conflicting data.

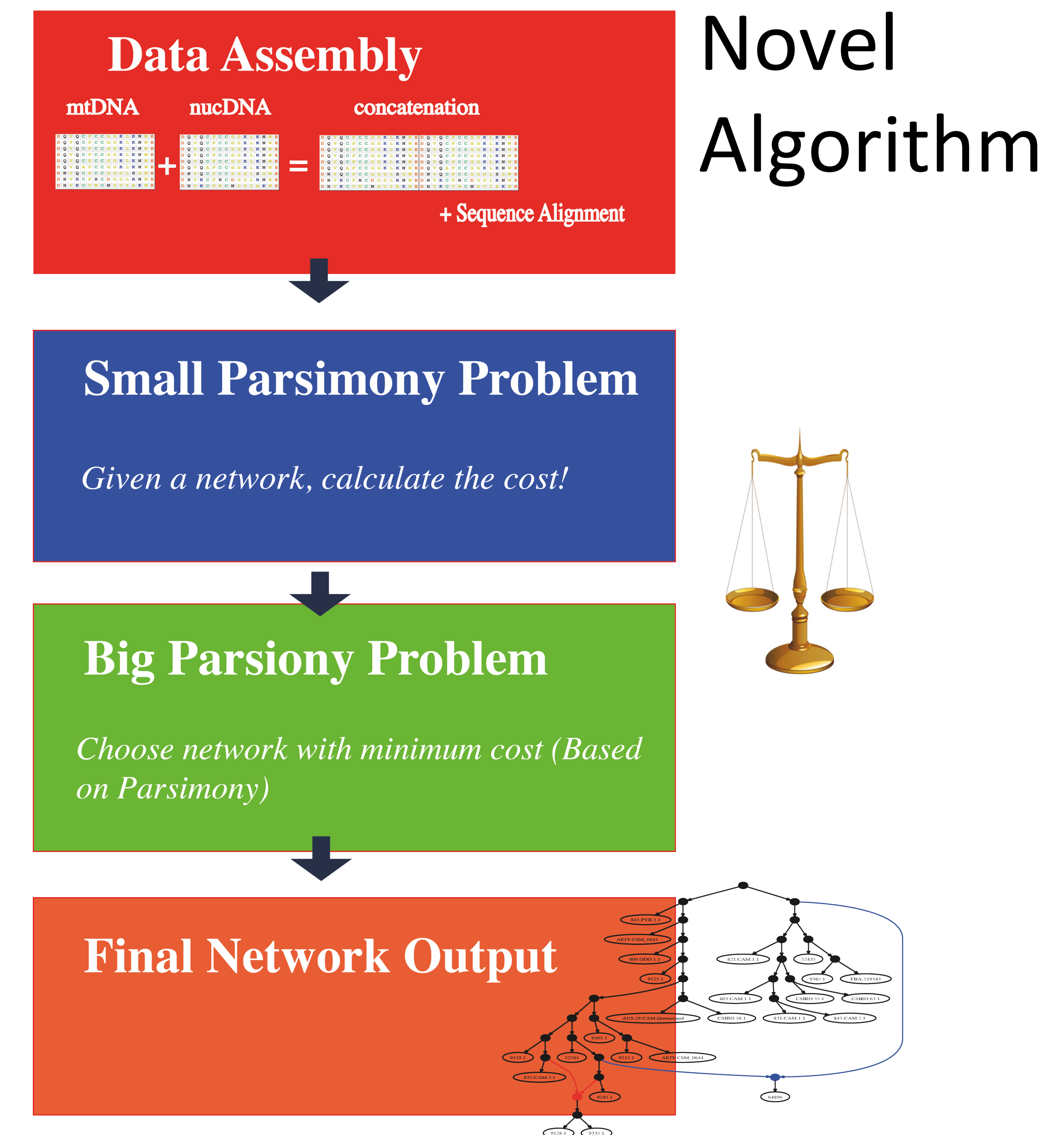


Figure 4: Process for determining a least-cost historical reconstruction, considering both trees and networks simultaneously.