Next-Generation Sequencing (NGS) Basics, Development and Application of High-throughput DNA Sequencing
Table of Contents

- Definition of gene sequencing
- Basic principles of gene sequencing
- Development and improvement of NGS techniques
- Fields of application
- Future of NGS
Definition of gene sequencing

**DNA sequencing** [CELL MOL] The determination of the sequence of nucleotides in deoxyribonucleic acid (DNA) molecules.

**Sequencing:** the act, or process, of determining the sequence of proteins or nucleic acids.

---


Basic principles of gene sequencing

- 1977: publication of two different approaches to gene sequencing

A new method for sequencing DNA
(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

Allan M. Maxam and Walter Gilbert
Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Contributed by Walter Gilbert, December 9, 1976

DNA sequencing with chain-terminating inhibitors
(DNA polymerase/nucleotide sequences/bacteriophage φX174)

F. Sanger, S. Nicklen, and A. R. Coulson
Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977
Chemical cleavage method (Maxam-Gilbert method)

- Principle: base-specific chemical cleavage

<table>
<thead>
<tr>
<th>Chemical agents</th>
<th>Cleaved base(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimethyl sulfate</td>
<td>G</td>
</tr>
<tr>
<td>Formic acid</td>
<td>G+A</td>
</tr>
<tr>
<td>Hydrazine</td>
<td>C+T</td>
</tr>
<tr>
<td>Hydrazine + 1,5 M NaCl</td>
<td>C</td>
</tr>
</tbody>
</table>

- Chemical agents modify certain bases
- Hot piperidine (90 °C) is used to cleave the modified DNA at the position of the modified base
- Electrophoresis and autoradiography
5' $^{32}\text{PACTGTCCGA}$ 3'

Base-specific chemical cleavage of the DNA strand in four different aliquots

Cleavage at G

$^{32}\text{PACTG}$
$^{32}\text{PACTGTCCCG}$

Cleavage at G+A

$^{32}\text{PACTG}$
$^{32}\text{PACTGTCCCG}$

Cleavage at T+C

$^{32}\text{PACTGTC}$
$^{32}\text{PACTGTCC}$

Cleavage at C

$^{32}\text{PACTGTC}$
$^{32}\text{PACTGTCC}$

Electrophoretic resolution/separation and autoradiographic verification
Gel runs in this direction

Read in this direction

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G+A</th>
<th>T+C</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5’ → 3’
Dideoxy (chain-termination) method (Sanger’s method)

- Principle: *in-vitro* strand synthesis and chain termination

2′-Deoxyribonucleoside-5′-Triphosphate (dNTP)  
2',3-dideoxyribonucleoside triphosphate (ddNTP)

- A dNTP possesses a hydroxyl group necessary for chain elongation
- ddNTP has no hydroxyl group
  - ➔ no chain elongation
  - ➔ chain termination
A DNA strand is elongated by the use of a DNA polymerase.

The reaction contains also necessary components für an *in-vitro* DNA synthesis – including one of the four different ddNTPs.

**sequencing reaction components**

<table>
<thead>
<tr>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer¹</td>
</tr>
<tr>
<td>DNA polymerase</td>
</tr>
<tr>
<td>dNTPs und ddNTPs²</td>
</tr>
<tr>
<td>DNA template¹</td>
</tr>
</tbody>
</table>

1 one of them being radioactively labelled
2 only one sort in each of the four aliquots

DNA polymerases can incorporate ddNTPs like dNTPs.

ddNTPs have no 3‘-OH group⇒ chain is terminated after the incorporation of a ddNTP (no formation of a phosphodiester linkage between two nucleotides).

The DNA template is divided into four separate sequencing reactions containing all four standard dNTPs and to each reaction is added only one of the four ddNTPs.
Advantages and disadvantages of the sequencing methods

Maxam-Gilbert method

+ The sequence within an unknown DNA base sequence can be determined with the help of a restriction map
+ The sequence stems from the original DNA molecule (no enzymetically produced copies) → no errors in the base sequence
  - Only short DNA sequences can be sequenced
  - In general, the reaction runs slower and is less reliable
  - The reactions require different harmful chemicals

Sanger’s method

+ Longer fragments can be sequenced
+ Fast and reliable method → easily automatable
+ Radioactive substances can be replaced by fluorescent dyes → faster and safer, as the steps for autoradiography are no longer required
  - A Primer is needed; part of the sequence must be known (solution: use of random primers and primer design (e.g. Edman degradation of proteins and the genetic code))

Improvement in Sanger’s method: radioactive labels, autoradiographic detection and manual data interpretation were replaced with fluorescent labels, laser-induced fluorescence detection and computer-based data analysis
<table>
<thead>
<tr>
<th><strong>Maxam-Gilbert</strong></th>
<th><strong>Chemical fragmentation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sanger</strong></td>
<td>Chain termination method</td>
</tr>
<tr>
<td></td>
<td>(later: capillary (gel) electrophoresis)</td>
</tr>
<tr>
<td><strong>Roche</strong></td>
<td>Shotgun sequencing</td>
</tr>
<tr>
<td><strong>Illumina</strong></td>
<td>Pyrosequencing</td>
</tr>
<tr>
<td><strong>Life Technologies</strong></td>
<td>Illumina sequencing</td>
</tr>
<tr>
<td><strong>Helicos BioSciences</strong></td>
<td>Sequencing by Oligonucleotide Ligation and Detection (SOLiD)</td>
</tr>
<tr>
<td><strong>Pacific Biosciences</strong></td>
<td>Ion semiconductor sequencing</td>
</tr>
<tr>
<td><strong>Oxford Nanopore Technologies</strong></td>
<td>True single molecule sequencing (tSMS)</td>
</tr>
<tr>
<td></td>
<td>Single molecule real time sequencing (SMRT)</td>
</tr>
<tr>
<td></td>
<td>Nanopore sequencing</td>
</tr>
</tbody>
</table>

**First-generation sequencing**

**Second-generation sequencing**

**Third-generation sequencing**

**Next-generation sequencing (NGS)**
First-generation sequencing

Second-generation sequencing

Third-generation sequencing

- *in-vitro* amplification of the template DNA → steps of molecular cloning are reduced
- simplified sample preparation (minimisation, automation and standardisation of working steps → better reproducibility and reduction of error sources), to some extent machines are maintained by only one technician/lab worker
- Improvement of sequencing accuracy
- Increase in number and quantity of samples
- Increase in read length
- Increase in data output
- Decrease of (overall) costs
Development of costs

Cost per Raw Megabase of DNA Sequence

taken from: Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). http://www.genome.gov/sequencingcosts/ (accessed January 6, 2014)
Development of costs

Cost per Genome

taken from: Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). http://www.genome.gov/sequencingcosts/ (accessed January 6, 2014)
Fields of application

Examples:
- Archaeology
- Anthropology
- Genetics and genomics (e.g. evolutionary genomics)
- Biotechnology
- Molecular biology
- Biochemistry
- Forensics

Key aspects:
- Immunogenetics/pharmacogenomics ➔ individualised medicines and therapies (minimise side effects and adverse events; maximise effects)
- Medical genetics/human genetics (detection of mutations, genetically determined (hereditary) diseases)
  - Oncology/oncogenomics ➔ better prediction of the individual course of disease; better therapy selection
- Epigenomics
Histone modification: Histone proteins package and order the DNA.

DNA methylation: addition of a methyl group to DNA nucleotides; influence on gene expression.

Transcription factors: proteins that bind to the DNA and regulate the transcription process.

ChIP-seq: technique for the analysis of Proteine-DNA-interaction.

Outlook and future of NGS

- Ten years after the Human Genome Project: qualified views

  *It is fair to say that all of these predictions have come true, with some caveats that offer important lessons about the best path forward for genomics and personalized medicine. The promise of a revolution in human health remains quite real. Those who somehow expected dramatic results overnight may be disappointed, but should remember that genomics obeys the First Law of Technology: we invariably overestimate the short-term impacts of new technologies and underestimate their longer-term effects.*

- For the future: growing importance of NGS
  - Improvement and development of old and new techniques respectively
  - New projects

References and further literature

- General information on DNA sequencing

• Human Genome Project


• Websites on NGS projects


• Sanger’s method


• Maxam-Gilbert method

Next-Generation Sequencing


- **Capillary electrophoresis**

[1] Karger, B. L.; Guttman, A. DNA Sequencing by Capillary Electrophoresis. *ELECTROPHORESIS* 2009, 30 (S1), S196.


- **Sequencing by Oligonucleotide Ligation and Detection (SOLiD™ technology)**


Ion semiconductor sequencing

VisiGen sequencing system

Pyrosequencing


- **Illumina sequencing**


- **True Single Molecule Sequencing (tSMS™ technology)**


- **Single molecule real time sequencing (SMRT™ technology)**


- **Nanopore sequencing**


