中央研究院 資訊科學研究所
Institute of Information Science, Academia Sinica

Distinguished Lecture Series

From Assembling Short DNA Reads to Protein Sequencing by Assembling Mass Spectra

Monday, November 9th, 2009 14:00
Auditorium 106 at new IIS Building

Pavel Pevzner

Director, Center for Algorithmic and Systems Biology, University of California at San Diego

Abstract

Increasing read length is viewed as the crucial condition for fragment assembly with next-generation sequencing technologies. However, introducing mate-paired reads (separated by a gap of length GapLength) opens a possibility to transform short mate-pairs into long mate-reads of length approximately GapLength, and thus raises the question as to whether the read length (as opposed to GapLength) even matters. We describe a new tool for assembling mate-paired short reads and use it to analyze the question of whether the read length matters. We further complement the ongoing experimental efforts to maximize read length by a new computational approach for increasing the effective read length. While the common practice is to trim the error-prone tails of the reads, we present an approach that substitutes trimming with error correction using repeat graphs. An important and counterintuitive implication of this result is that one may extend sequencing reactions that degrade with length "past their prime" to where the error rate grows above what is normally acceptable for fragment assembly.

We further address the problem of sequencing molecules that are not directly inscribed in the genomes (e.g., antibodies or antibiotics-like non-ribosomal peptides) and propose to assemble them from tandem mass spectra. We show that our Eulerian approach to DNA sequencing can be generalized to Shotgun Protein Sequencing (SPS). We illustrate applications of SPS to sequencing of snake venoms (collaborations with Karl Clauser at Broad Institute) and antibodies (collaboration with Jennie Lill at Genentech). We further show how mass-spectrometry enables de novo sequencing of peptide-like natural products.

This is a joint work with Nuno Bandeira (UCSD), Mark Chaisson (Pacific Biosciences) and Dima Brinza (Life Technologies).

For more infomation: http://www.iis.sinica.edu.tw/