
DNA sequence analysis

- DNA sequencing is the process of determining the accurate order of nucleotides along chromosomes and genomes.
- It includes any method or technology that is used to determine the order of the four bases—adenine, guanine, cytosine, and thymine—in a strand of DNA.
- Sequencing is used in molecular biology to study genomes and the proteins they encode. Information obtained using sequencing allows researchers to identify changes in genes, associations with diseases and phenotypes.

The Methodology for DNA Sequencing

- Basic methods:

A. The chain termination Sequencing (sanger method).

B. The chemical degradation Sequencing (Maxam and Gilbert method).

- Advanced method:

A. Shotgun sequencing.

B. Bridge PCR.

- Next generation sequencing:

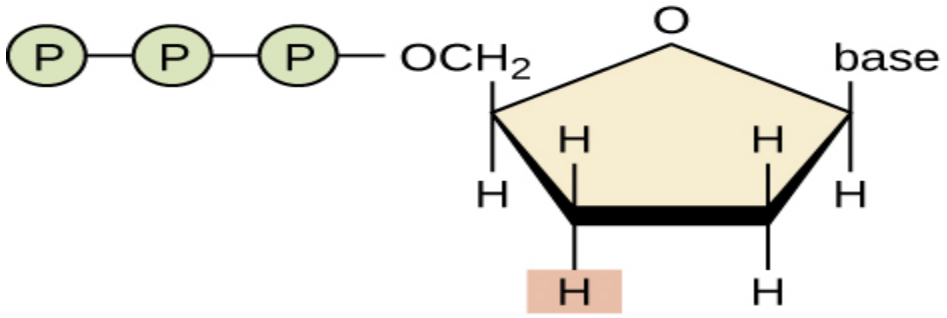
A. Illumina (Solexa) sequencing.

B. SOLiD sequencing.

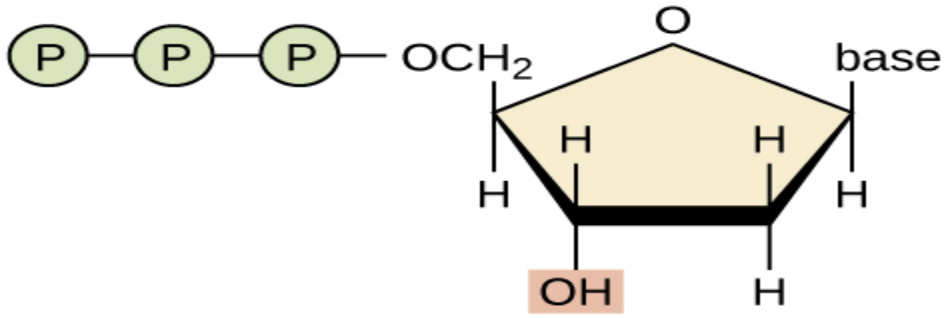
C. Pyrosequencing.

The Sanger Chain-Termination Sequencing Method

- Chain termination DNA sequencing is based on the principle that single-stranded DNA molecules that differ in length by just a single nucleotide can be separated from one another by polyacrylamide gel electrophoresis.
- The original method began with cloning the DNA into a vector, such as M13 phage or a phagemid, that would give the cloned DNA in single stranded form.
- To the single-stranded DNA one hybridizes an oligonucleotide primer about 20 bases long.
- The trick to Sanger's method is to carry out such DNA synthesis reactions in four separate tubes and to include in each tube a different chain terminator. The chain terminator is a dideoxy nucleotide such as dideoxy ATP (ddATP).



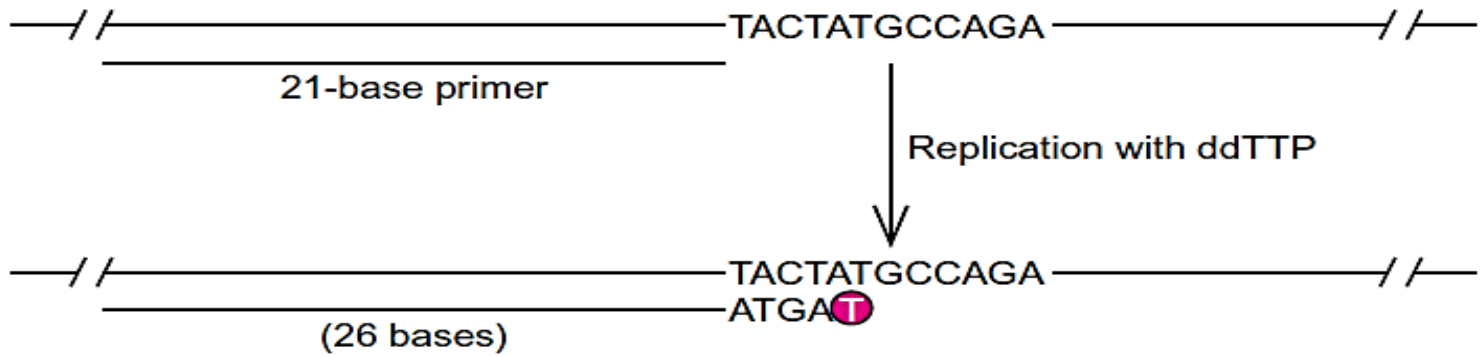
dideoxynucleotide (ddNTP)



deoxynucleotide (dNTP)

- Dideoxy nucleotides by themselves do not permit any DNA synthesis at all, so an excess of normal deoxy nucleotides must be used, with just enough dideoxy nucleotide to stop DNA strand extension once in a while at random.
- This random arrest of DNA growth means that some strands will terminate early, others later. Each tube contains a different dideoxy nucleotide: ddATP in tube 1, so chain termination will occur with A's; ddCTP in tube 2, so chain termination will occur with C's; and so forth.
- The result is a series of fragments of different lengths in each tube. In tube 1, all the fragments end in A; in tube 2, all end in C; in tube 3, all end in G; and in tube 4, all end in T.
- all four reaction mixtures are electrophoresed in parallel lanes in a high-resolution polyacrylamide gel under denaturing conditions, so all DNAs are single stranded. Finally, autoradiography is performed to visualize the DNA fragments, which appear as horizontal bands on an x-ray film.

(a) Primer extension reaction:



(b) Products of the four reactions:

Tube 1: Products of ddA reaction

Template: TACTATGCCAGA
(22) A
(25) ATGA
(27) ATGATA

Tube 3: Products of ddC reaction

Template: TACTATGCCAGA
(28) ATGATAC
(32) ATGATACGGTC

Tube 2: Products of ddG reaction

Template: TACTATGCCAGA
(24) ATG
(29) ATGATACG
(30) ATGATACGG

Tube 4: Products of ddT reaction

Template: TACTATGCCAGA
(23) AT
(26) ATGAT
(31) ATGATACGGT
(33) ATGATACGGTCT

(b) Products of the four reactions:

Tube 1: Products of ddA reaction

Template: TACTATGCCAGA
(22) A
(25) ATGA
(27) ATGATA

Tube 3: Products of ddC reaction

Template: TACTATGCCAGA
(28) ATGATAC
(32) ATGATACGGTC

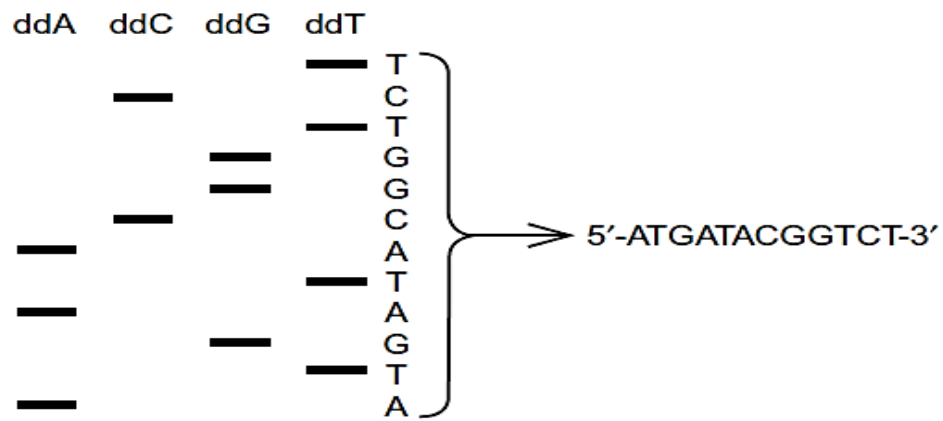
Tube 2: Products of ddG reaction

Template: TACTATGCCAGA
(24) ATG
(29) ATGATACG
(30) ATGATACGG

Tube 4: Products of ddT reaction

Template: TACTATGCCAGA
(23) AT
(26) ATGAT
(31) ATGATACGGT
(33) ATGATACGGTCT

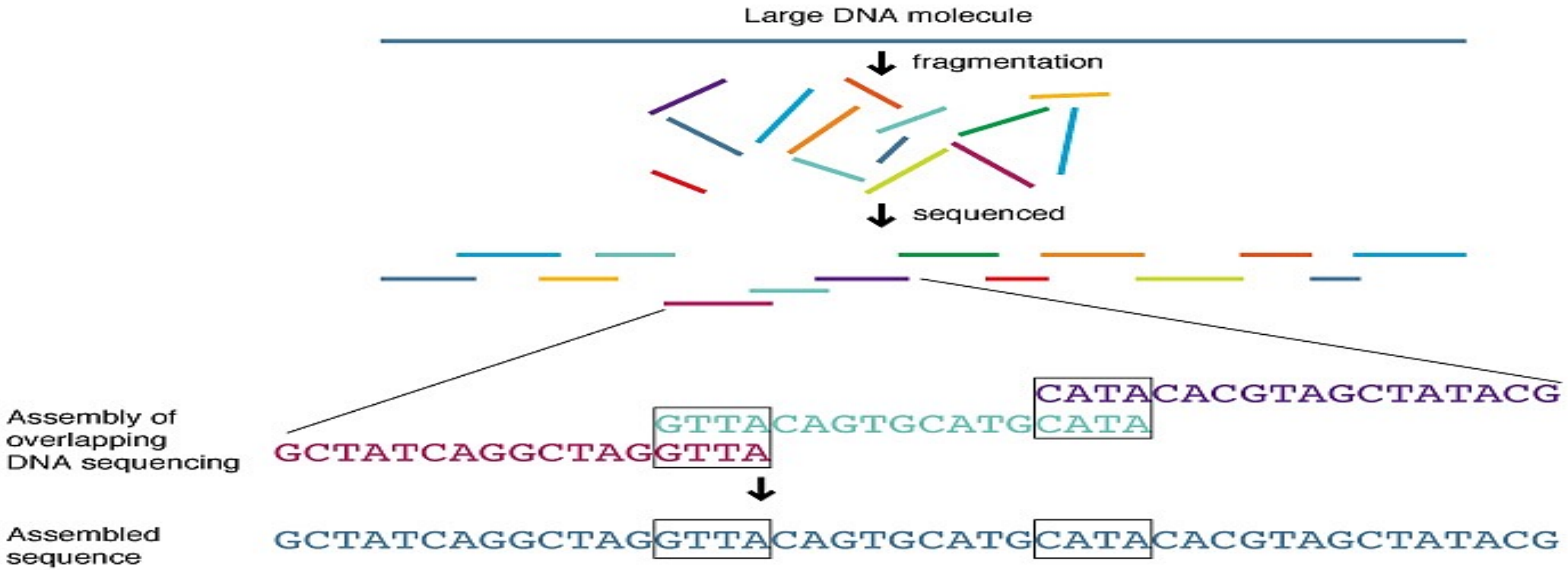
(c) Electrophoresis of the products:



Shotgun sequencing

- Shotgun sequencing is a method used for sequencing long DNA strands.
- The chain termination method of DNA sequencing ("Sanger sequencing") can only be used for short DNA strands of 100 to 1000 base pairs. Due to this size limit, longer sequences are subdivided into smaller fragments that can be sequenced separately, and these sequences are assembled to give the overall sequence.
- In shotgun sequencing, DNA is broken up randomly into numerous small segments, which are sequenced using the chain termination method to obtain reads.
- Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing. Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.
- Whole genome sequencing was obtained by this technique.

- It does not require any prior knowledge of the genome and so can be carried out in the absence of a genetic or physical map.
- In this method, The DNA molecule is broken into small fragments, each of which is sequenced.
- The master sequence is assembled by searching for overlaps between the sequences of individual fragments.

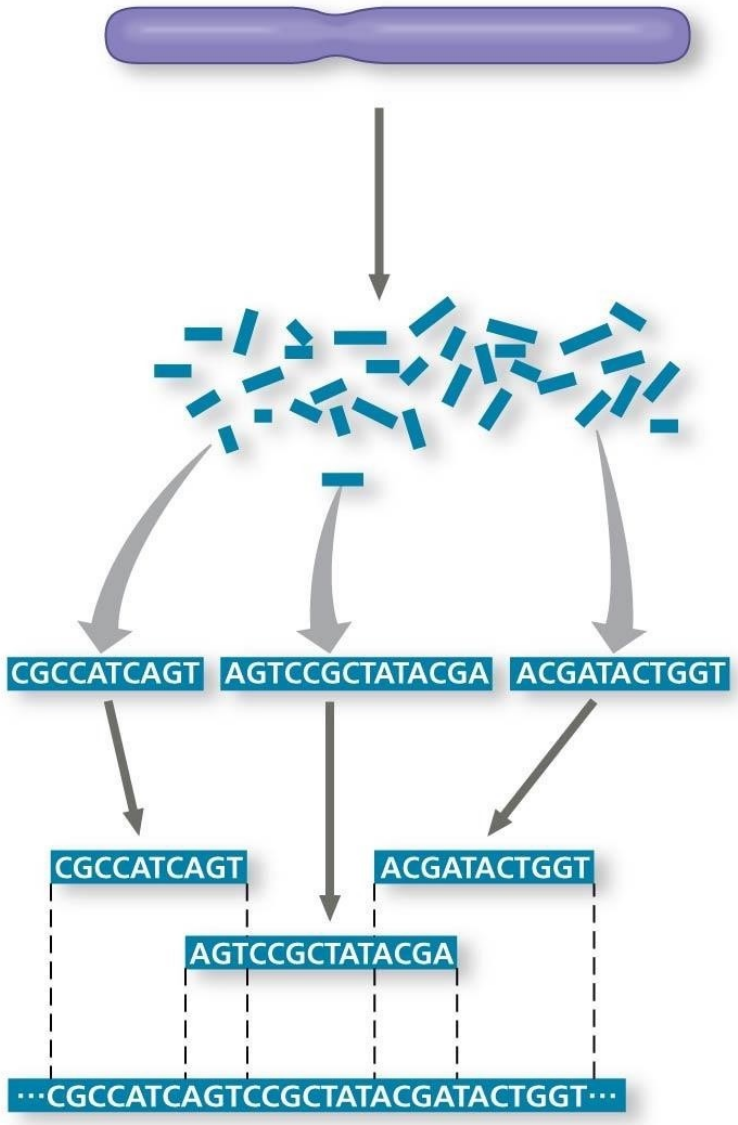


1. Cut the DNA into overlapping fragments short enough for sequencing.

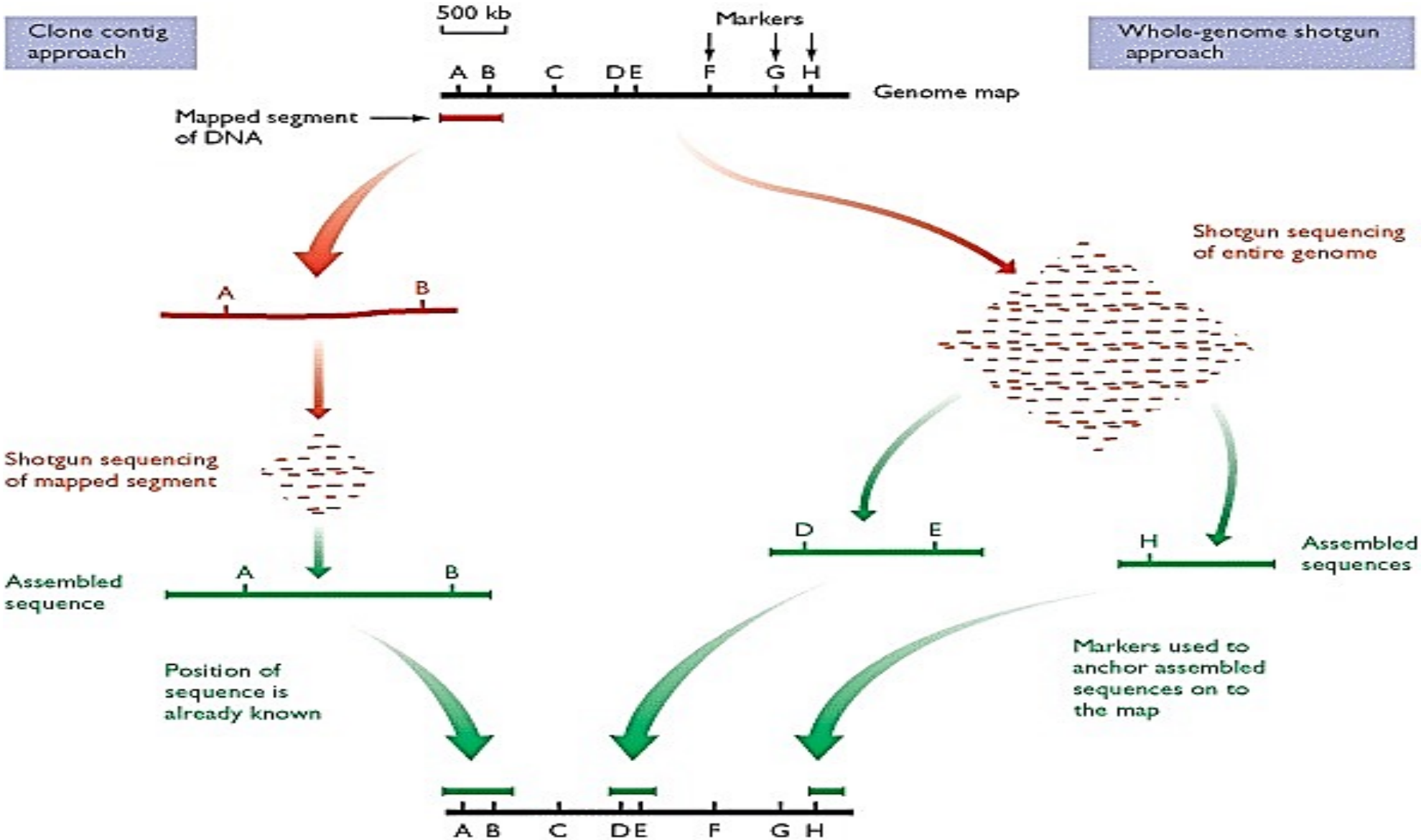
2. Clone the fragments in plasmid or other vectors.

3. Sequence each fragment.

4. Order the sequences into one overall sequence with computer software.



- The difficulties in applying the shotgun method to a large molecule that has a significant repetitive DNA content means that this approach cannot be used on its own to sequence a eukaryotic genome. Instead, a genome map must first be generated. A genome map provides a guide for the sequencing experiments by showing the positions of genes and other distinctive features. Once a genome map is available, the sequencing phase of the project can proceed in either of two ways.


Clone contig approach
500 kb
Markers
A B C D E F G H
Genome map
Whole-genome shotgun approach
Mapped segment of DNA
A
B
Shotgun sequencing of entire genome
Shotgun sequencing of mapped segment
D
E
H
Assembled sequences
A
B
Assembled sequence
Position of sequence is already known
Markers used to anchor assembled sequences on to the map
A B C D E F G H

Next-generation sequencing (pyrosequencing)

- This technique has the great advantages of speed and accuracy, and it does not require electrophoresis.
- The idea behind pyrosequencing is to allow DNA polymerase to replicate the DNA to be sequenced and follow the incorporation of each nucleotide in real time. "sequencing by synthesis" principle. Each nucleotide incorporation event results in the release of pyrophosphate (PPi), and that can be measured quantitatively by coupling it to the generation of light.
- "Sequencing by synthesis" involves taking a single strand of the DNA to be sequenced and then synthesizing its complementary strand enzymatically. The pyrosequencing method is based on detecting the activity of DNA polymerase with another chemoluminescent enzyme (luciferase is an enzyme which emits light in the presence of ATP).

- The single-strand DNA (ssDNA) template is hybridized to a sequencing primer and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and with the substrates adenosine 5´ phosphosulfate (APS) and luciferin.
- The light produced from the addition of each dNTP in a pyrosequencing run is recorded as a peak.
- Nucleotides that are not incorporated generate only a small amount of light. Incorporation of a single nucleotide yields a relative light intensity of 1. Incorporation of two, three, or four nucleotides of the same kind in a row generate relative light intensities of 2, 3, or 4, respectively.