Domain tree based analysis of protein architecture evolution

Kristoffer Forslund[1,†], Anna Henricson[2,†], Volker Hollich[2], and Erik L.L. Sonnhammer[1,2,*]

[1] Stockholm Bioinformatics Centre, Albanova, Stockholm University, SE-106 91 Stockholm, Sweden.
[2] Department of Cell and Molecular Biology, Karolinska Institutet, SE-171 77 Stockholm, Sweden.

[*] To whom correspondence should be addressed.
[†] These authors contributed equally to this work.

ABSTRACT

Understanding the dynamics behind domain architecture evolution is of great importance to unravel the functions of proteins. Complex architectures have been created throughout evolution by recombination and duplication events. An interesting question is how many times a particular architecture has been created, a form of convergent evolution or domain architecture reinvention. Previous studies have approached this issue by comparing architectures found in different species. We wanted to achieve a finer-grained analysis by reconstructing protein architectures on complete domain trees.

The prevalence of domain architecture reinvention in fully sequenced genomes was investigated with a novel domain tree based method that uses maximum parsimony for inferring ancestral protein architectures. To ensure robustness, we applied the method to bootstrap trees, and only considered results with strong statistical support. Domain architectures were taken from Pfam for 50 species that cover the kingdoms of life well.

We detected multiple origins for 9.9% of the examined architectures, indicating that this is a much more common phenomenon than previously thought. The results were used to determine which domains are most frequent in multiply created architectures, and to assess whether specific functions could be attributed to them. We found that the protein domains in question are often associated with signaling pathways and regulatory functions, and also a small but statistically significant enrichment of GO terms associated with such functions. This suggests that multiple independent creation of protein architectures is associated with processes involved in rapid adaptive evolution.