

# EULERIAN GRAPHS AND READING DNA SEQUENCES

*Michael Waterman*

*University of Southern California and Tsinghua University*

*Los Angeles, CA*

`msw@usc.edu`

## **Abstract**

In 1975 when Sanger was developing the DNA sequencing method for which he was to receive his second Nobel Prize, he found Staden who developed the first computer program to assemble longer DNA sequences from the reads. The reads were randomly located and oriented along the target DNA. Until recently all DNA sequence assembly programs were sophisticated elaborations of this approach. They often consist of three major steps: compare all pairs of reads, find an approximate arrangement of the significant overlaps, and multiple alignment for this arrangement. In 1995 an elegant and entirely new approach was proposed in which each read is broken down into shorter overlapping words, and then a certain graph is constructed so that Eulerian paths in this graph correspond to the target DNA sequence. This lecture will show how this graph is constructed and give some examples of its operation. Today for new-generation sequencing which is producing unprecedented volumes of genetic data, Eulerian graphs are the method of choice.