



Symposium: Beyond Sequence Alignment | 2017-10-16 | Abstracts

THE CHANGING LANDSCAPE OF BIOLOGICAL SEQUENCE ANALYSIS

Michael Waterman, Professor of Biological Sciences, Computer Science, and Mathematics
University of Southern California

This lecture has three components.

(1) Sequence assembly from shotgun reads began with Staden in conjunction with Fred Sanger at Cambridge. Soon the widely applied overlap-layout-consensus techniques were developed, which in the most sophisticated form resulted in whole genome assembly of the human genome. After 2000, next generation sequencing data made that approach infeasible. A method based in short words (25 letters say) from the reads was derived based on Eulerian graphs, known today as de Bruin assembly.

(2) Classical sequence alignment algorithms had their origins in the 1970s. Alignments of two and more sequences gave valuable insight into evolution and structure. Then local alignment became central and Blast was developed as a heuristic for Smith-Waterman alignment. But massive data sets from NGS have motivated alignment-free methods based on short word counts for the reads and contigs.

(3) What seems remarkable about the success of alignment-free methods is that somehow organisms have some control such content. How does an organism do that? As long as we have had sequences, people have applied Shannon information methods to the sequences, but there has been no success remotely like that in communication theory. What is biological information and how should we measure it? This remains mostly unknown.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity