



HAL
open science

Unique Robustness Properties of Balanced Minimum Evolution

Fabio Pardi

► **To cite this version:**

Fabio Pardi. Unique Robustness Properties of Balanced Minimum Evolution. SMBE: Society for Molecular Biology and Evolution, Jul 2010, Lyon, France. lirmm-01236891

HAL Id: lirmm-01236891

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01236891>

Submitted on 2 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIQUE ROBUSTNESS PROPERTIES OF BALANCED MINIMUM EVOLUTION

FABIO PARDI

LIRMM - Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier.

Background. In a classic paper in computational phylogenetics, Atteson (1999) studied the *robustness* of a number of methods for phylogenetic reconstruction. Informally, this is the ability to withstand noise without compromising the reconstruction of the correct tree.

For methods which reconstruct a tree based on a matrix of distances between taxa, the notion of robustness can be made very precise. The input distances can be seen as estimates of the “correct” evolutionary distances $\mathbf{D}^T = [d_{ij}^T]$ in the unknown evolutionary tree T for the taxa under consideration (where d_{ij}^T is simply the length of the path between i and j in T). In an ideal world, the input distances coincide with those in \mathbf{D}^T , in which case any sensible method is able to reconstruct the correct tree T from \mathbf{D}^T . If, more realistically, the input distances equal $\mathbf{D}^T + \boldsymbol{\epsilon}$, reconstruction of the correct topology of T can only be guaranteed if the noise terms $\boldsymbol{\epsilon}$ are sufficiently small. The robustness of a tree reconstruction method can be measured by the maximum “size” for $\boldsymbol{\epsilon}$ still allowing correct reconstruction of the topology of T .

Atteson (1999) showed that there is a theoretical upper bound Θ_T on $\|\boldsymbol{\epsilon}\|_\infty = \max_{i,j} |\epsilon_{ij}|$ beyond which no method can always reconstruct the correct tree topology (Θ_T is half the length of the shortest branch in T). However, he also proved that a number of algorithms, including neighbor-joining, are guaranteed to reconstruct the topology of T whenever $\|\boldsymbol{\epsilon}\|_\infty < \Theta_T$. Because no method can do better than that, these methods are said to have *optimal robustness*.

Results. Since the most important factor in determining the accuracy of tree reconstruction is the optimization principle used to evaluate alternative trees, we have recently started to investigate the robustness of a number of principles (which can be defined as the robustness of an algorithm reconstructing the optimal tree with respect to that principle).

For example, we have recently shown that *balanced minimum evolution* (BME), the principle underlying neighbor-joining, has optimal robustness, whereas another version of minimum evolution based on least squares has very limited robustness. This difference may partly explain the well-documented gap in reconstruction accuracy between these two approaches.

Here, I will announce a result that considerably strengthens the result above: I show that BME is in fact the *only* principle with optimal robustness, among all *linear* optimization principles, i.e., those that score tree topologies on the basis of linear functions of the input distances (each function usually represents the total branch length associated with a topology) — this includes all minimum evolution principles in the line initiated by Rzhetsky and Nei (1992). In other words, I show that a necessary (as well as sufficient) condition for a linear minimum evolution principle to have optimal robustness is that the coefficients of its tree-length functions coincide with those of BME (and therefore the principle itself coincides with BME).

Finally, I will deal with the practical relevance of this result. In particular the strengths and weaknesses of Atteson’s definition of robustness will be discussed.