Gel electrophoresis
Gel electrophoresis

Separates
- DNA fragments (single nucleotide resolution)
- proteins
- Protein-DNA complexes can be analyzed by gel electrophoresis

native - denatured gel-electrophoresis
agarose - polyacrylamide gel-electrophoresis
What is needed?

- Agarose - a polysaccharide made from seaweed. Agarose is dissolved in buffer and heated, then cools to a gelatinous solid with a network of crosslinked molecules.
- 1% agarose has ~ 200 nm diameter pores
Polyacrylamide gels have a pore size of only a few nm.
• Buffer - in this case TBE
• The buffer provides ions in solution to ensure electrical conductivity.
• Not only is the agarose dissolved in buffer, but the gel slab is submerged (submarine gel) in buffer after hardening
• Also needed are a power supply and a gel chamber
• Gel chambers come in a variety of models, from commercial through home-made, and a variety of sizes
Agarose gel electrophoresis apparatus
Polyacrylamide gel electrophoresis apparatus

A. Front view

B. Side view
How does it work?

• DNA is an organic acid, and is **negatively** charged
• When the DNA is exposed to an electrical field, the particles migrate toward the **positive** electrode
• Smaller pieces of DNA can travel further in a given time than larger pieces
A gel being run

- Positive electrode
- Agarose block
- DNA loaded in wells in the agarose
- Buffer
- Black background
  To make loading wells easier
- Comb
The results from a DNA gel

- Many samples can be run on one gel- but it is important to keep track
- Most gels have one lane as a ‘DNA ladder’ - DNA fragments of known size are used for comparison
A gel as seen under UV light - some samples show 2 Bands of DNA, while others had none or one. The slower one could be the protein-DNA complex
DNA Sequencing
DNA Sequencing

• DNA sequencing – used to determine the actual DNA sequence of an organism. Using a computer, one can identify an organism, and predict protein sequences and functions based on the nucleic acid data.
  – The most commonly used sequencing method is the dideoxy method.
    • This method uses dideoxynucleotide triphosphates (ddNTPs) which have an H on the 3’ carbon of the ribose sugar instead of the normal OH found in deoxynucleotide triphosphates (dNTPs). Dideoxynucleotides are chain terminators. In a synthesis reaction, if a dideoxynucleotide is added instead of the normal deoxynucleotide, the synthesis stops at that point because the 3’OH necessary for the addition of the next nucleotide is absent.
Deoxy versus dideoxy
DNA synthesis

[Diagram showing the process of DNA synthesis, including the daughter and template strands with base pairing and the 5'-to-3' direction of chain growth.]
In the dideoxy method of sequencing, the template DNA that is to be sequenced is mixed with a primer complementary to the template DNA and the four normal dNTPs, one of which is radioactively labeled for subsequent visualization purposes.

This mixture is then splint into four different tubes that are labeled A, C, G, and T. Each tube is then “spiked” with a different ddNTP (ddATP for tube A, ddCTP for tube C, ddGTT for tube G, or ddTTP for tube T).

DNA polymerase is added and using the DNA template and its’ complementary primer, the synthesis of new strands of DNA complementary to the template begins.

Occasionally a dideoxynucleotide is added instead of the normal deoxynucleotide and synthesis of that strand is terminated at that point.
DNA sequencing continued

• In the tube containing ddATP, some percentage of newly synthesized molecules will get a ddATP in each place that there is a T in the template DNA.

• The result is a set of new DNA molecules in tube A, each of which ends in an A.

• A similar type of reaction occurs in the three other tubes to result in molecules that end in C, G, and T in tubes C, G, and T respectively.

• After the synthesis reactions are complete, the products of the four different tubes are loaded onto four adjacent lane of a polyacrylamide gel and the different fragments are separated by size.

• The sequencing gel is able to resolve fragments that differ in size from each other by only one base.
DNA sequencing continued

- After electrophoresis to separate the fragments by size, the fragments are visualized by exposing the gel to photographic film (Remember that one nucleotide was radioactively labeled).
- All fragments in lane A will end in an A, fragments in lane C will all end in a C, fragments in lane G will all end in a G, and fragments in lane T will all end in a T.
- The sequence of the DNA is read from the gel by starting at the bottom and reading upward.
Dideoxynucleoside Sequencing

Figure 7.1  Human Molecular Genetics, 3/e. (© Garland Science 2004)
Dideoxynucleoside Sequencing

Figure 7-2  part 1 of 2  Human Molecular Genetics, 3/e.  © Garland Science 2004
Dideoxy DNA Sequencing

1. Set up four DNA polymerization reactions that contain the following components.
   - Template strand: 3'-GATGATGG-5'
   - DNA polymerase
   - Primer strand: 5' OH 3'
   - dTTP, dATP, dGTP, dCTP

2. Add one of the four 2', 3'-dideoxyribonucleoside triphosphate chain-terminators to each of the four reaction mixtures.

3. Denature the reaction products, load them on a polyacrylamide gel, separate the products based on size by gel electrophoresis, and expose the gel to X-ray film.
DNA sequencing

Figure 20.28  Photograph of an autoradiograph of a 2',3'-dideoxynucleotide chain-terminator sequencing gel. The sequence defined by the lower portion of the gel is shown on the right.
DNA sequencing

- Automated DNA sequencing – in automated DNA sequencing a radioactive deoxynucleotide is not used and all four dideoxy reactions are done in a single tube.
- This is possible because each ddNTPs is labeled with a different fluorescent dye.
- Therefore the dye present in each synthesized fragment corresponds to the dye attached to the dideoxynucleotide that was added to terminate the synthesis of that particular fragment.
- The contents of the single tube reaction are loaded onto a single lane of a gel and electrophoresis is done.
DNA Sequencing

- A fluorimeter and computer are hooked up to the gel and they detect and record the dye attached to the fragments as they come off the gel.
- The sequence is determined by the order of the dyes coming off the gel.
Automated DNA sequencing
Dideoxynucleoside Sequencing

Figure 7-3 part 1 of 2 Human Molecular Genetics, 3/e. © Garland Science 2004
Genome Sequencing Approaches

Figure 8-3 Human Molecular Genetics, 3/e. (© Garland Science 2004)
Pyrosequencing sheds light on DNA sequencing. Genome Res 2001
Pyrosequencing - Solid Phase

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001
Pyrosequencing - Liquid Phase

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001
Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res 2001
PicoTiter™ Plates

- Multiple optical fibers are fused to form an optical array
- Selective removal of core material leaves wells that serve as ‘test tubes’
- Reactions occurring in the ‘test tubes’ can be monitored optically, through the remaining fiber
- Well diameter: 3μ - 250μ, 44μ typical.
- Plate contains 1.6 million wells
454 LifeSciences Sequencer - Process Overview

1) Prepare Adapter Ligated ssDNA Library

2) Clonal Amplification on 28 μ beads

3) Load beads and enzymes in PicoTiter Plate™

4) Perform Sequencing by synthesis on the 454 Instrument
454 LifeSciences Sequencer - Process Overview

emPCR

A) Anneal Single Stranded template to an excess of DNA Capture beads

B) Emulsify beads and PCR reagents in water-in-oil microreactors

C) Break Microreactors, Enrich for DNA positive beads, Load DNA beads
454 LifeSciences Sequencer

Depositing DNA Beads into the PicoTiter™ Plate

Load beads into PicoTiter™ Plate → Load Enzyme Beads → Centrifuge Step

44 µm
454 LifeSciences Sequencer

454 Technology - Sequencing Instrument
# Read Results: Run Statistics

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome Size</th>
<th>GC Content</th>
<th>Reads</th>
<th>Bases</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. jejuni</em> (NCTC 11168)</td>
<td>1.6MB</td>
<td>31%</td>
<td>226,429</td>
<td>24,521,527</td>
<td>13.1</td>
</tr>
<tr>
<td><em>Escherichia coli</em> (K12)</td>
<td>4.6MB</td>
<td>51%</td>
<td>904,858</td>
<td>86,135,933</td>
<td>18.6</td>
</tr>
<tr>
<td><em>H. salinarum</em> (NRC-1)</td>
<td>2.0MB</td>
<td>68%</td>
<td>213,774</td>
<td>21,309,326</td>
<td>10.4</td>
</tr>
</tbody>
</table>
454 LifeSciences Sequencer

Read Results: Read Length

*C. jejuni* (NCTC 11168)
- 1.6MB Genome
- 31% GC Content
- 1 Run
- 24.5 Million Bases

*Escherichia coli* (K12)
- 4.6MB Genome
- 51% GC Content
- 4 Runs
- 86 Million Bases

*H. salinarum* (NRC-1)
- 2.0MB Genome
- 68% GC Content
- 1 Run
- 21 Million Bases
Solexa Sequencing
Genome Projects

1976/79  First viral genome – MS2/fX174
1980    Mitochondrion
1982    First shotgun sequenced genome – Bacteriophage lambda
1995    First prokaryotic genome – H. influenzae
1996    First unicellular eukaryotic genome – Yeast
1998    The first multicellular eukaryotic genome – C.elegans
2000    Drosophila melanogaster
2000    Arabidopsis thaliana
2001    Human Genome
2002    Mouse Genome

The Genome OnLine Database knows of 958 genome sequencing projects, of which 169 are completed
## Favourite and Model Organisms

### Multicellular Animals

<table>
<thead>
<tr>
<th>Mammals</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>3.5 Gb</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>3.2 Gb</td>
<td></td>
</tr>
<tr>
<td>Cow</td>
<td>3.0 Gb</td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>2.8 Gb</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>3.1 Gb</td>
<td></td>
</tr>
<tr>
<td>Chimp</td>
<td>3.5 Gb</td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>3.0 Gb</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Birds</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken</td>
<td>1.2 Gb</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frog</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Xenopus Laevis</td>
<td>1.7 Gb</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Insects</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Drosophila</td>
<td>165 Mb</td>
<td></td>
</tr>
<tr>
<td>Honey Bee</td>
<td>270 Mb</td>
<td></td>
</tr>
<tr>
<td>Yellow Fever Mosquito</td>
<td>780 Mb</td>
<td></td>
</tr>
<tr>
<td>Malaria Mosquito</td>
<td>278 Mb</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nematodes</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Caenorhabdites elegans</td>
<td>100 Mb</td>
<td></td>
</tr>
<tr>
<td>Caenorhabdites briggsae</td>
<td>80 Mb</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sea Urchin</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongylocentrotus purpuratus</td>
<td>800 Mb</td>
<td></td>
</tr>
</tbody>
</table>

### Multicellular Plants

<table>
<thead>
<tr>
<th>Arabidopsis thaliana</th>
<th>125 Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rice</td>
<td>430 Mb</td>
</tr>
</tbody>
</table>

Strachan and Read (2004) Chapter 8
The Human Genome I


β−globin
Exon 2
Exon 1
Exon 3

(Chromosome 11)

α globin

Myoglobin

DNA:
ATTGCCATGTCGATAATTGGACTATTTGGA

Protein:

aa aa aa aa aa aa aa aa aa aa

mitochondria

3.2*10⁹ bp

*5.000

6*10⁴ bp

*20

3*10³ bp

*10³

30 bp
<table>
<thead>
<tr>
<th>Category</th>
<th>Nuclear Genome</th>
<th>Mitochondria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly conserved - coding</td>
<td>1.5%</td>
<td>93%</td>
</tr>
<tr>
<td>Highly conserved - other</td>
<td>3.5%</td>
<td>5%</td>
</tr>
<tr>
<td>Transposon based repeats</td>
<td>45%</td>
<td>-</td>
</tr>
<tr>
<td>Heterochromatin</td>
<td>6.6%</td>
<td>-</td>
</tr>
<tr>
<td>Other non-conserved</td>
<td>44%</td>
<td>2%</td>
</tr>
<tr>
<td>Mendelian inheritance</td>
<td>1 (typically)</td>
<td>Possibly thousands</td>
</tr>
<tr>
<td>Recombination</td>
<td></td>
<td>No recombination</td>
</tr>
<tr>
<td>Gene Density:</td>
<td>1/130 kb</td>
<td>2 kb</td>
</tr>
<tr>
<td>Pseudogenes:</td>
<td>20000</td>
<td></td>
</tr>
<tr>
<td>Processed Pseudogenes:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Gene families

Clustered

α-globins (7), growth hormone (5), Class I HLA heavy chain (20), ...

Dispersed

Pyruvate dehydrogenase (2), Aldolase (5), PAX (>12), ...

Clustered and Dispersed

HOX (38 – 4), Histones (61 – 2), Olfactory receptors (>900 – 25), ...

Transposons

Classes of interspersed repeat in the human genome

<table>
<thead>
<tr>
<th>Class</th>
<th>Autonomous</th>
<th>Non-autonomous</th>
<th>Length</th>
<th>Copy number</th>
<th>Fraction of genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINEs</td>
<td></td>
<td></td>
<td></td>
<td>6–8 kb</td>
<td>21%</td>
</tr>
<tr>
<td>SINEs</td>
<td></td>
<td></td>
<td>100–300 bp</td>
<td>1,500,000</td>
<td>13%</td>
</tr>
<tr>
<td>Retrovirus-like elements</td>
<td></td>
<td></td>
<td>6–11 kb</td>
<td>450,000</td>
<td>8%</td>
</tr>
<tr>
<td>DNA transposon fossils</td>
<td></td>
<td></td>
<td>2–3 kb</td>
<td>300,000</td>
<td>3%</td>
</tr>
</tbody>
</table>

Genes and Gene Structures I

• Presently estimated Gene Number: 24,000 (reference:  )

• Average Gene Size: 27 kb

• The largest gene: Dystrophin 2.4 Mb - 0.6% coding – 16 hours to transcribe.

• The shortest gene: tRNA\textsuperscript{TYR} 100% coding

• Largest exon: ApoB exon 26 is 7.6 kb Smallest: <10bp

• Average exon number: 9

• Largest exon number: Titin 363 Smallest: 1

• Largest intron: WWOX intron 8 is 800 kb Smallest: 10s of bp

• Largest polypeptide: Titin 38,138 smallest: tens – small hormones.

• Intronless Genes: mitochondrial genes, many RNA genes, Interferons, Histones,..
Genes and Gene Structures II

Genes within Genes:

Intron 26 of neurofibromatosis type I (NF1) contains 3 internal (2 exons) genes in the opposite direction.

Overlapping Genes:

Class III region of HLA

Simple Eukaryotic
1. A challenge to automated annotation.
2. How widespread is it?
3. Is it always functional?
4. How does it evolve?

Cartegni, L. et al. (2002) “Listening to Silence and understanding nonsense: Exonic mutations that affect splicing” Nature Reviews Genetics 3.4:285-
## RNAs in the Genome

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>~200</td>
<td>snoRNA</td>
<td>small nucleolar, over 100 types - RNA modification and processing</td>
</tr>
<tr>
<td>~100</td>
<td>snRNA</td>
<td>small nuclear - involved in splicing</td>
</tr>
<tr>
<td>~200</td>
<td>miRNA</td>
<td>very small ~22bp, regulation</td>
</tr>
<tr>
<td>~175</td>
<td>28S, 5.8S, 5S</td>
<td>large cytosolic subunit</td>
</tr>
<tr>
<td>~175</td>
<td>18S</td>
<td>small mitochondrial subunit</td>
</tr>
<tr>
<td>~250</td>
<td>5S</td>
<td>large mitochondrial subunit</td>
</tr>
<tr>
<td>&gt;500</td>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>&gt;1500</td>
<td>Antisense RNA</td>
<td>&gt; 1500 types</td>
</tr>
<tr>
<td>≈10000</td>
<td>non-coding RNA</td>
<td>regulatory functions?</td>
</tr>
</tbody>
</table>

Applications of Ultra-Low-Cost Sequencing

- Sequencing individual human genomes as component of preventive care.
- Genotype-phenotype associations
- Comprehensive gene expression profiling in vitro and in situ at all stages of development of a multicellular organism
- Comprehensive analysis of mutations present in cancer clones.
- Mitochondrial heteroplasmy
- Microbial diversity (metagenomic studies)

Shendure N. Advanced sequencing technology: methods and goals. Nat Rev Gen 2004
Oligonucleotide Array Synthesis

- Photolithography
- In situ synthesis
- Photolabile protective groups (photomasking)
High-Density Oligonucleotide Arrays

Chee M. Assessing genetic information with high-density oligonucleotide arrays. Science 1996
High-Density Oligonucleotide Arrays

Chee M. Assessing genetic information with high-density oligonucleotide arrays. Science 1996