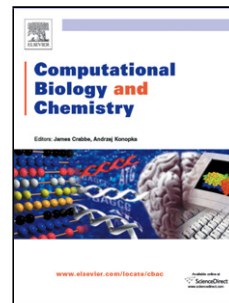# Accepted Manuscript

Title: Live Phylogeny with Polytomies: Finding the Most Compact Parsimonious Trees

Author: D. Papamichail A. Huang E. Kennedy J.-L. Ott A. Miller G. Papamichail

Please cite this article as: D. Papamichail, A. Huang, E. Kennedy, J.-L. Ott, A. Miller, G. Papamichail, Live Phylogeny with Polytomies: Finding the Most Compact Parsimonious Trees, <![CDATA[*Computational Biology and Chemistry*]]> (2017), http://dx.doi.org/10.1016/j.compbiolchem.2017.03.013

Live Phylogeny with Polytomies: Finding the Most Compact Parsimonious Trees

D. Papamichail[a,*], A. Huang[a], E. Kennedy[a], J.-L. Ott[a], A. Miller[a],
G. Papamichail[b]

[a]*Department of Computer Science, The College of New Jersey, Ewing, NJ, 08628*
[b]*Department of Computer Science, New York College, Athens, Greece*

Abstract

Construction of phylogenetic trees has traditionally focused on binary trees where all species appear on leaves, a problem for which numerous efficient solutions have been developed. Certain application domains though, such as viral evolution and transmission, paleontology, linguistics, and phylogenetic stemmatics, often require phylogeny inference that involves placing input species on ancestral tree nodes (live phylogeny), and polytomies. These requirements, despite their prevalence, lead to computationally harder algorithmic solutions and have been sparsely examined in the literature to date. In this article we prove some unique properties of most parsimonious live phylogenetic trees with polytomies, and their mapping to traditional binary phylogenetic trees. We show that our problem reduces to finding the most compact parsimonious tree for n species, and describe a novel efficient algorithm to find such trees without resorting to exhaustive enumeration of all possible tree topologies.

Keywords: Phylogenetics, Maximum Parsimony, Live Phylogeny, Polytomies.

1. Introduction

Phylogeny is the evolutionary history of a set of species whose relationships are often represented by a tree. Phylogenetic trees can be rooted or unrooted, and their edges are labelled with lengths that correspond to evolutionary distances between species.

Maximum Parsimony is a method that uses *characters*, associates a cost with each character mutation (*event*), and aims to build a tree with the smallest possible cost. In recent years, statistical methods [1, 2] have supplanted maximum parsimony approaches for constructing phylogenies in certain domains. However, maximum parsimony remains an effective and widely-used method to predict correct viral phylogenies based on genomic data [3, 4, 5], for morpho-

---

*Corresponding author