摘要：

Most evolutionary analyses or structure modeling are based upon pre-estimated multiple sequence alignment (MSA) models. From a computational point of view, it is too complex to estimate a correct alignment. Hence, increasing or identifying signal inside sequence alignment has intensified over the last few years. During the presentation, I would like to share our couple works on this topic.

The first part, transmembrane proteins (TMPs) constitute about 20~30% of all protein coding genes. We show how homology extension can be adapted and combined with a consistency based approach in order to significantly improve the multiple sequence alignment of alpha-helical TMPs. Then, homology and evolutionary modeling are the most common applications of MSAs. Both are known to be sensitive to the underlying MSA accuracy. We show how this problem can be partly overcome using the transitive consistency score (TCS), an extended version of the T-Coffee scoring scheme. Using this local evaluation function, we show that one can identify the most reliable portions of an MSA, as judged from BAliBASE and PREFAB structure-based reference alignments. Finally, We demonstrate that incorporating MSA induced uncertainty into bootstrap sampling can significantly increase correlation between clade correctness and its corresponding bootstrap value. Our procedure involves concatenating several alternative multiple sequence alignments of the same sequences, produced using different commonly used aligners. We then draw bootstrap replicates while favoring columns of the more unique aligner among the concatenated aligners.