Makarenkov, V., E. Bazin & P. Legendre. 2001. T-REX – Software for reconstructing phylogenetic trees and reticulation networks. The multiple regression on distance matrices as a comparative method. Pp. 209-210 in: N. El-Mabrouk, T. Lengauer & D. Sankoff [Eds.], *Currents in computational molecular biology 2001*. (Proceedings of the RECOMB 2001 Conference, Montréal.) Les Publications CRM, Montréal.

T-REX - Software for Reconstructing Phylogenetic Trees and Reticulation Networks. The Multiple Regression on Distance Matrices as a Comparative Method

Vladimir Makarenkov[*†]   *Eric Bazin*[*]   *Pierre Legendre*[*]

1 Description of T-REX

T-REX (tree and reticulogram reconstruction) is an application to reconstruct phylogenetic trees and reticulation networks from distance matrices. The application includes a number of tree fitting methods like NJ, UNJ or ADDTREE which have been very popular in phylogenetic analysis. At the same time, the software comprises several new methods of phylogenetic analysis as for example: tree reconstruction using weights, tree inferring from incomplete distance matrices or modeling a reticulation network between objects or species considered. Reticulation networks can be useful to detect hybridization or mutation events, or represent reticulate evolution process. T-REX also allows the user to a visualize obtained tree or network structures using Hierarchical, Radial or Axial types of tree drawing and manipulate them interactively.

A tree is a formal structure of the representation of the process of evolution. The leaves represent the species under study, the interior nodes represent virtual ancestors and the edges represent the evolutionary events. In biology, this tree is called a *phylogenetic tree*, or an *additive tree* if tree edges have the valuations. The principal goal of phylogenetic reconstruction is to infer an additive tree from imperfect contemporary data, which do not correspond directly to any tree topology. Consequently, one should utilize available fitting methods to obtain the data corresponding to phylogenetic tree. Five fitting algorithms for inferring a phylogenetic tree from distance matrices are provided by T-REX. Recently a number of methods for reconstruction of phylogenetic trees from partial distance or dissimilarity data have been proposed. T-REX allows the user to carry out four of these recently developed methods.

From the biological point of view a reticulogram (or a reticulation network) represents an evolutionary structure in which the objects may be related in a non-unique way to a common ancestor. A phylogenetic tree cannot represent such a structure. Reticulate patterns have been found in nature in some phylogenetic problems: (1) In bacterial evolution, lateral gene transfer (LGT) produces reticulate evolution; (2) Reticulate evolution also occurs in plants where allopolyploidy may lead to the instantaneous appearance of a new species possessing the chromosome complement of its two parent species; (3) It is also found in within-species micro-evolution in sexually reproducing eukaryotes. Reticulate patterns may also occur in non-phylogenetic problems such as host-parasite relationships involving host transfer and in the field of ecological biogeography. When biological data are analyzed the reticulations linking the species or their ancestors can be interpreted as mutation or hybridization events that have occurred during the evolution process. They can also represent the phenomenon of homoplasy or parallel evolution that might have taken place in the past. Applications of the

[*]Département de sciences biologiques, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec H3C 3J7, Canada. makarenv@magellan.umontreal.ca, p0659728@magellan.umontreal.ca, Pierre.Legendre@umontreal.ca

[†]Institute of Control Sciences, 65 Profsoyuznaya, Moscow 117806, Russia.

reticulogram reconstruction algorithm implemented in T-REX can be found in Makarenkov and Legendre [4,5] or in the Special Section of the Journal of Classification [2] dedicated to the Reticulate Evolution including the contributions from F. J. Lapointe, P. Legendre, J. Rohlf, P. Smouse and P. Sneath. As a practical application T-REX has been used in a number of biological studies. For instances, it was employed to analyze Case's data of the immunological distances different between species of frogs. In a recent study, Makarenkov and Legendre [5] considered two applications of reticulation networks constructed using T-REX application. The first example produced a spatially-constrained reticulogram representing the postglacial dispersal of freshwater fishes in the Quebec Peninsula. The second example depicted the morphological differentiation of muskrats in a river valley in Belgium. Makarenkov and Legendre [5] used T-REX to explore how the new reticulation algorithm could be applied to represent homoplasy in the phylogenetic tree of primates.

2 Implementation and Availability

The Windows 9x/NT version of T-REX was written by Vladimir Makarenkov in Visual C++ 6.0 environment. The Macintosh version of T-REX was implemented in C++ and Pascal by Vladimir Makarenkov and Philippe Casgrain. Our software is available free of charges at the T-REX web site at http://www.fas.umontreal.ca/BIOL/Casgrain/en/labo/t-rex/index.html.

3 Multiple Regression on Distance Matrices (DM)

We also would like to pay a particuliar attention to a new comparative method: the multiple regression on distance matrices, which was first used in a predictive way [3]. Let us consider a set of n species and let X_1 and X_2 be distance matrices among them (ex: phenotypic distance matrices). In our recent study we made use of the Euclidian distances. In particular, we considered a distance matrix (DA) associated with a phylogenetic tree representing the evolution of the species considered. The main idea is to regress X_1 on X_2 and DA to obtain the following model: $X_1 = b_0 + b_1^* X_2 + b_2^* DA$.

The goal of the method is to remove the phylogeny effect from the distance matrices (X_1 and X_2) among species. We have run several simulations to test our method under the hypothesis that characters evoluate along the tree with brownian motion. We have compared this method with Felsenstein's Independent Contrasts [1]. We have shown that the Independent Contrasts are almost perfect under the hypothesis of brownian motion and that they work generally better than our met110d. Nethertheless, we have shown that DM can remove some effect of phylogenetic autocorrelation from phenotypic distance matrices. It is worth noting that the hypothesis of brownian motion is not always realistic, in particular, with quantitative characters. Thus, the obtained results have encouraged us to look for other evolutionary contexts, in which our method is well-adapted.

References

[1] Felsenstein, J. 1985. Phylogenies and comparative method. *Am. Nat.* 126. pp. 1–15.

[2] Legendre, P., Lapointe, F. J., Rohlf, J., Smouse, P., and Sneath, P. 2000. Special Section dedicated to the reticulate evolution. *Journal of Classification*, 17. pp. 153–193.

[3] Legendre, P., Lapointe, F. J., and Casgrain, P. 1994. Modeling brain evolution from behavior: a permutational regression approach. *Evolution*, 48. pp. 1487–1499.

[4] Makarenkov, V. and Legendre, P. 1999. General Network Representation of a Dissimilarity Matrix: Adding Reticulations to an Additive tree. Submitted to *Molecular Biology and Evolution*.

[5] Makarenkov, V. and Legendre, P. 2000. Improving the additive tree representation of a dissimilarity matrix using reticulations. In Kiers H. A. L., Rasson J.-P., Groenen P. J. F., and Schader M. (Eds.), *Data Analysis Classification and Related Methods*, Springer. pp. 35–40.