

LocTree2: a web server for predicting localization in all domains of life

Motivation: The knowledge of the sub-cellular localization of a protein can help in elucidating its function, as a protein's localization can provide hints about its functional role in a cell. Despite advances in high-throughput imaging, localization maps remain importantly incomplete. Several methods have been developed that accurately predict localization, yet many challenges remain to be tackled.

Results: Here, we introduce LocTree2, a web-server for the prediction of localization in all three domains of life, spanning water-soluble globular and membrane proteins. It predicts three localization classes for Archaea, six classes for Bacteria, and eighteen classes for Eukaryota. LocTree2 uses a hierarchical system of Support Vector Machines (SVMs) implemented to imitate the cascading mechanism of cellular sorting. The classification is made using sequence information only. The method reaches high levels of sustained performance for both Eukaryota (Q18=65%; Q18 is eighteen-state accuracy for classifying proteins to eighteen localization classes) and Bacteria (Q6=84%; Q6 is six-state accuracy). Our method also accurately distinguishes between membrane and non-membrane proteins. LocTree2 works well even for protein fragments; these may result from erroneous assemblies or wrong gene predictions that are common in genome projects. In our hands, LocTree2 compared favorably to other state-of-the-art methods when tested on new data. The LocTree2 server includes whole-proteome predictions from a number of bacterial and eukaryotic genomes.

Availability: Online or as standalone version at

<http://www.rostlab.org/services/loctree2> or via the PredictProtein service (predictprotein.org).