1. How old is your fold?

Charlotte M. Deane

University of Oxford, Department of Statistics

We have created the first relative age estimation technique for protein folds. The ages presented show correlation with other protein age estimators and are used to investigate evolutionary pressure on fold topology and complexity. This shows for example very different age patterns of alpha/beta folds compared to small folds.

Introduction:
At present there exists no age estimate for the different protein structures found in nature. It has become clear from occurrence studies that different folds arose at different points in evolutionary time[1]. An estimation of the age of different folds would be a starting point for many investigations into protein structure evolution: how we arrived at the set of folds we see today. It would also be a powerful tool in protein structure classification allowing us to reassess the available hierarchical methods and perhaps suggest improvements.

Methods:
We have created the first relative age estimation technique for protein folds. Our method is based on fold occurrence patterns on completed genomes. The genome assignments were obtained by PSI-BLAST[2] searches and from the SUPERFAMILY[3] database. The superfamily and fold definitions were obtained from SCOP[4]. Firstly we built phylogenetic trees from the completed genomes using neighbour joining and parsimony methods. Ages were then calculated from occurrence patterns on the tree following the method of Mirkin et al.[5].

Using parsimony to estimate relative ages is shown to be more robust against false assignments and horizontal gene transfer, than simply assigning the lowest node to which all occurrences converge. We compare our relative age measure with other previously used protein age estimators. In addition we use it to investigate evolutionary pressure on fold topology, complexity and folding rates.

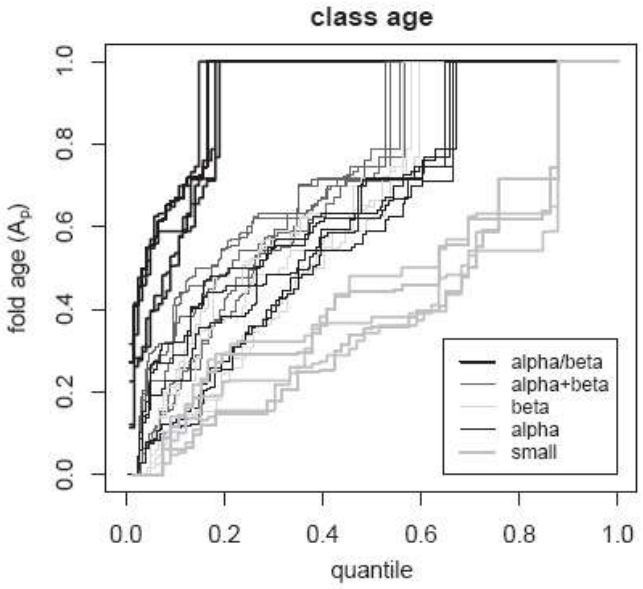

Figure 1: Parsimony age against fold quantile (fraction of folds smaller than the given age) for all trees using SUPERFAMILY assignments. The alpha/beta folds appear on average to be far older than the small folds. The age distribution of each fold class is shown as 6 lines: {superfamily- and fold based trees} × {parsimony, Jaccard and Bray-Curtis distance methods}.

Results & Discussion:

The relative fold ages presented are shown to be robust to various trees, built with different methods and built from either superfamily or fold occurrence patterns. These ages show correlations with other previously used protein age estimators: the maximum number of interactions per fold and the mean number copies of a fold on a genome. In both cases older folds are able to have a higher number of interactions or copies than younger folds, but a low number of copies or interaction does not necessarily indicate a young relative age. The age estimates given are not absolutes but they already offer intriguing insights for different fold topologies, like the very different age patterns of alpha/beta folds compared to small folds. The alpha/beta folds appear on average to be far older than their small fold counterparts.

An age predicting applet, example trees and additional material are available at:
http://www.stats.ox.ac.uk/~abeln/foldage

1. Abeln, S. & Deane, C. M. (2005). Proteins, 60(4):690-700.
2. Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. (1997). Nucleic Acids Res, 25 (17), 3389–3402.
3. Gough, J. & Chothia, C. (2002). Nucleic Acids Res, 30 (1), 268–272.
4. Murzin, A., Brenner, S., Hubbard, T., & Chothia, C. (1995). J Mol Biol, 247 (4), 536–540.
5. Mirkin, B., Fenner, T., Galperin, M., & Koonin, E. (2003). BMC Evol Biol, 3 (1), 2.