PCR, gel electrophoresis and Sanger sequencing

Theo H. M. Smits

COST 1104 Training Course, 24.09.2015

«Early» gene technology: a problem and a solution

- Before 1985: tedious labor to get target DNA
  - Selection using Southern blotting
  - Cloning into vector
  - Use of restriction sites to create expression vectors

Kary Mullis
- 1983: first results with a new method
- 1987: patent granted
- 1993: Nobel Price
- Polymerase Chain Reaction
PCR – definition and requirements

- The polymerase chain reaction is a scientific technique to amplify a selected piece of DNA to generate thousands of millions of exact copies.

  - **Prerequisite:** DNA Sequence is known
  - **Equipment:** Thermocycler
  - **Material:** DNA, Polymerase, nucleotides, primers, buffer

The method

![Diagram of PCR process]
The method

- Exponential amplification of DNA pieces
- After 35 cycles: $2^{35}$ copies (68 billion) when starting with 2 copies
- Trace amounts of template can be detected

A standard PCR program

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturing</td>
<td>5 Min</td>
<td>95°C</td>
</tr>
<tr>
<td>Cycling (25-40 x):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturing</td>
<td>30 Sec</td>
<td>95°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>30 Sec</td>
<td>$T_m - 5^\circ$C</td>
</tr>
<tr>
<td>Extension</td>
<td>1 Min/kb</td>
<td>72°C</td>
</tr>
<tr>
<td>Final extension</td>
<td>10 Min</td>
<td>72°C</td>
</tr>
<tr>
<td>Cooling</td>
<td>Infinite</td>
<td>4-10°C</td>
</tr>
</tbody>
</table>
A standard PCR setup

PCR buffer
• MgCl\textsubscript{2} concentration between 0.5 and 5 mM, ideally 1.5 mM

dNTPs
• Each dNTP at 0.2 mM

Primer 1
• 0.1 - 1.0 $\mu$M

Primer 2
• 0.1 - 1.0 $\mu$M

Template DNA
• 0.1 - 1 $\mu$g/100 $\mu$l

DNA polymerase
• Taq: 0.5 - 2.5 U/100 $\mu$l

PCR additives
• DMSO 0 - 10%
• Betaine (Qiagen Q-solution) 1 M

DNA templates

Directly on cells
• E.g. for when doing 16S rRNA gene amplification or clone selection
• Can be inefficient due to inhibition by proteins and other cell components

„Fast“ chromosomal DNA
• Dissolve cells in sterile water
• Heat 5 Min at 95°C
• Centrifuge
• Transfer supernatant to new tube
• Use as template

Isolated DNA
• Best for difficult PCRs
• More work, but better results
DNA polymerase

- Taq DNA polymerase (from *Thermus aquaticus*)
  - Fast (around 150 bases/s)
  - Relatively high error rate (1 error in 1,000-10,000 bases)
  - No proofreading
  - Terminal adenylation

- Proofreading enzymes
  - E.g. Pfu DNA polymerase (from *Pyrococcus furiosus*)
  - Slower (around 25 bases/s)
  - High fidelity (1 error in 1 Mio bases)

- Modified enzymes
  - Hotstart Taq \(\rightarrow\) slow release of Taq to enhance stability over a run

Designing primers

The most critical parameter for a successful PCR!

In general, a good primer should match following:

- 16 – 30 nucleotides
- G+C content is between 20 – 80%
- No long stretches of identical nucleotides (especially G’s)
- No more than 2 G’s or C’s at 3’ end
- \(T_m\) between 50°C and 60°C
- The 3’ end should not contain self-complementaries (e.g. –CATG-3’)
- No mismatch at 3’ end (5-7 bases)

The design of primers is generally difficult to match with all the above rules dependent on

- Specificity
- Aim of primer
Calculating the melting temperature

- Wallace formula
  \[ T_m = \frac{\#(A+T) \times 2 + \#(G+C) \times 4}{820} > 60 \]

- Salt-corrected method
  \[ T_m = 100.5 + 41 \times \frac{\#(G+C)}{\#(T+A+G+C)} - \frac{820}{\#(T+A+G+C)} + 16.6 \times \log_{10}([Na^+]) \]

- Software for \( T_m \) calculation available
- Indicated by DNA synthesis companies

Rules for primer pairs

- \( T_m \) as close as possible together
  - Adaptation of annealing temperature to primer with lowest \( T_m \)
- No primer-primer complementaries
- No primer-dimer formation
- No side products
  - Difficult to avoid
  - Can be result of one or two primers
Specificity of primers

- Amplification at 3’ end of primer
- 5’ end does not deal in specificity

Mismatch in:
- First 5 - 7 bases from 3’ end: 
  - no amplification
- Base 6 - 12: 
  - lower efficiency of annealing
- Rest: 
  - nearly no influence

Specificity is defined clearly at the 3’ end!!!!

What if PCR fails?

- Step 1: repeat PCR
  - Forgotten something?
- Step 2: check starting material
  - Genomic DNA quality
  - dNTPs
- Step 3: check primer concentrations and sequences
- Step 4: check thermostable polymerase
- Step 5: start again
Other common problems

- Additional bands
  - Optimize annealing temperature
  - Optimize elongation time
  - Optimize PCR mix
  - Try with one primer each
  - Check primer sequences

- Smear
  - Too much template
  - Too many cycles
  - Degraded sample DNA

Contamination

Modern labs devote tremendous effort to avoiding cross-contamination

Example:

“A technician setting up a PCR reaction (with blood from the crime scene) after pipetting the positive control (with blood from the suspect) could cross contaminate the samples, even if the technician changed pipette tips between samples. A few blood cells could volatilize in the pipette, adhere to the inside of the pipette, and get ejected into the next sample.”
Gel electrophoresis

Theo H. M. Smits
COST 1104 Training Course, 24.09.2015

Agarose gel electrophoresis

- Separation of DNA according to molecular weight of the DNA fragment
- DNA is negatively loaded, and thus runs towards plus-pole
Staining agarose gels

- Two options:
  - Staining after running
  - In-gel staining
- Intercalating dyes
  - Ethidium bromide
  - GelRed

Influence of agarose concentration

- Running distance dependent on agarose concentration

<table>
<thead>
<tr>
<th>% Agarose</th>
<th>Detection range (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>10 - 0.8</td>
</tr>
<tr>
<td>1.0</td>
<td>7 - 0.5</td>
</tr>
<tr>
<td>1.2</td>
<td>6 - 0.4</td>
</tr>
<tr>
<td>1.5</td>
<td>4 - 0.2</td>
</tr>
<tr>
<td>2.0</td>
<td>3 - 0.1</td>
</tr>
</tbody>
</table>

- Potentially carcinogenic
Why not use circular fragments for gel electrophoresis

- Different forms of circular DNA present in cell

Agarose Gel Electrophoresis of DNA

(slot to introduce DNA sample into gel)

Cathode

Form I (linear)

Form II (supercoiled)

Anode

Direction of Migration of DNA
Sanger sequencing

Theo H. M. Smits
COST 1104 Training Course, 24.09.2015

Definition of sequencing

• The process of determining the precise order of nucleotides within a DNA molecule

• 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

• The advent of rapid DNA sequencing methods has greatly accelerated biological and medical research and discovery.
**History of sequencing**

- Maxam/Gilbert DNA sequencing
  - First widely-adopted method for DNA sequencing (1977)
  - Method based on radioactive nucleotides and chemical separation after certain bases
  - Replaced by Sanger sequencing after improvement of chain termination method
  - Technical complexity prohibiting its use in standard molecular biology kits
  - Extensive use of hazardous chemicals
  - Difficulties with scale-up

---

**The Maxam-Gilbert method of sequencing**

- First widely-adopted method for DNA sequencing (1977)
- Method based on radioactive nucleotides and chemical separation after certain bases
- Got replaced by Sanger sequencing after improvement of chain termination method
- Technical complexity prohibiting its use in standard molecular biology kits
- Extensive use of hazardous chemicals
- Difficulties with scale-up
Chain termination sequencing

Also known as Sanger sequencing

Frederic Sanger
- British chemist (1918-2013)
- Only person with two noble prizes in chemistry
  - 1958: AA sequence and structure of insulin
  - 1980: DNA sequencing
- Published the first complete sequence of an organism in 1975: bacteriophage φX174

Chain termination sequencing

- Modified sugar residue in nucleotides
- Random incorporation of ddNTPs in elongating chains
Chain termination sequencing

- Four tubes, each containing one ddNTP at low concentration
- Initially done in presence of radiolabeled nucleotides
  - $^{32}$P-dATP
  - $^{35}$S-dATP
- Separation on polyacrylamide gels

Automated gel sequencers

- Improvements based on use of fluorescently labeled primers
- Advantages:
  - Good readable sequences
  - Long reads (up to 1000 bp)
- Disadvantages:
  - 4 reactions/template
  - 4 lanes/template
  - 1 run/day possible
  - Small inconsistencies in gel make automated reading complicated
  - Requires often manual reading
Fluorescent-labeled dye sequencing

- Use of fluorescent dyes linked to the ddNTPs
- Different dyes for different bases

- Advantages:
  - More sensitive method
  - Requires less template
  - One reaction/template
  - One lane/template

- Disadvantages:
  - Platform-dependent
  - Patent-protected chemistry

Fluorescence-labeled dye sequencing – chemistry

- Every base different dye with different emission spectrum
- Automated correction for overlaps
Fluorescence-labeled dye sequencing – method

- Allows single lane detection

First-generation fluorescence-labeled dye sequencers

- Good quality of reads
- Gel-based system
Second-generation fluorescence-labeled dye sequencers

- Use of capillaries to improve consistency of reads
- User-friendly
  - No gel preparation
  - Automation of
    - Capillary filling
    - Sample loading

Quality control: Phred scores

- BaseCaller indicated how reliable the call is
- Low score → low security
Phred scores

<table>
<thead>
<tr>
<th>Phred quality score</th>
<th>Probability that the base is called wrong</th>
<th>Accuracy of the base call</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1 in 10</td>
<td>90%</td>
</tr>
<tr>
<td>20</td>
<td>1 in 100</td>
<td>99%</td>
</tr>
<tr>
<td>30</td>
<td>1 in 1,000</td>
<td>99.9%</td>
</tr>
<tr>
<td>40</td>
<td>1 in 10,000</td>
<td>99.99%</td>
</tr>
<tr>
<td>50</td>
<td>1 in 100,000</td>
<td>99.999%</td>
</tr>
</tbody>
</table>

- The higher the score, the lower the chance that a base call is wrong
- Logarithmic scale!
- Output: in .qual or .fastq files

Sequence assembly of Sanger reads

- In bioinformatics, sequence assembly refers to aligning and merging fragments of a (much longer) DNA sequence in order to reconstruct the original sequence
- To have confidence in what you sequenced, you always need to sequence both strands of your DNA sequence
  - Forward
  - Reverse
- With small DNA fragments, this is easily done
Sequence assembly of Sanger reads

- Forward and reverse read combined into a single sequence
- Overhangs left and right are primer sequences
- Coverage: twofold
- Possibility of correcting sequencing errors
- End product is a consensus sequence

Error correction

- Partial peak view of two reads
- Visible errors
  - 1) Weak identification of peaks
  - 2) Missing base
- Possible to correct with good reverse sequence
Will Sanger sequencing die out soon?

- Next Generation Sequencing (2nd/3rd) is very efficient in sequencing large amounts of genomes at the time
  - Mb to Tb amounts per run
  - Getting cheaper and cheaper
  - Very efficient bioinformatics

- Other sequencing technologies emerge

- Nevertheless: Sanger sequencing is still more than sufficient for a whole lot of small-scale applications

Coming back...

Jerry is REALLY old. Apparently he once actually poured and ran a sequencing gel.