Heterogeneous Maximum Likelihood Methods for the Detection of Adaptive Evolution.

Jennifer M. Commins[1], **Dr Peter Foster**[2], **Dr James O. McInerney**[3]

Keywords: methods, heterogeneous maximum likelihood, adaptive evolution.

1 Introduction.

Proteins evolve by means of random genetic drift (Kimura 1983) and adaptive evolution. Adaptive events result in an organism being more fit than previous members of the same group (Hughes 1999). Our aim is to identify where these events have occurred using a Maximum Likelihood approach. Maximum Likelihood (ML) is a complex statistical method (Edwards 1972). Current implementations of this type of analysis appear to contain certain weaknesses and limitations (Suzuki and Nei 2001). As such, ML and Bayesian methods have come under increasing criticism regarding their reliability; empirical evidence suggests that they may not be using realistic models of sequence evolution (**Yu** *et al*. 2003).

2 Aims and Objectives.

The aim of this work is to design new methods for robustly inferring the evolutionary history of extant sequences and for precisely identifying signatures of adaptive evolution events by using rigorous computational optimisation procedures. Our approach requires the minimum of user input in order to find adaptive evolution events. A phylogenetic tree is constructed from a sequence alignment and assumed to be correct. For each internal node of the tree, we evaluate silent (Ds) and replacement (Dn) substitutions between it and adjoining nodes, both ancestral and descendent. Heterogeneity is achieved by dividing the data into categories according to rules describing the substitution process (Foster, 2003). Unlike other approaches, this method performs analysis across entire lineages of a phylogenetic tree as apposed to a single branch (Fig 1). Additionally, we examine whether or not the replacement substitutions were subsequently changed. We perform a similar analysis for silent substitutions. In this way, a path through the tree is maximised, identifying where the number silent and replacement substitutions differ significantly from each other, indicating evolutionary events.

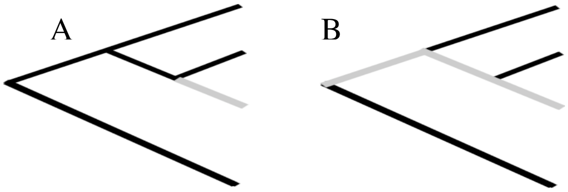

Figure 1: A) Alternative approach B) Lineage analysis implemented in the method described here.

---

[1] National University of Ireland, Maynooth, Co. Kildare, Ireland. ✉ jennifer.commins@may.ie.

[2] Natural History Museum, London, UK. ✉ p.foster@nhm.ac.uk.

[3] National University of Ireland, Maynooth, Co. Kildare, Ireland. ✉ james.o.mcinerney@may.ie

3 Discussion and Further Work.

This approach directly addresses claims that current ML methods are sensitive to violations of the assumptions regarding which model of evolution to use. The software will be tested on both real and simulated data sets, where the likelihood can be verified using the existing Adaptive Evolution Database. The output of this work will be a software product that is capable of performing analyses on multiple sequence alignments, using methods that are closer to biological reality and more user-friendly than existing methods.

4 References and bibliography.

[3] Edwards, A. W. F. 1972. Likelihood. Cambridge University Press, Cambridge.

[6] Foster, P. 2003. Personal Correspondence.

[2] Hughes, A. L. 1999. Adaptive Evolution of Genes and Genomes. Oxford University Press, New York, Oxford.

[1] Kimura, M. 1983. The neutral theory of molecular evolution. Evolution of Genes and Proteins. Cambridge University, New York, London.

[4] Suzuki, Y., and M. Nei. 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. 18:2179-2185.

[5] Yu, Y.-K., J. C. Wootton, and S. F. Altschul. 2003. The compositional adjustment of amino acid substitution matrices. PNAS 100:15688-15693.