
Ecological class prediction: A new method of discriminant analysis, phylogeny-aware and applicable in large dimension.

Anaïs Duhamel^{*1}, Gilles Escarguel², and Julien Clavel³

¹Laboratoire de Géologie de Lyon - Terre, Planètes, Environnement [Lyon] – École Normale Supérieure - Lyon, Université Claude Bernard Lyon 1, Centre National de la Recherche Scientifique – France

²Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés (LEHNA) – CNRS : UMR5023, Université Claude Bernard - Lyon I (UCBL) – 43, boulevard du 11 novembre 1918 Bât Darwin C 69622 Villeurbanne Cedex, France

³University Lyon, UMR CNRS 5023 LEHNA, F-69622, Villeurbanne, France – CNRS : UMR5023 – France

Résumé

Understanding the link between ecological and morphological traits in extant species is a key issue, particularly in paleontology, as it allows the ecology of extinct species to be inferred from morphological features. Such predictions are classically performed using a linear discriminant analysis (LDA): an approach that consists first in fitting a model to a training set of individuals for which both categorical traits (e.g. ecological classes) and continuous traits (e.g. morphology) are known, and then predicting the class of individuals for which only continuous traits are known.

However, existing discriminant analyses either fail to address the fact that individuals are not independent observations from each other due to their shared evolutionary history, or attempt to correct for the phylogenetic signal that is thought to bias the prediction. Instead, we argue that the phylogenetic position of an individual can also be informative for inferring its ecology, as closely related species often share the same ecological classes. In addition, "classical" discriminant analysis methods significantly lose statistical power as the number of morphological traits (p) approaches the number of species (n), and are not applicable to datasets for which p is greater than n (high-dimensional datasets), which are now commonplace with the rise of 2D and 3D geometric morphometrics.

Here we introduce a new discriminant analysis that is both phylogeny-informed and applicable to high-dimensional datasets through penalized likelihood techniques. The performance of this newly implemented method was assessed on simulated and empirical datasets. It appears that this new method outperforms, in many situations, conventional discriminant approaches when applied to comparative datasets (e.g., phylogenetically related species).

Mots-Clés: Ecomorphology, Generalized least squares, High dimensional data sets, Linear discriminant analysis, Penalized likelihood

*Intervenant