Genome-wide new domain discovery in Sponges

In our analysis, 43 new domains in Sponges are discovered with predicted structures. Data folder contains all data generated for the 43 clusters (domains).

Methods:

1. Potential new domain discovery in sequence level

→ secom seed 6, seqs >= 5

2. Structural prediction of new domains

→ raptorX p-value < 1e-4

3. Functional annotation and validate of new domains

→ blastp (Protein-Protein BLAST 2.2.30+), default , NCBI nr database

→ hmmer, version 3.1b1 e-value 0.01, species selected based on phylogenetic tree.

→ tmalign, tmscore 0.5, length = query

→ spalign, spscore 0.5

→ final PDB annotation.

Results:

Domains discovery with predicted structures

Distribution of domains in other species

Functions of new domains

Discussion

New domains ...