DNA SEQUENCING TECHNOLOGY

Natapol Pornputtapong
17 January 2018
WHY NEED TO KNOW DNA SEQUENCES
WHAT CAN YOU DO WITH A DNA SEQUENCE?

• Geneticists are now able to understand the function of genes by finding distinctive coding regions such as DNA-binding sites, receptor recognition sites and transmembrane domains.

• Scientists have been able to better predict homology among species. Evolutionary biology describes how organisms are related.

• Criminal investigators can use DNA profiling to identify suspects, or exonerate the accused.
DNA SEQUENCE TECHNOLOGY

• First generation: Low throughput methods

• Second generation: High throughput methods

• Third generation: Long read and single molecule methods
FIRST GENERATION

- Sequence by synthesis
  - Sanger sequencing
  - Capillary sequencing
SANGER SEQUENCING:
CHAIN-TERMINATION METHOD

https://binf.snipcademy.com/lessons/dna-sequencing-techniques
# SANGER SEQUENCING: 
## CHAIN-TERMINATION METHOD

<table>
<thead>
<tr>
<th>ddATP</th>
<th>ddCTP</th>
<th>ddGTP</th>
<th>ddTTP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3’\[\text{T}\text{C}\text{G}\text{T}\text{A}\text{C}\text{T}\text{A}]\]

5’
CAPILLARY SEQUENCING

Primer
ACGTACGTACTCAGATGCT
ACGTACGTACTCAGATGC
ACGTACGTACTCAGATG
ACGTACGTACTCAGAT
ACGTACGTACTCAGA
ACGTACGTACTCAG
ACGTACGTACTCA
ACGTACGTACTC
ACGTACGTACT
ACGTACGTAC
ACGTACGTA

Capillary Electrophoresis

Readout: T C G T A G A C T C A
GENOME SIZE

Time:

http://www.pacb.com/blog/data-release-54x-long-read-coverage-for/
LONG TEMPLATE SEQUENCING:
PRIMER WALKING

Sequence of interest

P1

P2

P3

P4

https://binf.snipacademy.com/lessons/dna-sequencing-techniques
SHOTGUN SEQUENCING

Starting fragments:

- ATCAGTA
- AGTATCA
- TAGCTTGCA
- ATCAGTA
- CAGTCAG
- CAGTATAGC

Reconstruct based on overlapping regions with assembler

scaffold

contig 1
- ATCAGTA
- AGTATCA
- CAGTCAG

contig 2
- CAGTATAGC
- TAGCTTGCA

Aligned sequence:

ATCAGTATCAGTCAGTATAGCTTGCA

https://binf.snipcademy.com/lessons/dna-sequencing-techniques
SECOND GENERATION

- Sequence by ligation
  - SOLiD
  - BGI

- Sequence by synthesis
  - Roche: Pyrosequencing
  - Illumina
  - Ion torrent
STEPS IN SECOND GEN

Library preparation
Amplification
Sequencing reaction
Signal detection
Bioinformatic analysis
LIBRARY PREPARATION

Genomic DNA → Fragmentation → Linear DNA molecules → Adaptor ligation → DNA library

RNA → Reverse transcription → cDNA → Fragmentation → Linear DNA molecules → Adaptor ligation → DNA library
AMPLIFICATION:

EMULSION PCR

Emulsion PCR
(454 (Roche), SOLiD (Thermo Fisher), GeneReader (Qiagen), Ion Torrent (Thermo Fisher))

Emulsion
Micelle droplets are loaded with primer, template, dNTPs and polymerase

On-bead amplification
Templates hybridize to bead-bound primers and are amplified; after amplification, the complement strand disassociates, leaving bead-bound ssDNA templates

Final product
100–200 million beads with thousands of bound template

AMPLIFICATION: NANOBALL

Goodwin, 2016. *Nature Reviews Genetics*
AMPLIFICATION:
BRIDGE
AMPLIFICATION

**b Solid-phase bridge amplification (Illumina)**

- **Template binding**
  - Free templates hybridize with slide-bound adapters

**Bridge amplification**
Distal ends of hybridized templates interact with nearby primers where amplification can take place

**Cluster generation**
After several rounds of amplification, 100–200 million clonal clusters are formed

**Patterned flow cell**
Microwells on flow cell direct cluster generation, increasing cluster density

Goodwin, 2016. *Nature Reviews Genetics*
SEQUENCING:
BY LIGATION

SEQUENCING:
BY LIGATION

Goodwin, 2016. *Nature Reviews Genetics*
SEQUENCING:
BY SYNTHESIS

Semiconductor sequencing
As a base is incorporated, a single H⁺ ion is released, which is detected by a CMOS-ISFET sensor

Single nucleotide addition
Only one dNTP species is present during each cycle; several identical dNTPs can be incorporated during a cycle, increasing the emitted ions

SEQUENCING: BY SYNTHESIS

# SECOND GEN MACHINE

<table>
<thead>
<tr>
<th>Technology</th>
<th>Amplification</th>
<th>Sequencing</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOLiD</td>
<td>Emulsion PCR, Template walking</td>
<td>Ligation</td>
<td>Imaging</td>
</tr>
<tr>
<td>BGI</td>
<td>Nanoball</td>
<td>Ligation</td>
<td>Imaging</td>
</tr>
<tr>
<td>IonTorrant</td>
<td>Emulsion PCR</td>
<td>Synthesis: SNA</td>
<td>Voltage measurement</td>
</tr>
<tr>
<td>Illumina</td>
<td>Bridge PCR</td>
<td>Synthesis: CRT</td>
<td>Imaging</td>
</tr>
<tr>
<td>Platform</td>
<td>Read length (bp)</td>
<td>Throughput</td>
<td>Reads</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Sequencing by ligation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOLiD 5500 Wildfire</td>
<td>50 (SE)</td>
<td>80 Gb</td>
<td>~700 M*</td>
</tr>
<tr>
<td></td>
<td>75 (SE)</td>
<td>120 Gb</td>
<td>~700 M*</td>
</tr>
<tr>
<td></td>
<td>50 (SE)*</td>
<td>160 Gb*</td>
<td></td>
</tr>
<tr>
<td>SOLiD 5500 xl</td>
<td>50 (SE)</td>
<td>160 Gb</td>
<td>~1.4 B*</td>
</tr>
<tr>
<td></td>
<td>75 (SE)</td>
<td>240 Gb</td>
<td>~1.4 B*</td>
</tr>
<tr>
<td></td>
<td>50 (SE)*</td>
<td>320 Gb*</td>
<td></td>
</tr>
<tr>
<td>BGISEQ-500 FCS‡</td>
<td>50–100 (SE/PE)*</td>
<td>8–40 Gb*</td>
<td>NA‖</td>
</tr>
<tr>
<td>BGISEQ-500 FCL‡</td>
<td>50–100 (SE/PE)*</td>
<td>40–200 Gb*</td>
<td>NA‖</td>
</tr>
</tbody>
</table>

*Goodwin, 2016. *Nature Reviews Genetics*
### Summary

<table>
<thead>
<tr>
<th>Platform</th>
<th>Read length (bp)</th>
<th>Throughput</th>
<th>Reads</th>
<th>Runtime</th>
<th>Error profile</th>
<th>Instrument cost (US$)</th>
<th>Cost per Gb (US$, approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequencing by synthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ion PGM 318</td>
<td>200 (SE)</td>
<td>600 Mb–1 Gb</td>
<td>4–5.5 M*</td>
<td>4 h</td>
<td>1%, indel‡</td>
<td>$49‡</td>
<td>$450–800‡</td>
</tr>
<tr>
<td></td>
<td>400 (SE)*</td>
<td>1–2 Gb*</td>
<td></td>
<td>7.3 h*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ion Proton</td>
<td>Up to 200 (SE)</td>
<td>Up to 10 Gb*</td>
<td>60–80 M*</td>
<td>2–4 h*</td>
<td>1%, indel‡</td>
<td>$224‡</td>
<td>$80‡</td>
</tr>
<tr>
<td>Ion S5 540</td>
<td>200 (SE)*</td>
<td>10–15 Gb*</td>
<td>60–80 M*</td>
<td>2.5 h*</td>
<td>1%, indel‡</td>
<td>$65 (Ref. 158)</td>
<td>$300*</td>
</tr>
</tbody>
</table>

Goodwin, 2016. *Nature Reviews Genetics*
<table>
<thead>
<tr>
<th>Platform</th>
<th>Read length (bp)</th>
<th>Throughput</th>
<th>Reads</th>
<th>Runtime</th>
<th>Error profile</th>
<th>Instrument cost (US$)</th>
<th>Cost per Gb (US$, approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequencing by synthesis: CRT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illumina MiniSeq Mid output</td>
<td>150 (SE)*</td>
<td>2.1–2.4 Gb*</td>
<td>14–16 M*</td>
<td>17 h*</td>
<td>&lt;1%, substitution‡</td>
<td>$50,000 (Ref. 118)</td>
<td>$200–300 (Ref. 118)</td>
</tr>
<tr>
<td>Illumina MiniSeq High output</td>
<td>75 (SE)</td>
<td>1.6–1.8 Gb</td>
<td>22–25 M (SE)*</td>
<td>7 h</td>
<td>&lt;1%, substitution‡</td>
<td>$50,000 (Ref. 118)</td>
<td>$200–300 (Ref. 118)</td>
</tr>
<tr>
<td></td>
<td>75 (PE)</td>
<td>3.3–3.7 Gb*</td>
<td>44–50 M (PE)*</td>
<td>13 h</td>
<td>&lt;1%, substitution‡</td>
<td>$50,000 (Ref. 118)</td>
<td>$200–300 (Ref. 118)</td>
</tr>
<tr>
<td></td>
<td>150 (PE)*</td>
<td>6.6–7.5 Gb*</td>
<td>24 h*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illumina MiSeq v2</td>
<td>36 (SE)</td>
<td>540–610 Mb</td>
<td>12–15 M (SE)</td>
<td>4 h</td>
<td>0.1%, substitution‡</td>
<td>$99,000‡</td>
<td>~$1,000</td>
</tr>
<tr>
<td></td>
<td>25 (PE)</td>
<td>750–850 Mb</td>
<td>24–30 M (PE)*</td>
<td>5.5 h</td>
<td></td>
<td>$99,000‡</td>
<td>$996</td>
</tr>
<tr>
<td></td>
<td>150 (PE)</td>
<td>4.5–5.1 Gb</td>
<td>24 h*</td>
<td></td>
<td></td>
<td>$996</td>
<td>$212</td>
</tr>
<tr>
<td></td>
<td>250 (PE)*</td>
<td>7.5–8.5 Gb*</td>
<td>39 h*</td>
<td></td>
<td></td>
<td>$996</td>
<td>$142‡</td>
</tr>
<tr>
<td>Illumina MiSeq v3</td>
<td>75 (PE)</td>
<td>3.3–3.8 Gb</td>
<td>44–50 M (PE)*</td>
<td>21–56 h*</td>
<td>0.1%, substitution‡</td>
<td>$99,000‡</td>
<td>$250</td>
</tr>
<tr>
<td></td>
<td>300 (PE)*</td>
<td>13.2–15 Gb*</td>
<td>21–56 h*</td>
<td></td>
<td></td>
<td>$99,000‡</td>
<td>$110‡</td>
</tr>
<tr>
<td>Illumina NextSeq 500/550 High output</td>
<td>75 (SE)</td>
<td>25–30 Gb</td>
<td>400 M (SE)*</td>
<td>11 h</td>
<td>&lt;1%, substitution‡</td>
<td>$250‡</td>
<td>$43</td>
</tr>
<tr>
<td></td>
<td>75 (PE)</td>
<td>50–60 Gb</td>
<td>800 M (PE)*</td>
<td>18 h</td>
<td></td>
<td>$250‡</td>
<td>$41</td>
</tr>
<tr>
<td></td>
<td>150 (PE)*</td>
<td>100–120 Gb*</td>
<td>29 h*</td>
<td></td>
<td></td>
<td>$250‡</td>
<td>$33‡</td>
</tr>
</tbody>
</table>

Goodwin, 2016. *Nature Reviews Genetics*
THIRD GENERATION

• Sequence by synthesis
  • PACBIO

• Sequence by reading
  • Oxford Nanopore
2\textsuperscript{ND} TO 3\textsuperscript{RD} GEN PROTOCOL

Library preparation

Amplification

Sequencing reaction

Signal detection

Bioinformatic analysis

Library preparation

Sequencing reaction

Signal detection

Bioinformatic analysis
PACBIO

Goodwin, 2016. *Nature Reviews Genetics*
OXFORD NANOPORE

<table>
<thead>
<tr>
<th>Platform</th>
<th>Read length (bp)</th>
<th>Throughput</th>
<th>Reads</th>
<th>Runtime</th>
<th>Error profile</th>
<th>Instrument cost (US$)</th>
<th>Cost per Gb (US$, approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific BioSciences RS II</td>
<td>~20 Kb</td>
<td>500 Mb–1 Gb*</td>
<td>~55,000*</td>
<td>4 h*</td>
<td>13% single pass, ≤1% circular consensus read, indel‡</td>
<td>$695‡</td>
<td>$1,000‡</td>
</tr>
<tr>
<td>Pacific Biosciences Sequel</td>
<td>8–12 Kb⁶⁹</td>
<td>3.5–7 Gb*</td>
<td>~350,000*</td>
<td>0.5–6 h*</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxford Nanopore MK 1 MinION</td>
<td>Up to 200 Kb¹⁵⁹</td>
<td>Up to 1.5 Gb¹⁵⁹</td>
<td>&gt;100,000 (Ref. ¹⁵⁹)</td>
<td>Up to 48 h¹⁶⁰</td>
<td>~12%, indel¹⁵⁹</td>
<td>$1,000*</td>
<td>$750*</td>
</tr>
<tr>
<td>Oxford Nanopore PromethION</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

QUESTIONS?