



Contribution ID: 54

Type: **Talk**

# Protein structure evolution and convergence

*Monday, 12 February 2024 17:05 (25 minutes)*

The number of known folds is limited to a few thousand and this number is surprisingly low, several orders of magnitude lower than the number of sequences in the biosphere. Biological or physical constraints may considerably limit the repertoire of folds. In this case, structural convergence should be frequent. However, several studies showed that distribution in proteomes may be a global proxy to build phylogeny and recent experiments of protein design tend instead to show that the number of observed folds is very small compared to the number of possible stable folds. To address these apparent contradictions, we have mapped SCOP CATH and ECOD folds onto a sample of 210 species across the tree of life (TOL). We have assessed congruence using retention index of each fold for the TOL, and principal component analysis for deeper branches. Among the folds, 20% are universally present in our TOL, while 54% are clade-specific, especially among the Eukaryotic clades.

Reconstructed ancestral states coupled with dating of each node on the tree of life provided fold appearance rates. The rate is on average twice higher within Eukaryota than within Bacteria or Archaea. The highest rates are found in the origins of eukaryotes, holozoans, metazoans, metazoans stricto sensu, and vertebrates: the roots of these clades correspond to bursts of fold evolution. We could correlate the functions of some of the fold synapomorphies within eukaryotes with significant evolutionary events. Among them, we find evidence for the rise of multicellularity, adaptive immune system, or virus folds which could be linked to an ecological shift made by tetrapods.

### Submitting to:

Integrative Computational Biology workshop

**Primary authors:** Prof. LECOINTRE, Guillaume (MNHN); Dr CHOMILIER, Jacques (CNRS); ROMELI, Martin (CNRS); CARPENTIER, Mathilde (SU - MNH - CNRS)

**Presenter:** CARPENTIER, Mathilde (SU - MNH - CNRS)

**Session Classification:** Computational Methods and Deep Learning