Remote homology detection through structural alignment.

Alberto M. Dávila

Lab de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, FIOCRUZ

Remote homology detection is a challenging problem in Bioinformatics. Arguably, profile Hidden Markov Models (pHMMs) are one of the most successful approaches in addressing this important problem. pHMM packages present a relatively small computational cost, and perform particularly well at recognizing remote homologies. This raises the question of whether structural alignments could impact the performance of pHMMs trained from proteins in the Twilight Zone, as structural alignments are often more accurate than sequence alignments at identifying motifs and functional residues. Next, we assess the impact of using structural alignments in pHMM performance. We observed that pHMMs derived from structural alignments performed significantly better than pHMMs derived from sequence alignment in low-identity regions, mainly below 20%. We believe this is because structural alignment tools are better at focusing on the important patterns that are more often conserved through evolution, resulting in higher quality pHMMs. On the other hand, sensitivity of these tools is still quite low for these low-identity regions. Our results suggest a number of possible directions for improvements in this area.