Scaling properties in protein evolution

E. Alejandro Herrada[a], Claudio J. Tessone[b], Víctor M. Eguíluz[a], Emilio Hernández-García[a] and Carlos M. Duarte[c]

[a] *IFISC, Instituto de Física Interdisciplinar y Sistema Complejos CSIC-Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain*

[b] *ETH Zürich, KPL F 31.1, Kreuzplatz 5, CH-8032 Zürich, Switzerland*

[c] *IMEDEA, Instituto Mediterráneo de Estudios Avanzados (CSIC-UIB) C/ Miquel Marqués 21, E-07190 Esporles, Spain*

Many biological processes, from cellular metabolism to population dynamics, are characterized by allometric scaling (power-law) relationships between size and rate. During the last years different research groups[1,2] have developed statistical analysis tools, based on allometric scaling concepts, for the study of tree-like networks (for example to characterize how branching properties change with network size), mostly in the context of transportation networks. Among the biological processes naturally described in terms of a tree-like topology, the diversification of proteins leading to protein families, happening during the evolution of organisms, is conveniently represented in terms of phylogenetic trees, representing the evolutionary relationships among the different proteins. Here we introduce allometric scaling approaches for the statistical analysis of protein families phylogenies. Most of our analysis rely on the PANDIT database[3,4].

A protein phylogeny is considered as a group of tips and nodes linked by branches. Each node ultimately represents a diversification (mutation) event. For each node i, a subtree S_i is made up of a root at node i and all the descendant nodes bellow i. That point of view allows us to understand how much the protein (sub)family members diversify from i, through the subtree size A_i, and how is this diversity arranged, through several topological measures characterizing the shape of the subtree S_i. Universal scaling relationships between shape and size of protein phylogenies are revealed and discussed.

---

[1] J. R. Banavar, A. Maritan and A. Rinaldo. Nature **399**, 130 (1999).

[2] D. Garlaschelli, G. Caldarelli and L. Pietronero. Nature **423**, 165 (2003).

[3] S. Whelan, P. I. W. de Bakker, E. Quevillon, N. Rodriguez and N. Goldman, Nucleic Acids Research **34**, D327 (2006).

[4] `http://www.ebi.ac.uk/goldman-srv/pandit`