A New Algorithm for DNA Sequence and Assembly

Written by Idury and Waterman

Introduction
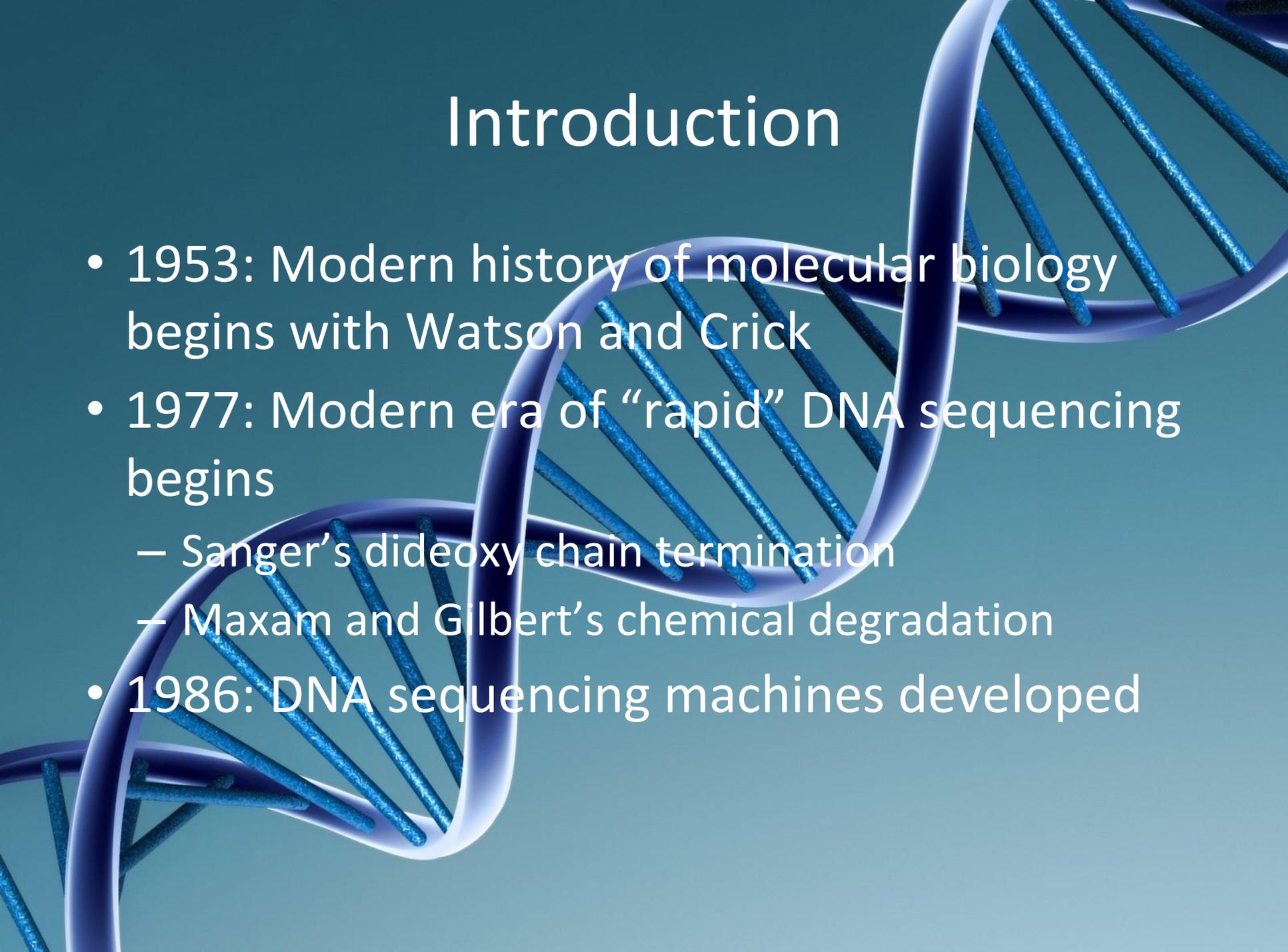
- 1953: Modern history of molecular biology begins with Watson and Crick
- 1977: Modern era of "rapid" DNA sequencing begins
  - Sanger's dideoxy chain termination
  - Maxam and Gilbert's chemical degradation
- 1986: DNA sequencing machines developed

Shotgun Sequencing

- Examines overlapping fragments and determines the encompassing DNA sequence through the overlaps

Genomic DNA

Shotgun

Sequencing

Assembly

ACGATCGATCGATCGTAATTTATAGCATGCTAGCTACTGACGGGCTTTTACGGCGTTAGATATATATCGATCGATCGATGCTATATAGCGTGACTGATCGTAGCTGTAGCTAGCTGTAGCTAGCT
Genomic Sequence
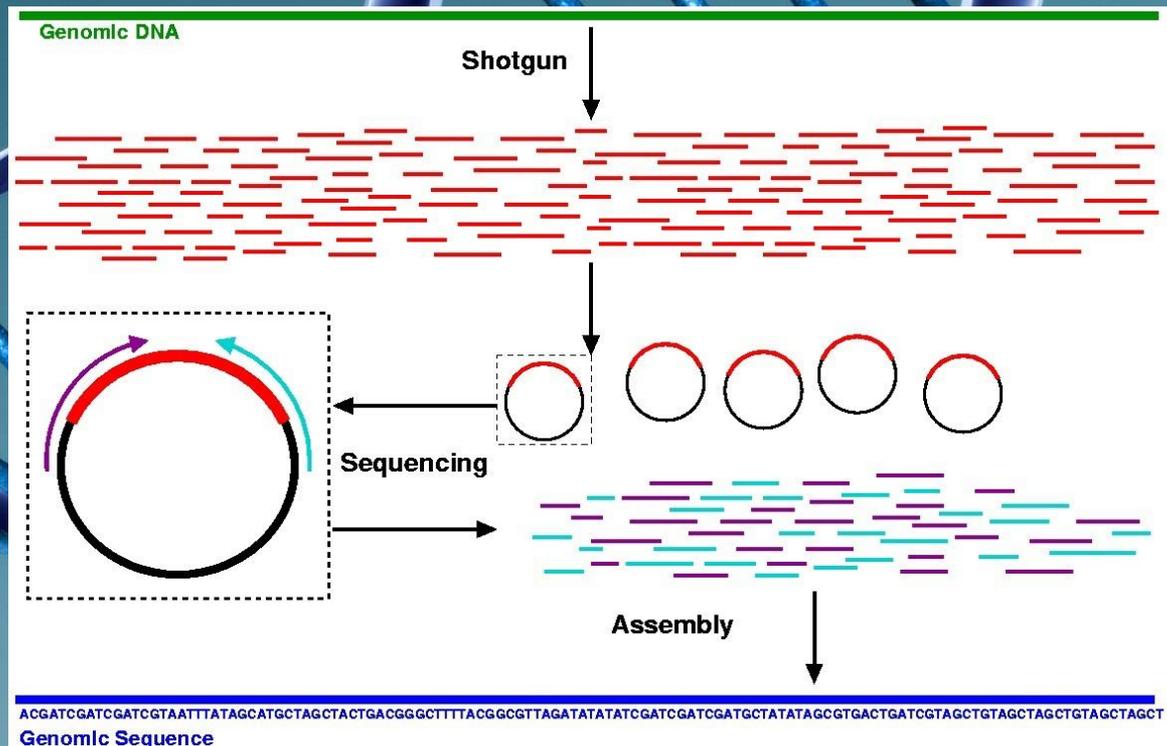
Sequencing by Hybridization
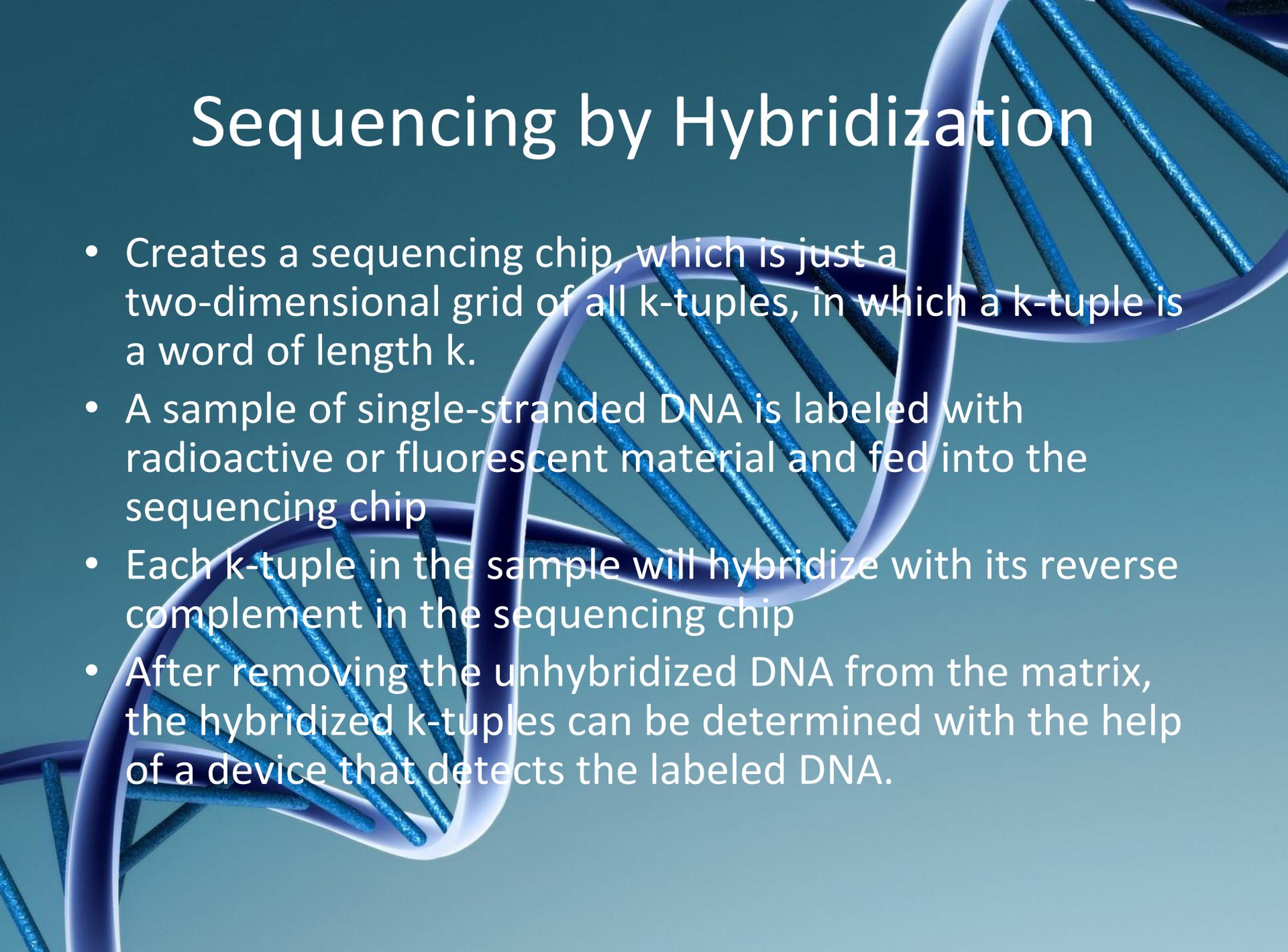
- Creates a sequencing chip, which is just a two-dimensional grid of all k-tuples, in which a k-tuple is a word of length k.
- A sample of single-stranded DNA is labeled with radioactive or fluorescent material and fed into the sequencing chip
- Each k-tuple in the sample will hybridize with its reverse complement in the sequencing chip
- After removing the unhybridized DNA from the matrix, the hybridized k-tuples can be determined with the help of a device that detects the labeled DNA.

Idury and Waterman Algorithm

- From each fragment, all k-tuples are determined
- Next, a directed graph is built based on the k-tuples.
  - The graph has a vertex set V containing (k-1) tuples and edge set containing the k-tuples.
  - A (k-1)-tuple *a* is joined by a directed edge to another (k-1) tuple *b* if there exists a k-tuple whose first (k-1)-tuple is *a* and second (k-1)-tuple is *b*.
- Finally, we perform Eulerian tours on the constructed graph to determine possibly encompassing sequences.
  - An Eulerian path visits every edge in the graph exactly once.
  - If there are multiple Eulerian paths, there are multiple possible sequences.