

Combining methods of remote homology detection can increase coverage by 10% in the midnight zone

Adam Reid

Abstract:

"A recent development in sequence-based remote homologue detection is the introduction of profile-profile comparison methods. They are much more powerful than previous technologies and frequently detect potentially homologous relationships not captured by structural classifications such as CATH and SCOP. As structural classifications traditionally act as the gold standard of homology, this poses a challenge in benchmarking them. We present a novel approach which allows an accurate benchmark of these methods against the CATH structural classification. We then apply this approach to assess the accuracy of a range of publicly available methods of remote homology detection including several profile-profile methods (COMPASS, HHSearch, PRC) from two perspectives. Firstly in distinguishing homologous domains from non-homologues. Secondly in annotating proteomes with structural domain families. PRC is shown to be the best method for distinguishing homologues. We show that SAM is the best practical method for annotating genomes, whilst using COMPASS for the most remote homologues would maximise data quality. Finally we introduce a simple approach to increase the sensitivity of remote homology detection by up to 10%. This is achieved by combining multiple methods with a jury vote."