

Comsan Chanma 2007: A Sub-Segment Frequency-Based Multiple Alignment Algorithm for Biological Sequences. Master of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Assistant Professor Panpiti Piamsa-nga, D.Sc. 100 pages.

Multiple sequence alignment has become a necessary task of biologists for sequence analysis. Because of fast increasing of biology data, biologists not only need accurate tools but fast tools are also required for multiple sequence alignment of biology data. In this paper, we propose a sub-segment's frequency-based processing technique to improve speed of alignment process. This technique is pre-process a sub-segment's frequency matrix between all pair of sequences with n-gram technique. This provides us with a distance matrix that can be used to guide the progressive alignment. Accuracy and computing time are measured using BALiBASE and PREFAB benchmark. The CPU time is drastically reduced as compared with ClustalW. On the PREFAB benchmark alignment databases, our result is 6.24 times faster than ClustalW with comparable accuracy.

Comsan Chanma

Student's signature

Panpiti Piamsa-nga

Thesis Advisor's signature

18 / May / 2007