日程: 2011 年 06 月 24 日 (金) 14:30～
場所: 臨海副都心センター別館 8 階コラボレーションコーナー
講演者/発表者: 清水 佳奈 (RNA 情報工学チーム)
主催チーム: RNA 情報工学チーム

タイトル
SlideSort: Fast and exact algorithm for Next Generation Sequencing data analysis

概要
Next Generation Sequencing (NGS) technology calls for fast and accurate algorithms that can evaluate sequence similarity for a huge amount data. In this study, we designed and implemented exact algorithm SlideSort that finds all similar pairs whose edit-distance does not exceed a given threshold from NGS data, which helps many important analyses, such as de novo genome assembly, identification of frequently appearing sequence patterns and accurate clustering.
Using an efficient pattern growth algorithm, SlideSort discovers chains of common k-mers to narrow down the search. Compared to existing methods based on single k-mer, our method is more effective in reducing the number of edit-distance calculations. In comparison to state-of-the-art methods, our method is much faster in finding remote matches, scaling easily to tens of millions of sequences. Our software has an additional function of single link clustering, which is useful in summarizing NGS data for further processing.