

Distributions of topological tree metrics between a species tree and a gene tree

Jing Xi¹ · Jin Xie² · Ruriko Yoshida³

Received: 23 June 2015 / Revised: 27 October 2015 / Published online: 15 March 2016
© The Institute of Statistical Mathematics, Tokyo 2016

Abstract In order to conduct a statistical analysis on a given set of phylogenetic gene trees, we often use a distance measure between two trees. In a statistical distance-based method to analyze discordance between gene trees, it is a key to decide “biologically meaningful” and “statistically well-distributed” distance between trees. Thus, in this paper, we study the distributions of the three tree distance metrics: the edge difference, the path difference, and the precise K interval cospeciation distance, between two trees: First, we focus on distributions of the three tree distances between two random unrooted trees with n leaves ($n \geq 4$); and then we focus on the distributions the three tree distances between a fixed rooted species tree with n leaves and a random gene tree with n leaves generated under the coalescent process with the given species tree. We show some theoretical results as well as simulation study on these distributions.

Keywords Coalescent · Phylogenetics · Tree metrics · Tree topologies

✉ Ruriko Yoshida
ruriko.yoshida@uky.edu

Jing Xi
jxi2@ncsu.edu

Jin Xie
jin.xie@uky.edu

¹ Department of Mathematics, North Carolina State University, 2108 SAS Hall,
2311 Stinson Drive, Raleigh, NC 27695, USA

² Statistics Department, University of Kentucky, Multidiplinary Science Building,
Lexington, KY 40506-0082, USA

³ Statistics Department, University of Kentucky, 325D Multidiplinary Science Building,
Lexington, KY 40506-0082, USA