Chapter 5
Chemical Synthesis, Sequencing, and Amplification of DNA

1. Chemical Synthesis of DNA

Chemical Synthesis of DNA

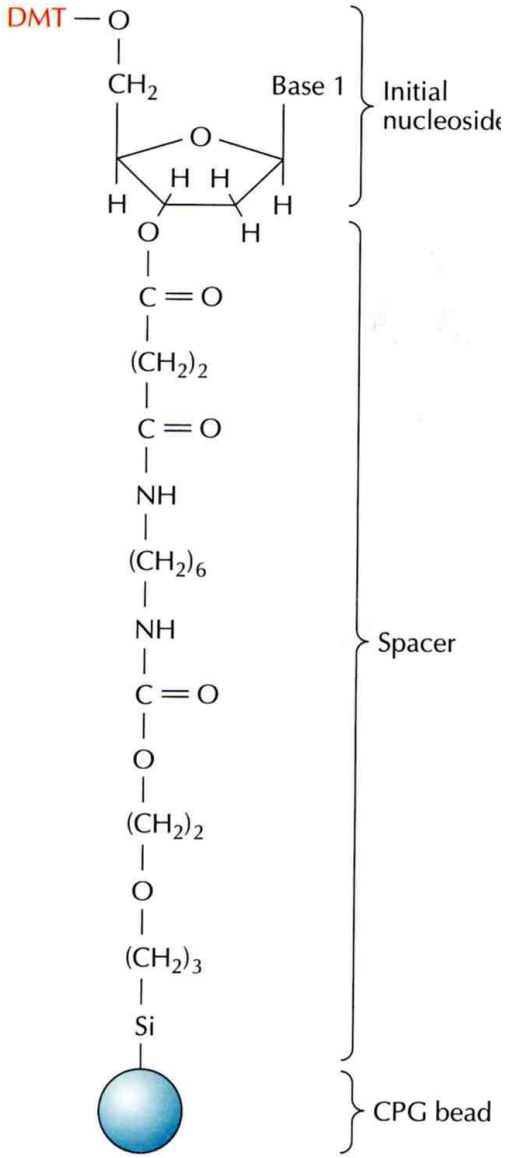
- Usage of chemically synthesized DNA oligonucleotides
  - PCR
  - Mutagenesis
  - Library screening
  - DNA sequencing
  - Gene cloning
- DNA synthesizer: gene machine
  - Synthesis of DNA oligos
  - Coupling of incoming nucleotide to 5’ OH terminus of the growing chain

The Phosphoramidite Method

- Attachment of the first nucleoside (3’ end)
- Synthesis cycle → n cycle
- Removing oligonucleotide from column
- Purifying oligonucleotide

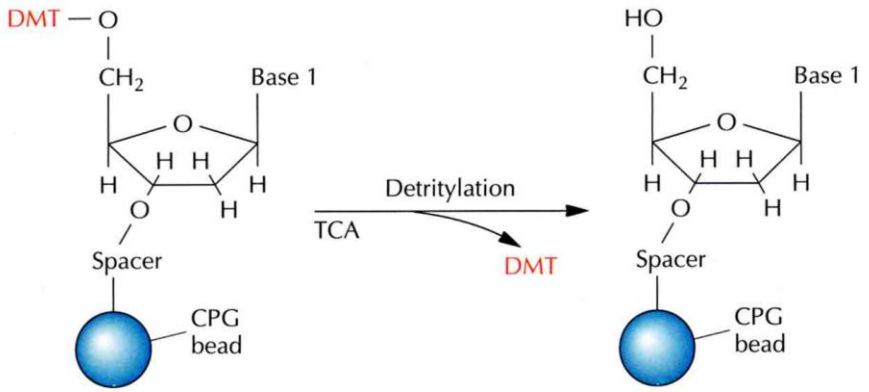
The Phosphoramidite Method

- Attachment of the first nucleoside (3’ end)
  - Nucleoside
    - Linked to spacer molecule through 3’ OH
    - Attachment of dimethoxytrityl (DMT) group to 5’ OH
  - Attachment of spacer to inert support (controlled pore glass (CPG) bead)



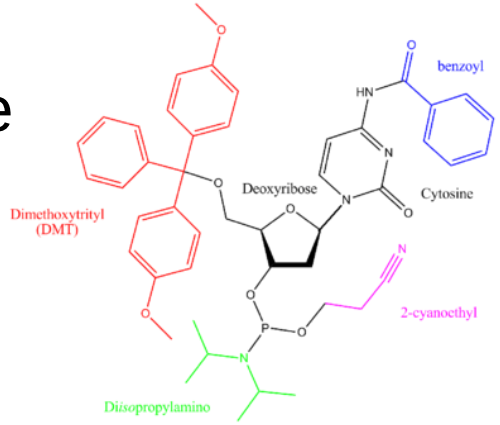
The Phosphoramidite Method

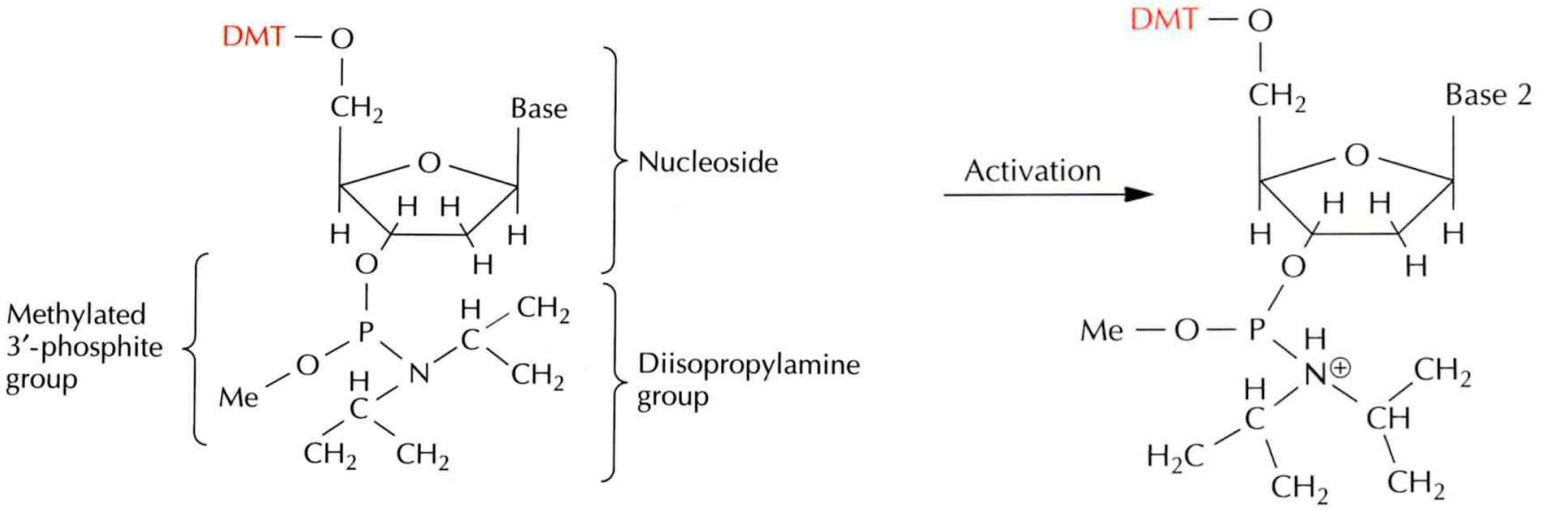
- Synthesis cycle
  - Washing reaction column with an anhydrous reagent (acetonitrile)
  - Detritylation
    - Removal of DMT by trichloroacetic acid (TCA) treatment
  - Washing
  - Activation and coupling
  - Washing
  - Capping
  - Oxidation



Activation and Coupling

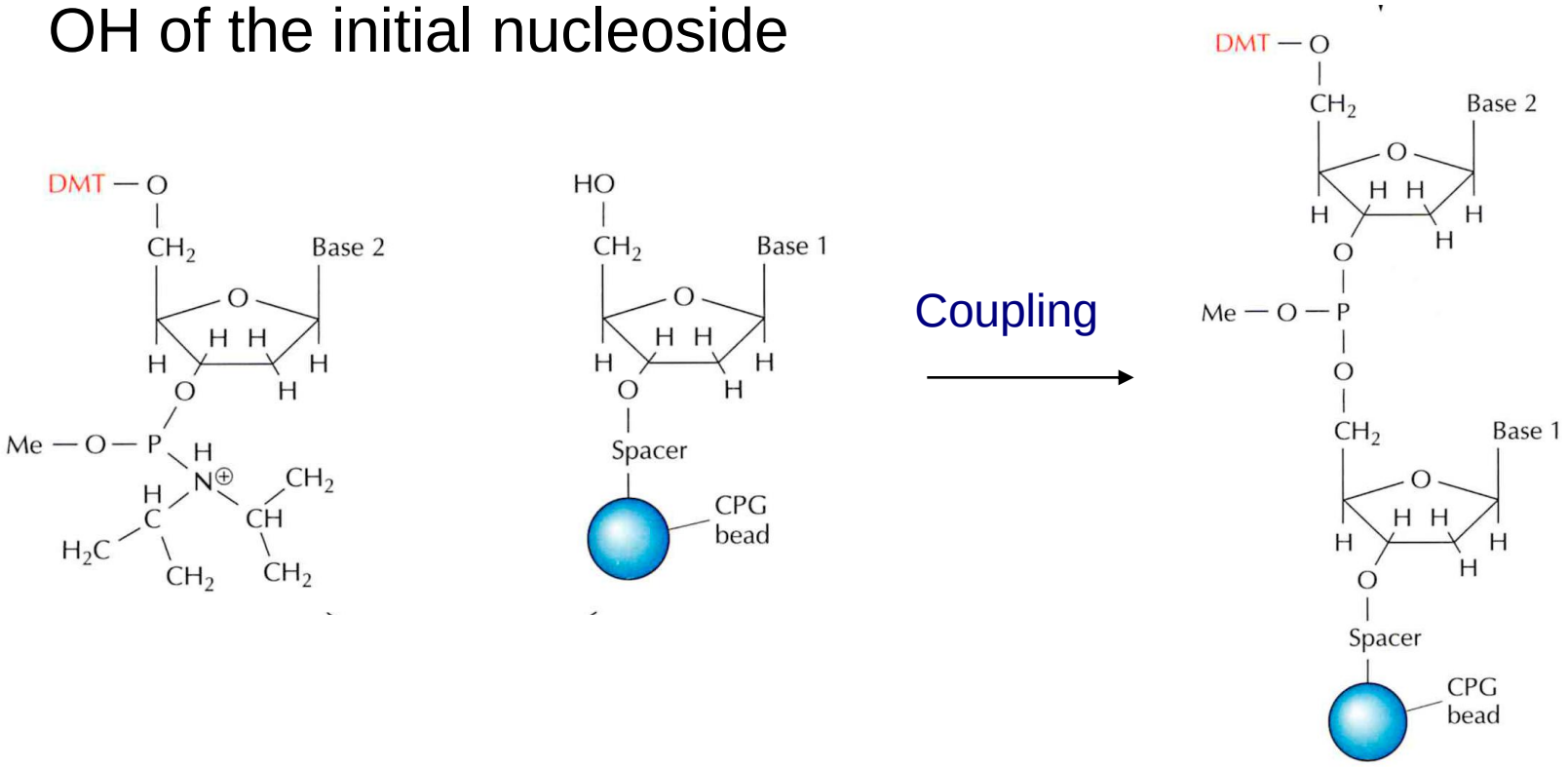
- Nucleoside phosphoramidite as substrate
  - Modification of amino groups of bases
    - A: benzoyl, G: isobutyryl, C: benzoyl
- Activation and coupling by tetrazole





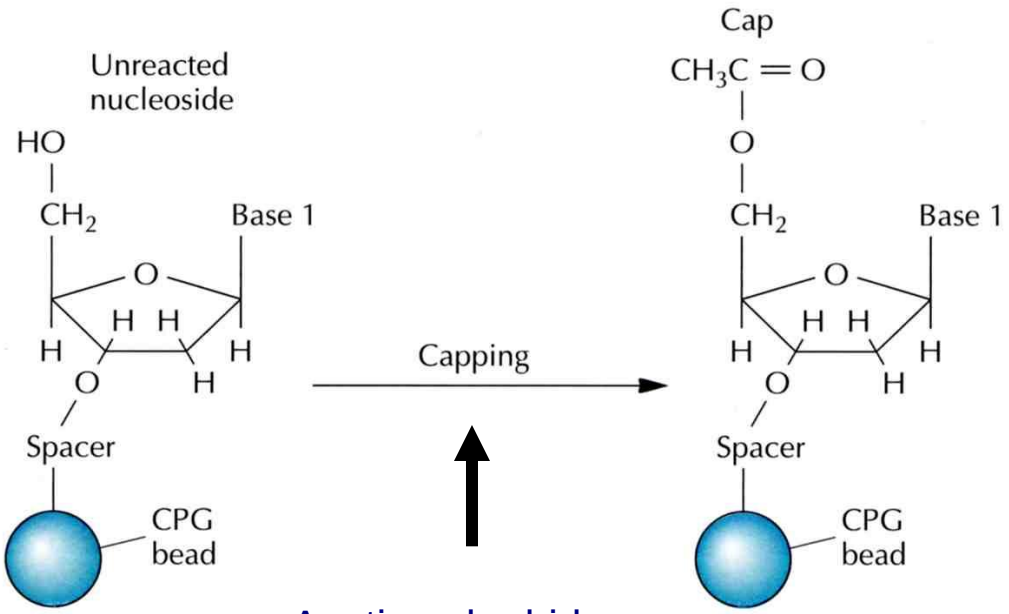
Activation and Coupling

- Covalent bonding of 3’ phosphite with 5’ OH of the initial nucleoside



Capping

- Blocking unlinked residues by acetylation of unreacted 5’ OH

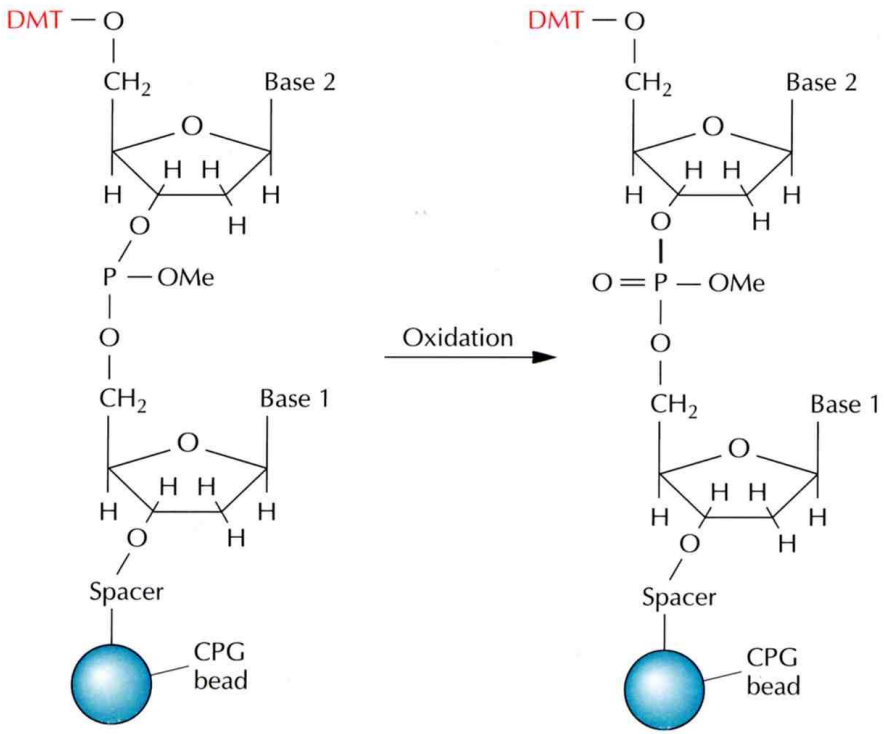


Acetic anhydride

Dimethyl-aminopyridine

Oxidation

- Oxidation of unstable phosphite triester bond with an iodine mixture → pentavalent phosphate triester

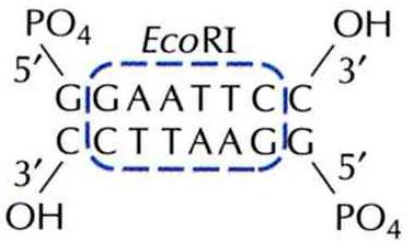


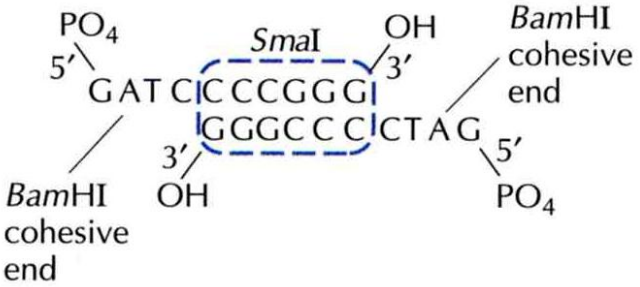
Purification of synthesized oligonucleotide

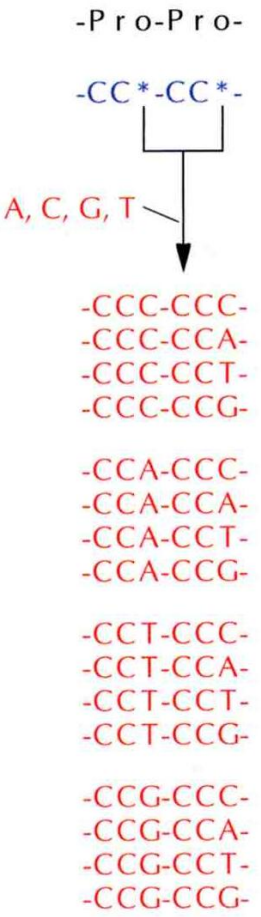
- Removing oligonucleotide from column
  - Removal of methyl group
  - Cleavage and elution of DNA strands from the spacer
  - Removal of benzoyl and isobutyryl group from bases
  - Detritylation
  - Phosphorylation of 5’ terminus
- Purification by reverse-phase high-pressure liquid chromatography or gel electrophoresis
  - Yield: depending on coupling efficiency
  - 99% efficiency, 20 mer → 82% final product (0.99^{20} X 100)

Uses of Synthesized Oligonucleotides

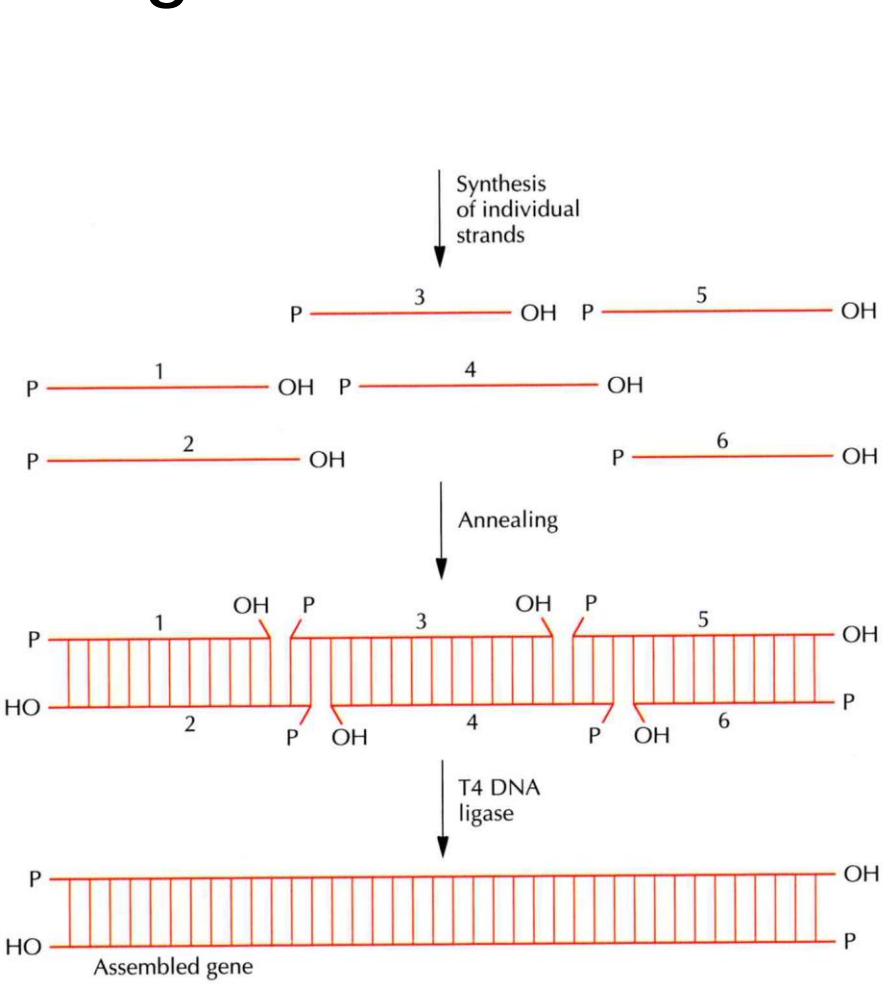
- Degenerate primers for library screening (20~40 mer)
- DNA sequencing and PCR (17~24mer)
- Usage for cloning
  - Linkers
    - Double-stranded, blunt-ended DNA with restriction enzyme site
  - Adaptors
    - Annealed oligos generating restriction enzyme-digested ends

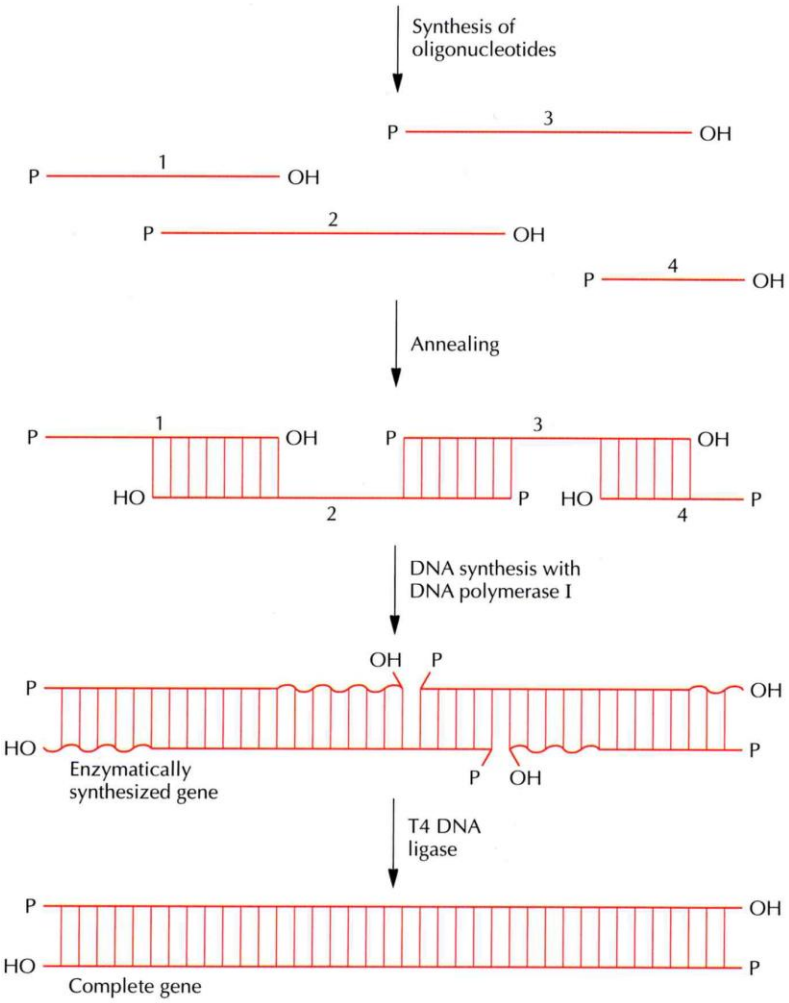






Building Synthetic Genes with Oligonucleotides



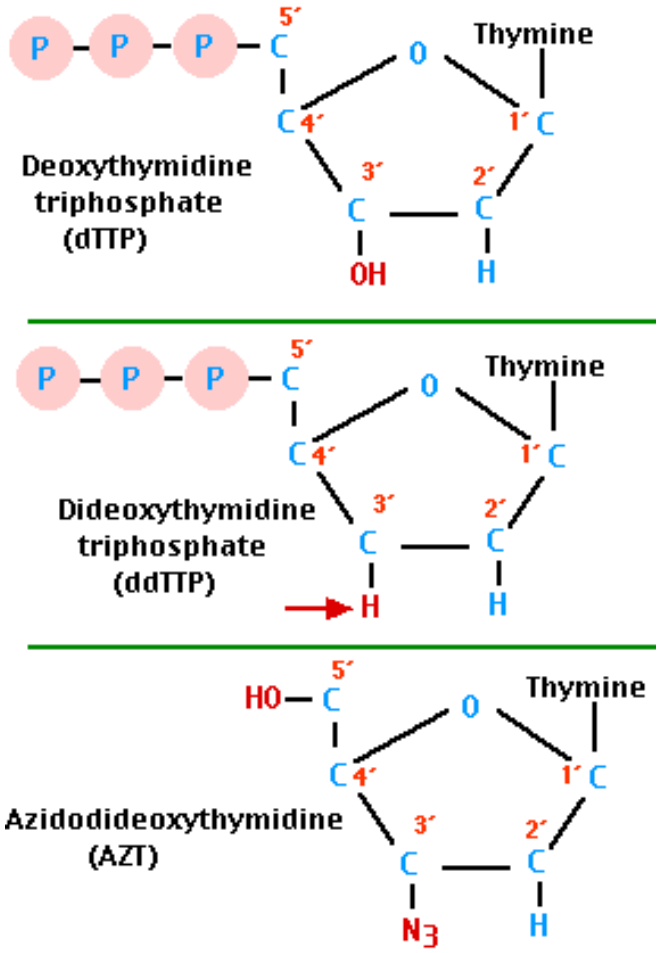


2. DNA Sequencing Techniques

DNA Sequencing Methods (1977)

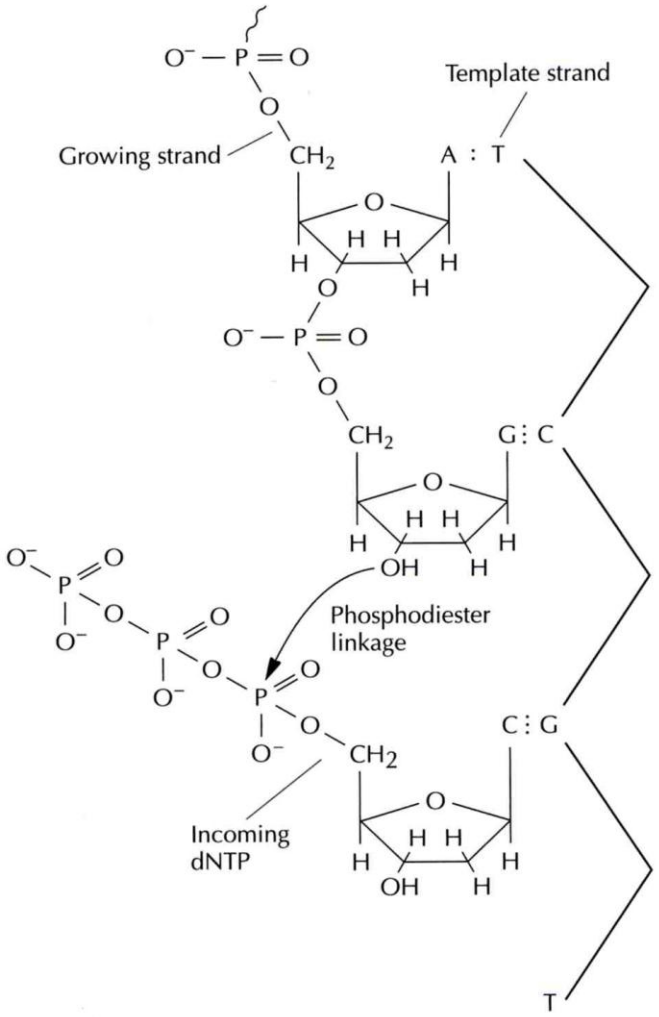
- Alan Maxam and Walter Gilbert
  - Chemical protocol
- Fred Sanger
  - Enzymatic procedures
  - Chain termination
  - Dideoxynucleotide method

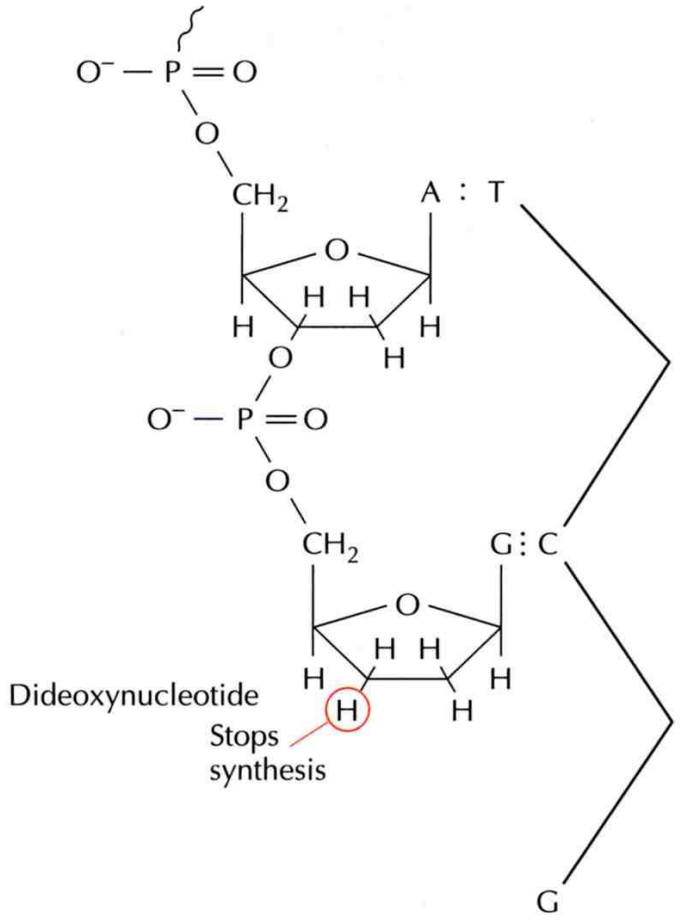
Dideoxynucleotide (ddNTP)



Anti-AIDS drug

Chain Termination by Dideoxynucleotide

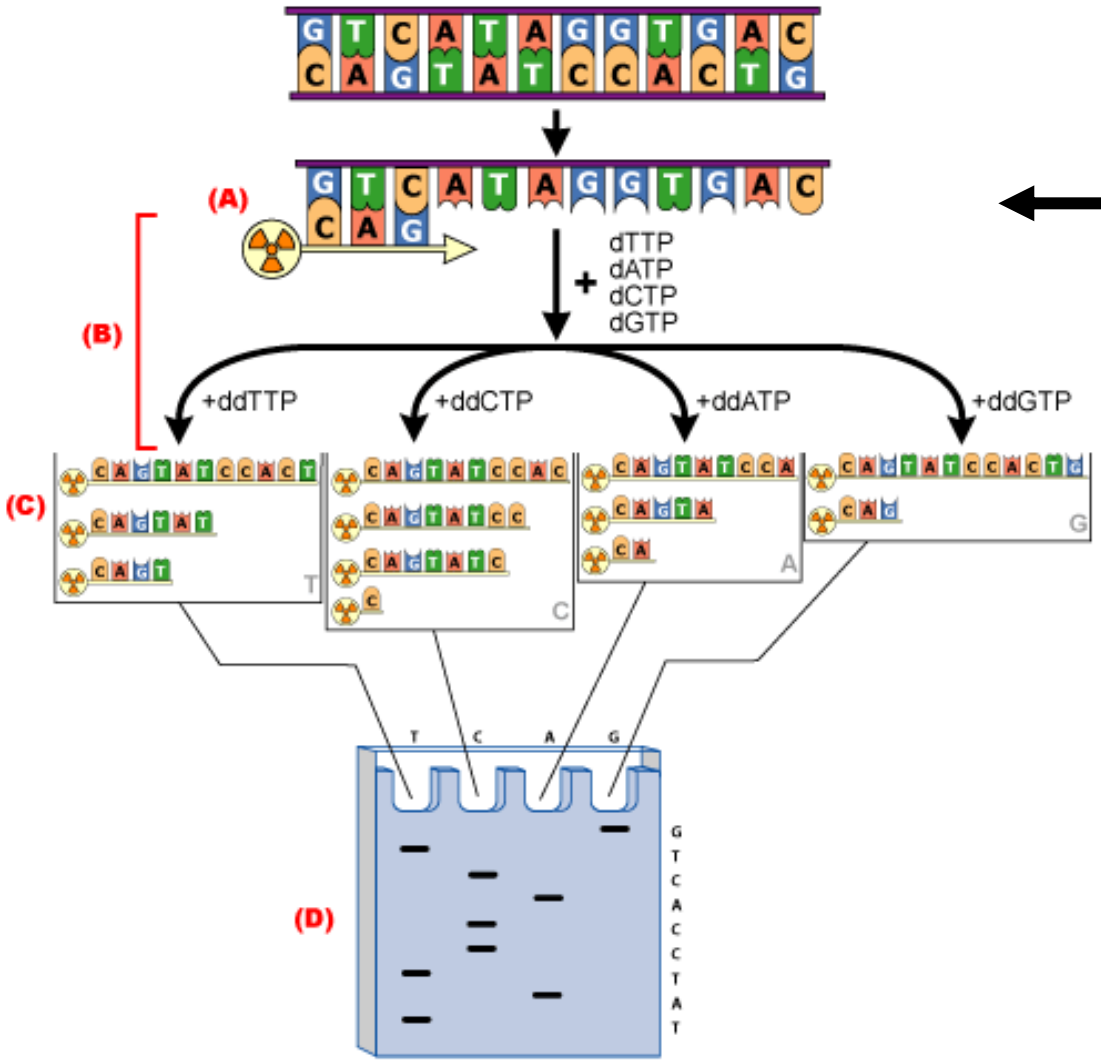




Chain Termination Sequencing Procedure

- Annealing of sequencing primers (17 ~ 24 mer) to a template
- 4 reaction tubes
  - dNTPs (one of which is radio-labeled, α-^{35}S dATP)
  - One of the four ddNTPs
- DNA synthesis with DNA polymerase
- Stop the reaction by formamide
- Separation of DNA by denaturing PAGE
- Autoradiography

Chain Terminator Sequencing

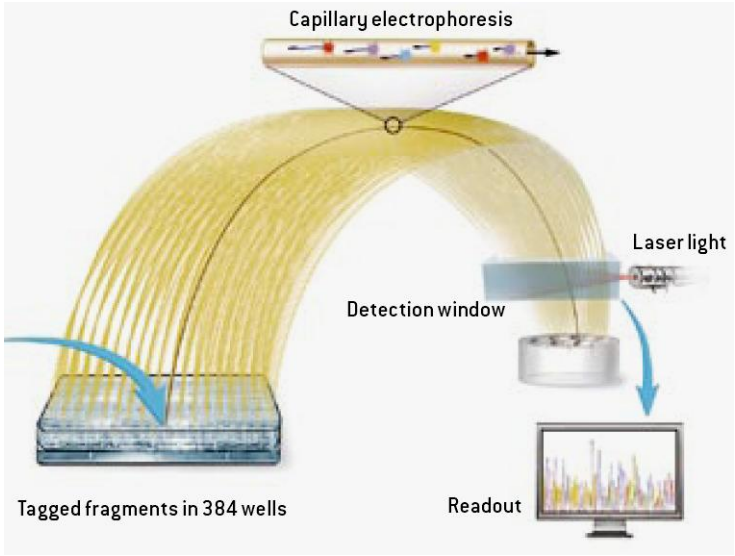


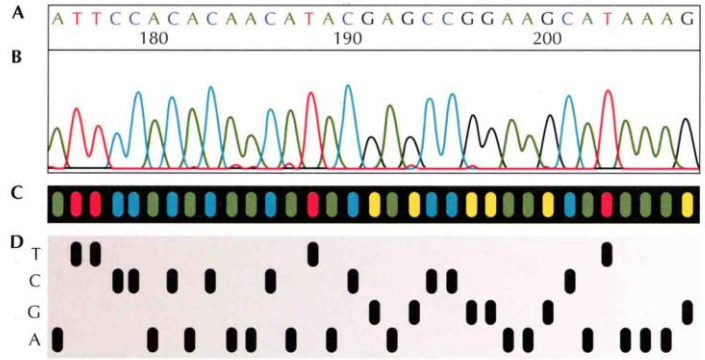
T7 DNA polymerase (Sequenase)

Taq polymerase

Automated DNA Sequencing

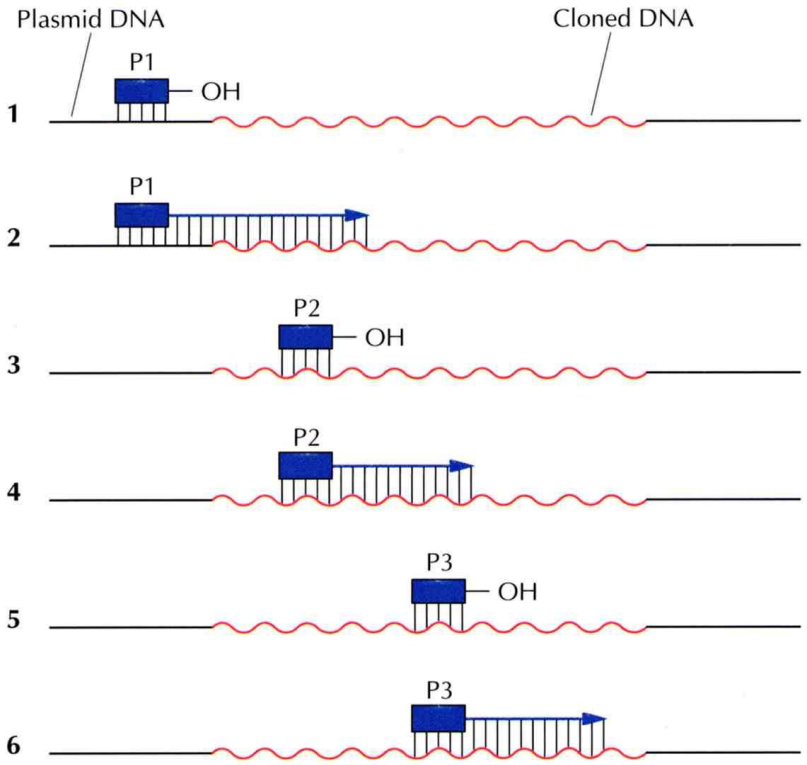
- Labeling ddNTPs with four different fluorescence dyes
  - 4-color, 1-lane detection
  - Run samples in one lane of polyacrylamide or polymer-filled capillary tube
- Detection of the fluorescence by an argon ion laser beam at the bottom of the electrophoretic matrix





Sequencing of Large DNA

- Restriction mapping and subcloning to sequence
- Primer walking
  - Designing serial sequencing primers based on the previous sequence data



3. Next Generation Sequencing

Recent advances in sequencing methods

- Sequencing by hybridization
- Sequencing by ligation
- Mass spectrometry-based sequencing
- Sequence-specific detection of single-stranded DNA using engineered nanopores
- Sequencing by synthesis (SBS)
  - Pyrosequencing
  - Using cleavable fluorescent nucleotide reversible terminators

The Era of Personal Genomics

nature Vol 437|15 September 2005|doi:10.1038/nature03959

ARTICLES

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies[1]*, Michael Egholm[1]*, William E. Altman[1], Said Attiya[1], Joel S. Bader[1], Lisa A. Bemben[1], Jan Berka[1], Michael S. Braverman[1], Yi-Ju Chen[1], Zhoutao Chen[1], Scott B. Dewell[1], Lei Du[1], Joseph M. Fierro[1], Xavier V. Gomes[1], Brian C. Godwin[1], Wen He[1], Scott Helgesen[1], Chun He Ho[1], Gerard P. Irzyk[1], Szilveszter C. Jando[1], Maria L. I. Alenquer[1], Thomas P. Jarvie[1], Kshama B. Jirage[1], Jong-Bum Kim[1], James R. Knight[1], Janna R. Lanza[1], John H. Leamon[1], Steven M. Lefkowitz[1], Ming Lei[1], Jing Li[1], Kenton L. Lohman[1], Hong Lu[1], Vinod B. Makhijani[1], Keith E. McDade[1], Michael P. McKenna[1], Eugene W. Myers[2], Elizabeth Nickerson[1], John R. Nobile[1], Ramona Plant[1], Bernard P. Puc[1], Michael T. Ronan[1], George T. Roth[1], Gary J. Sarkis[1], Jan Fredrik Simons[1], John W. Simpson[1], Maithreyan Srinivasan[1], Karrie R. Tartaro[1], Alexander Tomasz[3], Kari A. Vogt[1], Greg A. Volkmer[1], Shally H. Wang[1], Yong Wang[1], Michael P. Weiner[4], Pengguang Yu[1], Richard F. Begley[1] & Jonathan M. Rothberg[1]

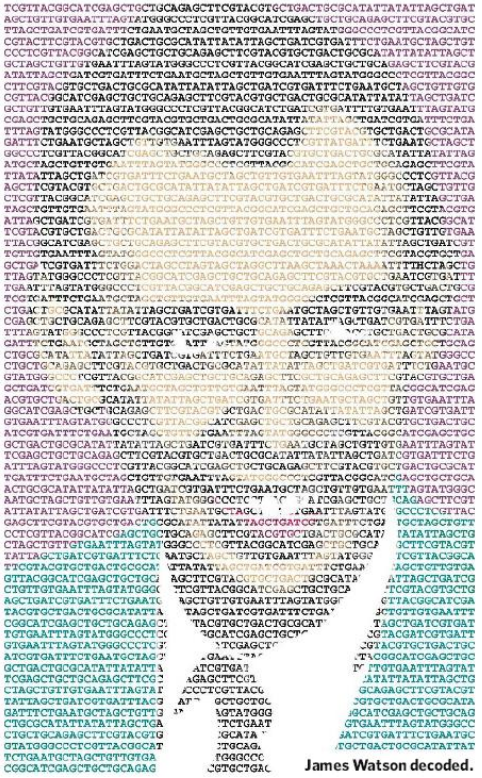
The proliferation of large-scale DNA-sequencing projects in recent years has driven a search for alternative methods to reduce time and cost. Here we describe a scalable, highly parallel sequencing system with raw throughput significantly greater than that of state-of-the-art capillary electrophoresis instruments. The apparatus uses a novel fibre-optic slide of individual wells and is able to sequence 25 million bases, at 99% or better accuracy, in one four-hour run. To achieve an approximately 100-fold increase in throughput over current Sanger sequencing technology, we have developed an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes. Here we show the utility, throughput, accuracy and robustness of this system by shotgun sequencing and *de novo* assembly of the *Mycoplasma genitalium* genome with 96% coverage at 99.96% accuracy in one run of the machine.

nature Vol 452|17 April 2008|doi:10.1038/nature06884

LETTERS

The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler[1]*, Maithreyan Srinivasan[2]*, Michael Egholm[2]*, Yufeng Shen[1]*, Lei Chen[1], Amy McGuire[3], Wen He[2], Yi-Ju Chen[2], Vinod Makhijani[2], G. Thomas Roth[2], Xavier Gomes[2], Karrie Tartaro[2]†, Faheem Niazi[2], Cynthia L. Turcotte[2], Gerard P. Irzyk[2], James R. Lupski[4,5,6], Craig Chinault[4], Xing-zhi Song[1], Yue Liu[1], Ye Yuan[1], Lynne Nazareth[1], Xiang Qin[1], Donna M. Muzny[1], Marcel Margulies[2], George M. Weinstock[1,4], Richard A. Gibbs[1,4] & Jonathan M. Rothberg[2]†



Genome sequencing of James Watson in two months

The Era of Personal Genomics

2008.11.6

Vol 456|6 November 2008|doi:10.1038/nature07517 nature

ARTICLES

Accurate whole human genome sequencing using reversible terminator chemistry

A list of authors and their affiliations appears at the end of the paper

DNA sequence information underpins genetic research, enabling discoveries of important biological or medical benefit. Sequencing projects have traditionally used long (400–800 base pair) reads, but the existence of reference sequences for the human and many other genomes makes it possible to develop new, fast approaches to re-sequencing, whereby shorter reads are compared to a reference to identify intraspecies genetic variation. Here we report an approach that generates several billion bases of accurate nucleotide sequence per experiment at low cost. Single molecules of DNA are attached to a flat surface, amplified *in situ* and used as templates for synthetic sequencing with fluorescent reversible terminator deoxyribonucleotides. Images of the surface are analysed to generate high-quality sequence. We demonstrate application of this approach to human genome sequencing on flow-sorted X chromosomes and then scale the approach to determine the genome sequence of a male Yoruba from Ibadan, Nigeria. We build an accurate consensus sequence from >30× average depth of paired 35-base reads. We characterize four million single-nucleotide polymorphisms and four hundred thousand structural variants, many of which were previously unknown. Our approach is effective for accurate, rapid and economical whole-genome re-sequencing and many other biomedical applications.

nature Vol 456|6 November 2008|doi:10.1038/nature07484

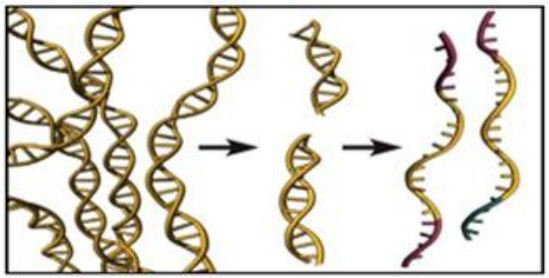
ARTICLES

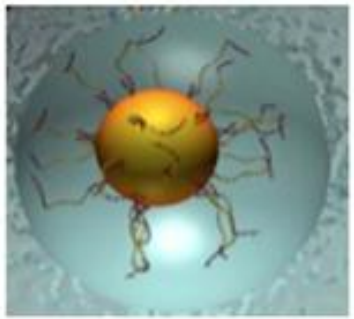
The diploid genome sequence of an Asian individual

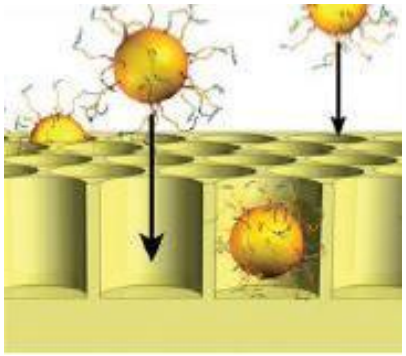
3.1. The 454 Life Sciences Sequencing Technology

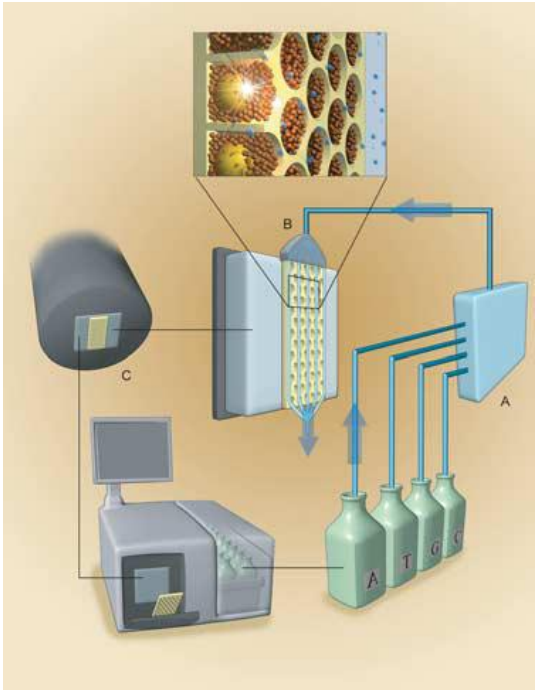
The 454 Life Sciences Sequencing Technology

- DNA library preparation : 4. 8 h
  - ssDNA with adaptors
- emPCR (emulsion PCR) : 8 h
  - 1 DNA/ bead
  - Clonal amplification in microreactor
- Sequencing : 7.5 h
  - Pyrosequencing of DNA in each bead



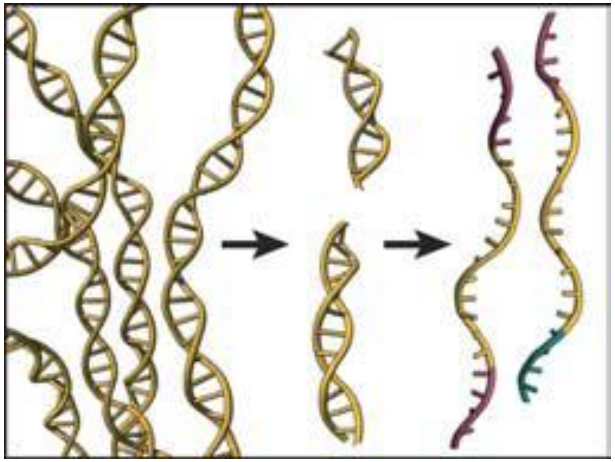


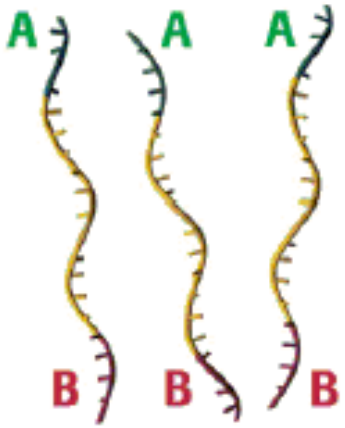




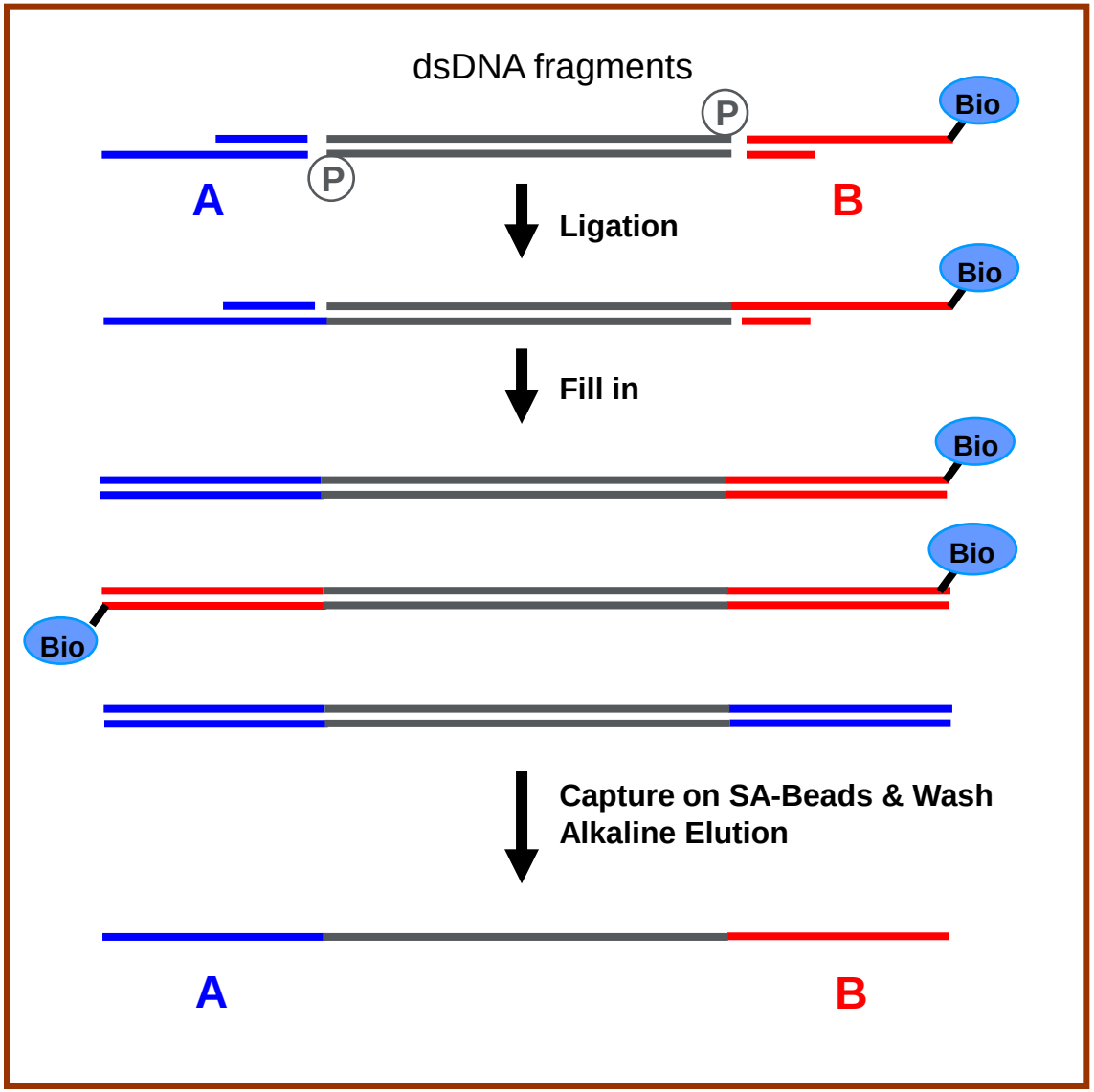
1. DNA library Preparation

- Starting materials :genomic DNA, PCR products, cDNA
  - Fragmentation by nebulization
  - 300-800 base-pair long fragments
- Ligation of A, B adaptor
- Selection of ssDNA containing both A and B adaptor





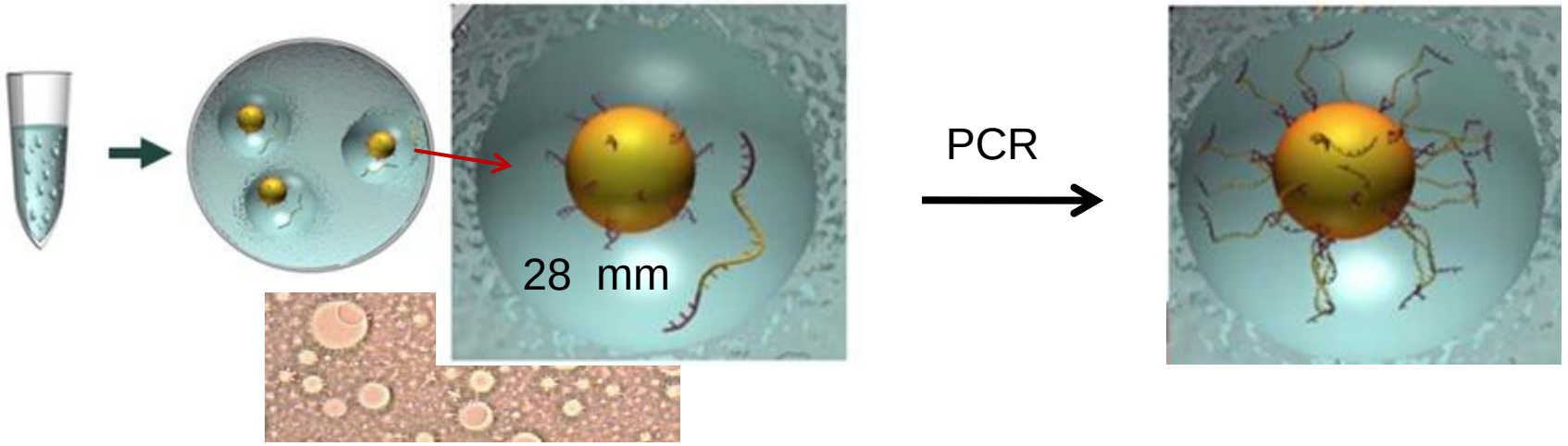
1. DNA library Preparation



- DNA end polishing
  - T4 DNA polymerase, Klenow, T4 kinase
- Ligation with adaptors A, B-biotin
  - Filling in nicks
  - Bst DNA polymerase strain displacement activity
- Isolation of ssDNA containing both A and B adaptor
  - Streptavidin beads
  - Alkaline elution

2. emPCR (emulsion PCR)

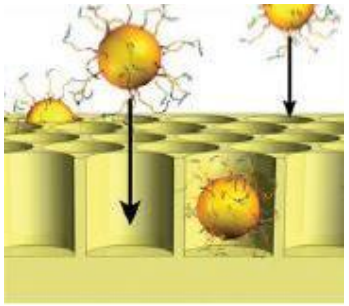
- Immobilization of a single ssDNA onto beads using adaptor (use excess of DNA capture beads)
- Emulsify beads and PCR reaction inside of the emulsion (water-in-oil microreactor)
- Clonal amplificaion inside microreactors

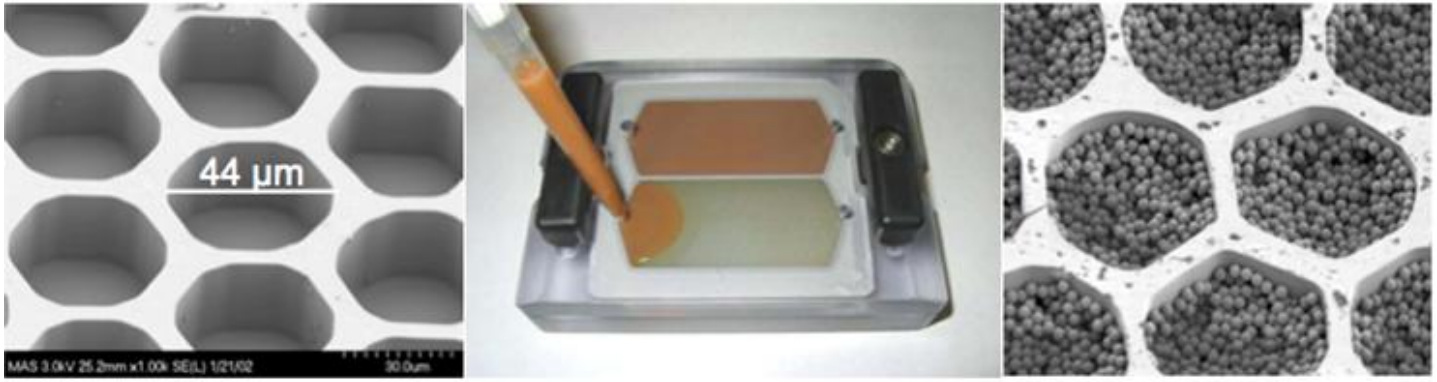


- Break microreactors and enrich for DNA-positive beads

3. Depositing DNA beads into PicoTiterDevice

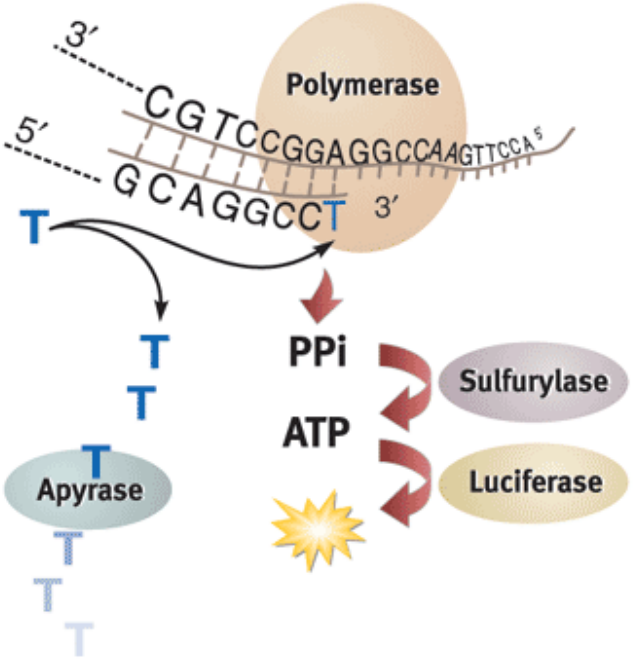
- Loading the ssDNA library beads onto the fibre-optic PicoTiterDevice for sequencing (1.6 x 10^7 wells).
  - 1 bead/well
- Layered with enzyme beads containing sulfurylase and luciferase
- 400,000 reads obtained in parallel

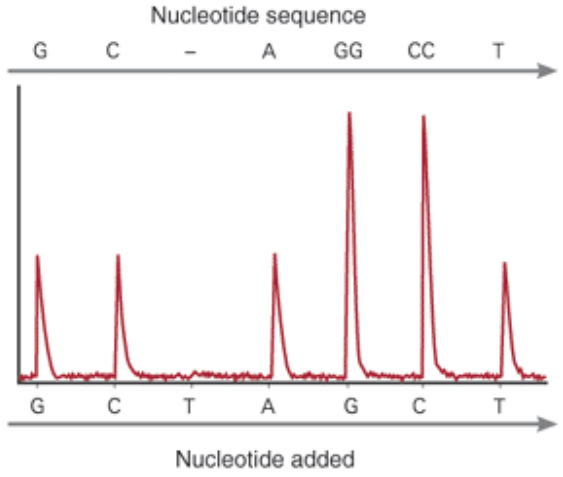




4. Sequencing

- Pyrosequencing
  - (NA)n + dNTP —Polymerase→ (NA)n+1 + PPi
  - PPi + APS (adenosine phosphosulfate) —ATP sulfurylase→ ATP + SO_4^{2-}
  - ATP + luciferin + O_2 —Luciferase→ AMP + PPi + Oxyluciferin + CO_2 + Light
  - ATP (dNTP) —Apyrase→ AMP (dNMP) + 2Pi

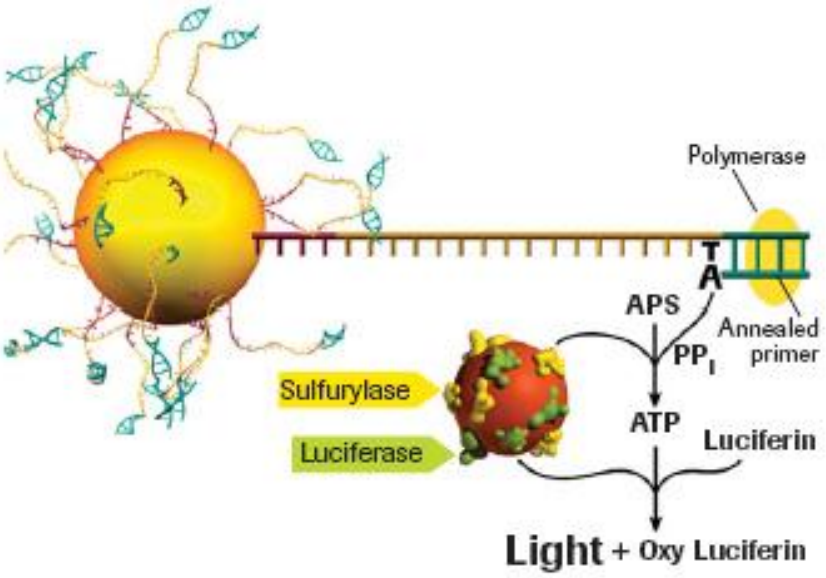




R.England & M. Pettersson, 2005, Nature Methods

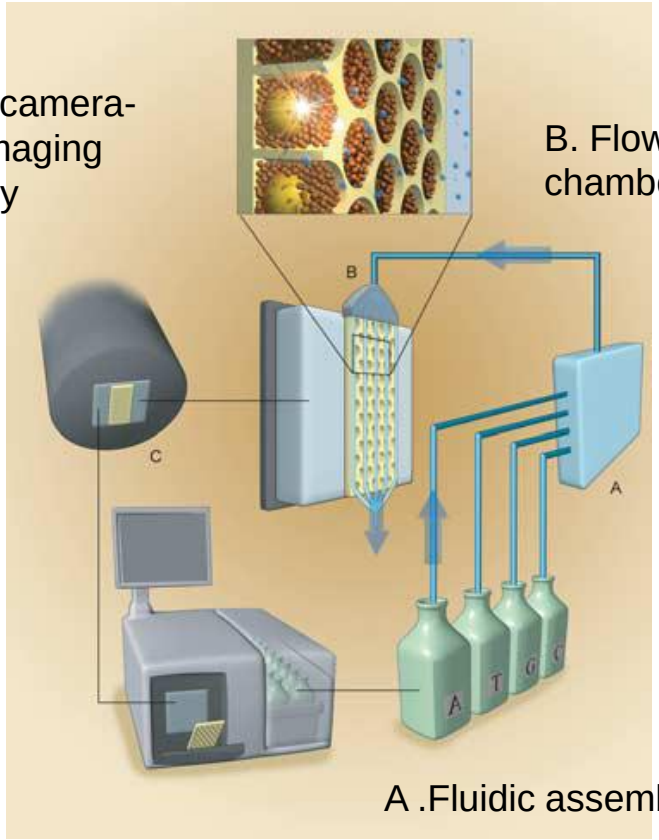
4. Sequencing

- (TACG) cycles
- Recording of a light signal generated by addition of nucleotide during polymerization reaction by CCD camera



C. CCD camera-based imaging assembly

B. Flow chamber

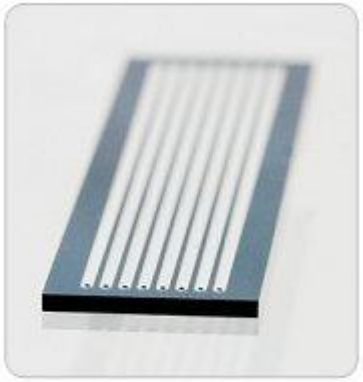


A .Fluidic assembly

3.2. Illumina’s Solexa Sequencing

Illumina's Solexa Sequencing

FIGURE 1: ILLUMINA GENOME ANALYZER FLOW CELL

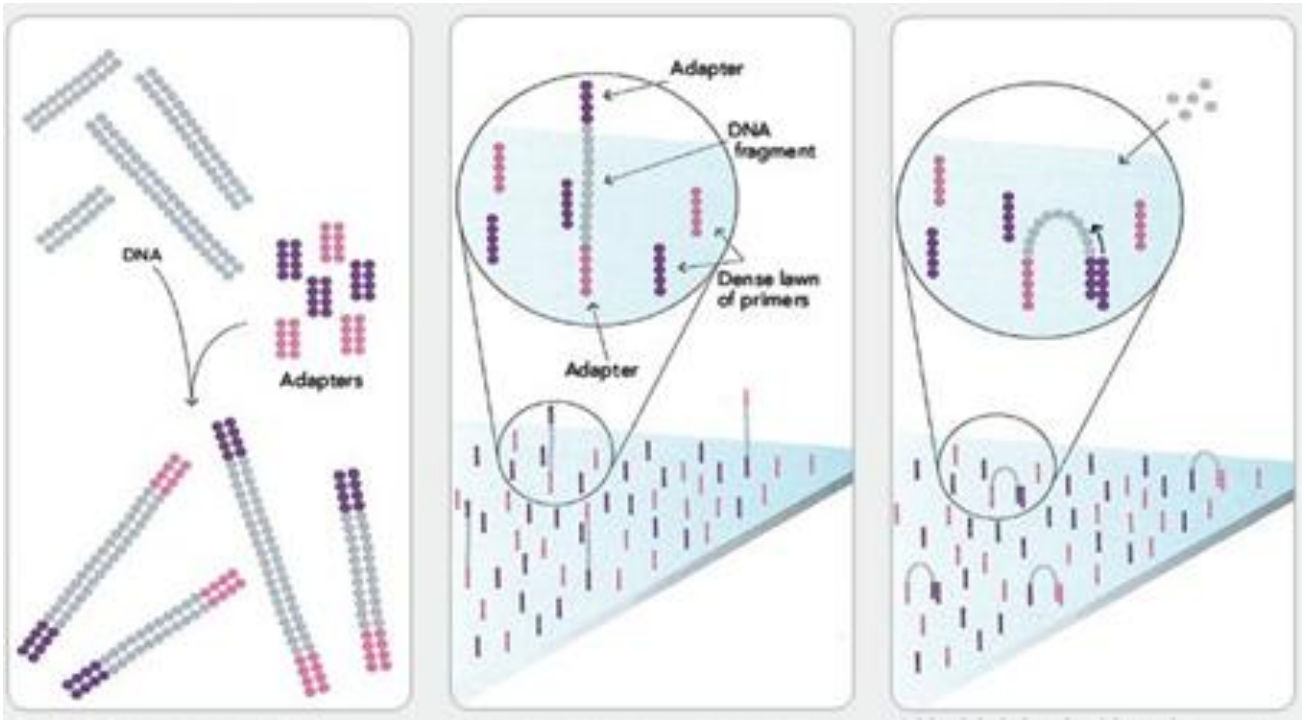


Up to eight samples can be loaded onto the flow cell for simultaneous analysis on the Illumina Genome Analyzer.

Preparation of genomic DNA

Attach DNA to surface

Bridge amplification



- Random fragmentation
- Attachment of adpators

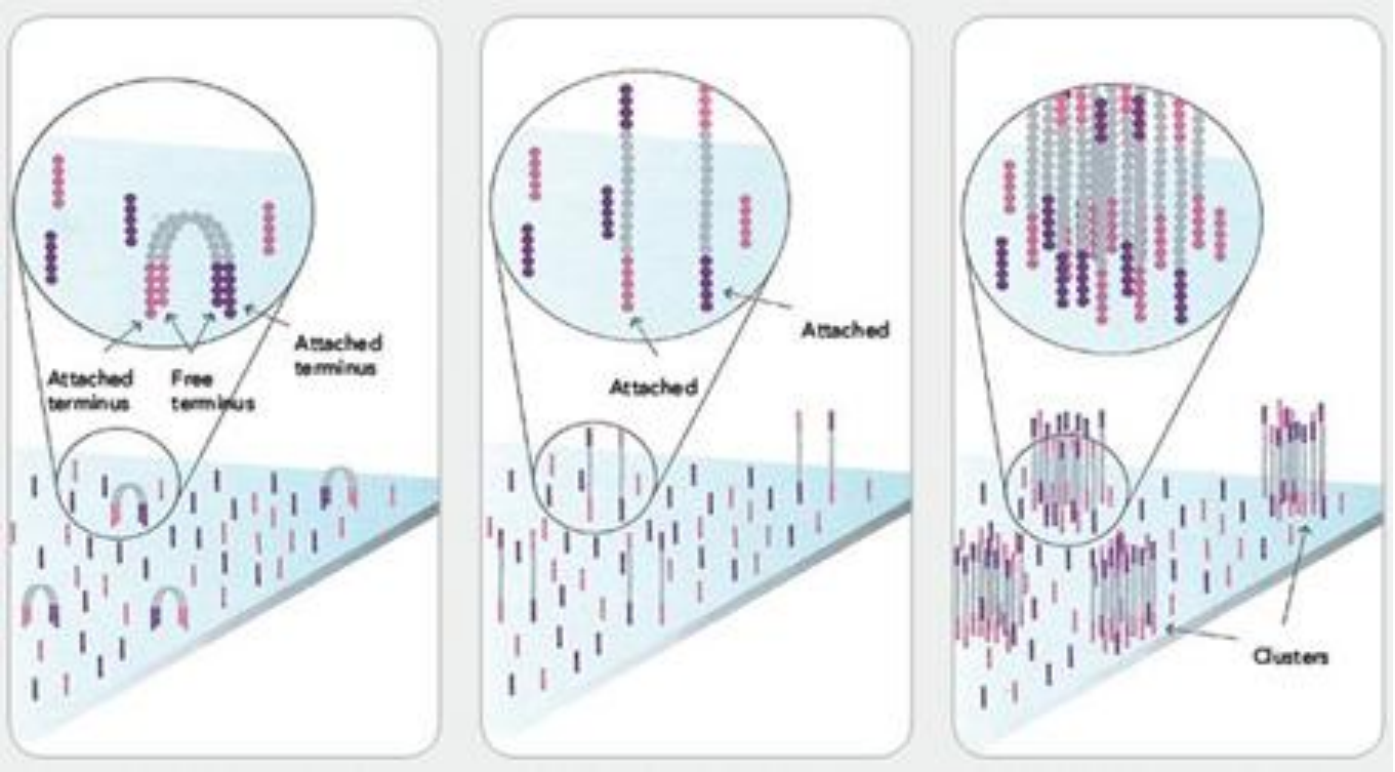
- Inside of flow cell

Illumina's Solexa Sequencing

Generation of dsDNA fragment

Denaturation of dsDNA

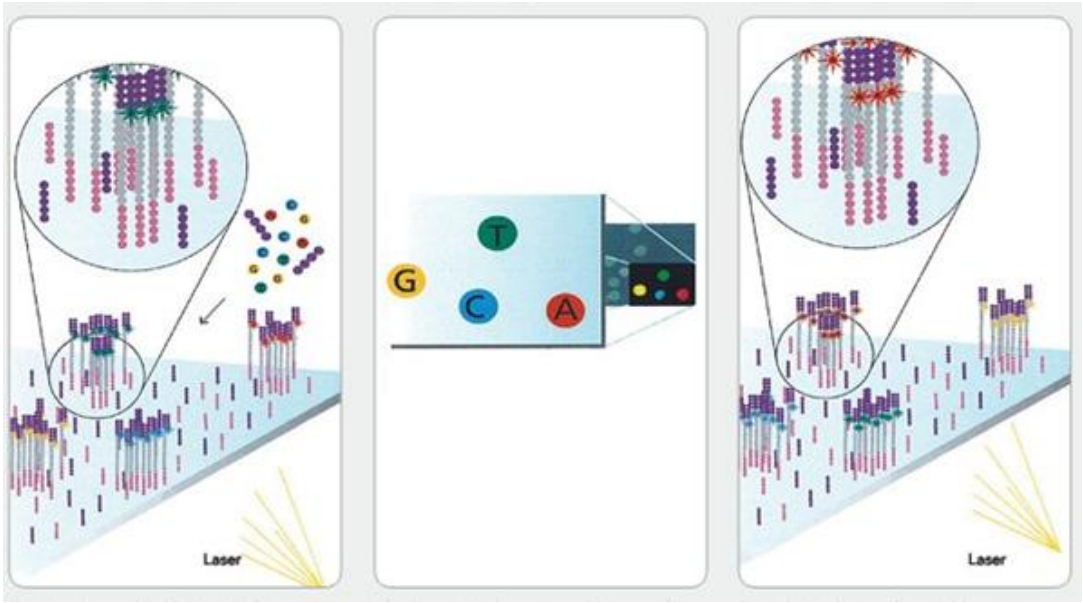
Repeat the cycle

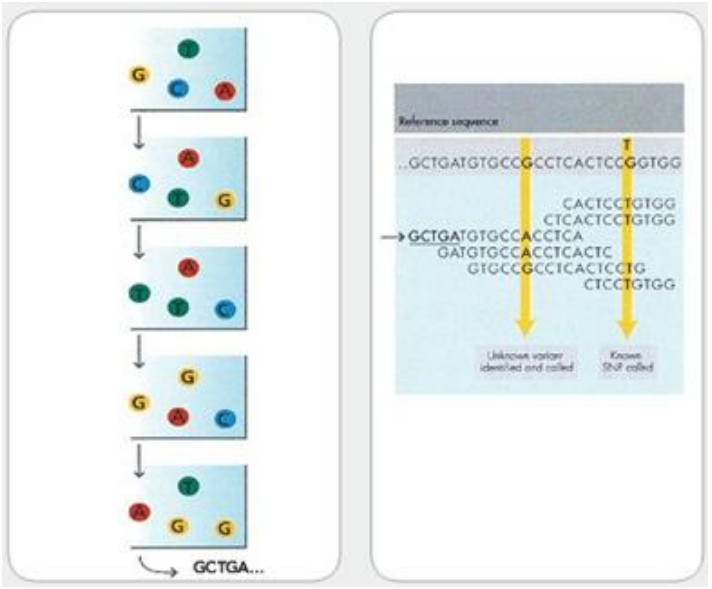


- Generation of clusters
- ≥50 million
- ~ 1000 copies each

Illumina's Solexa Sequencing

First chemistry cycle | Imaging | Second chemistry cycle | Multiple chemistry cycle | Align data

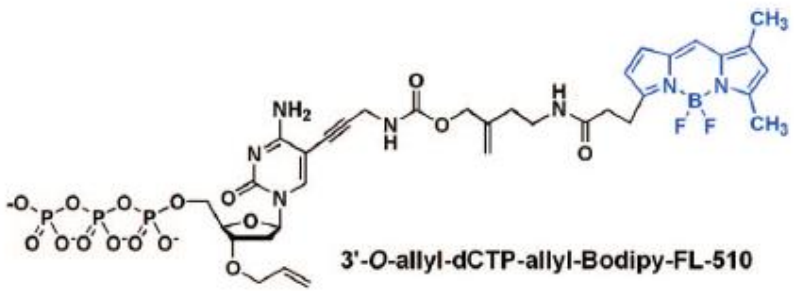


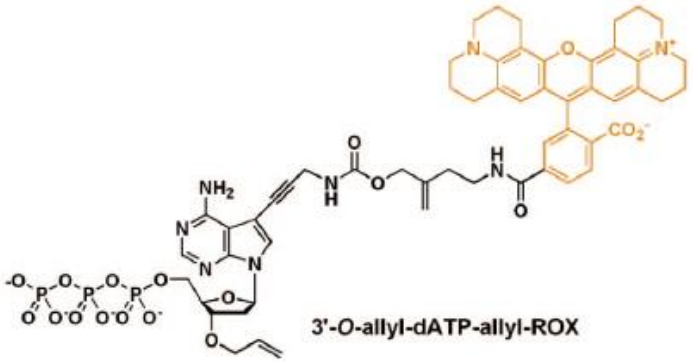


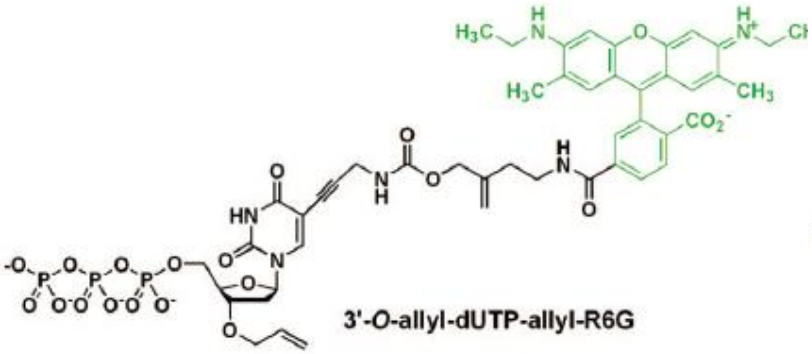
- Sequencing by 4 labeled reversible terminators

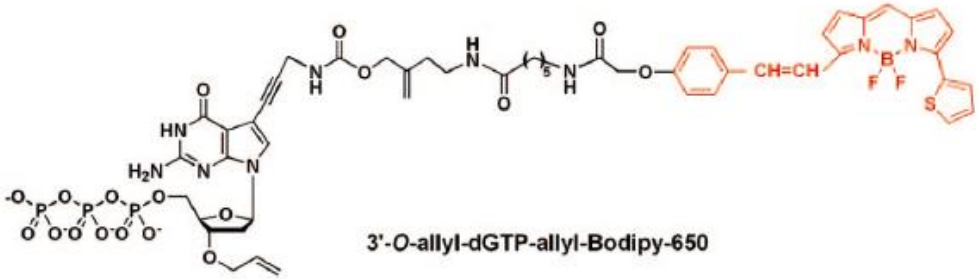
Cleavable Fluorescent Nucleotide Reversible Terminators

Fluorophore and 3’-O-allyl group can be cleaved by Pd-catalyzed deallylation



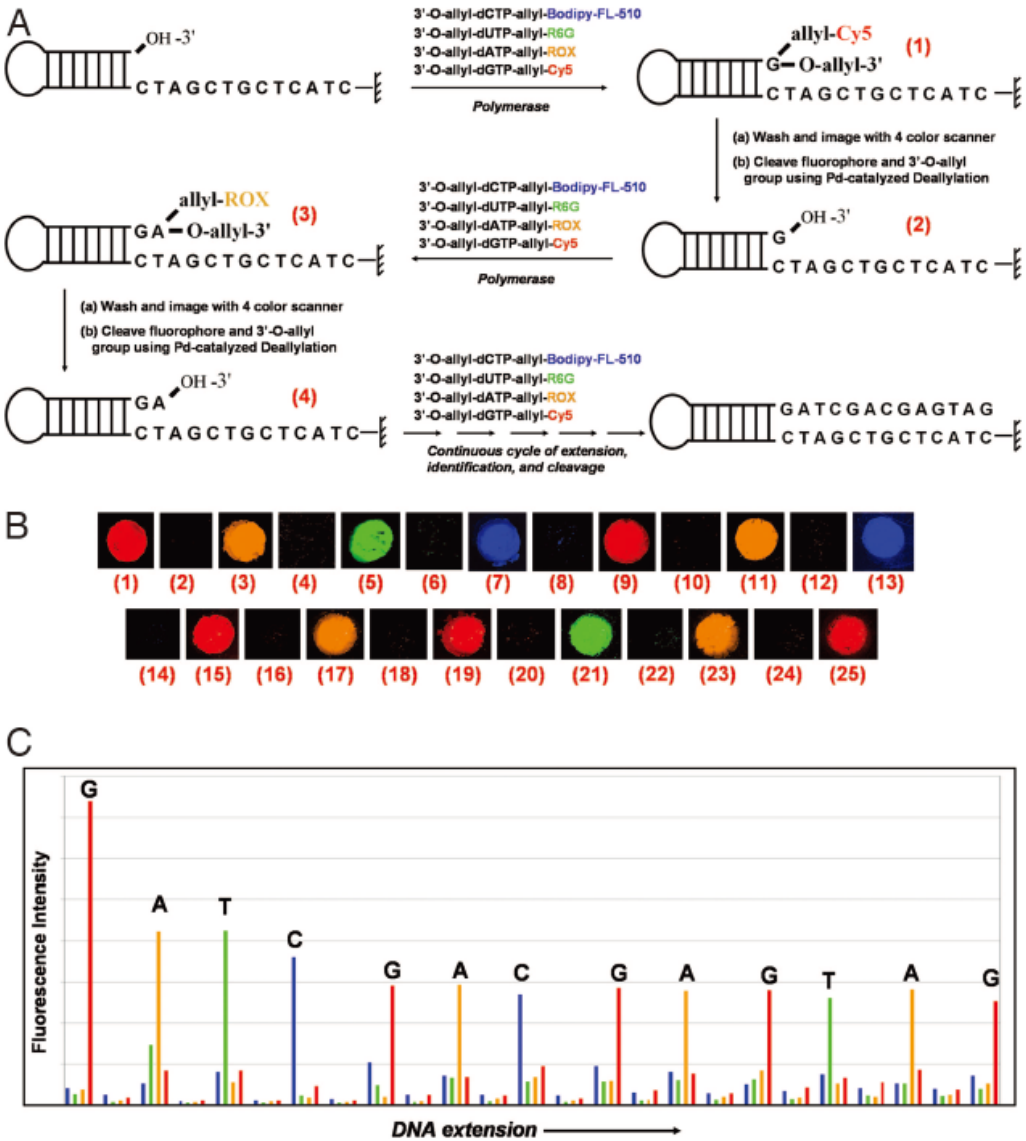






J. Ju et al., 2006, PNAS

Four-Color DNA Sequencing



J. Ju et al., 2006, PNAS

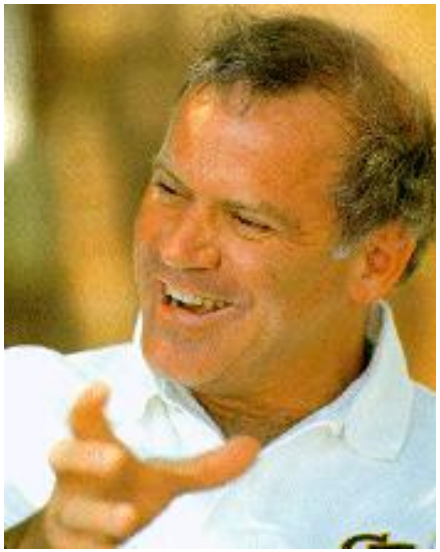
4. PCR

PCR: Polymerase Chain Reaction

Amplification of a specific DNA fragment

Kary Mullis

- Invented PCR in 1983
- Nobel Prize in Chemistry in 1993

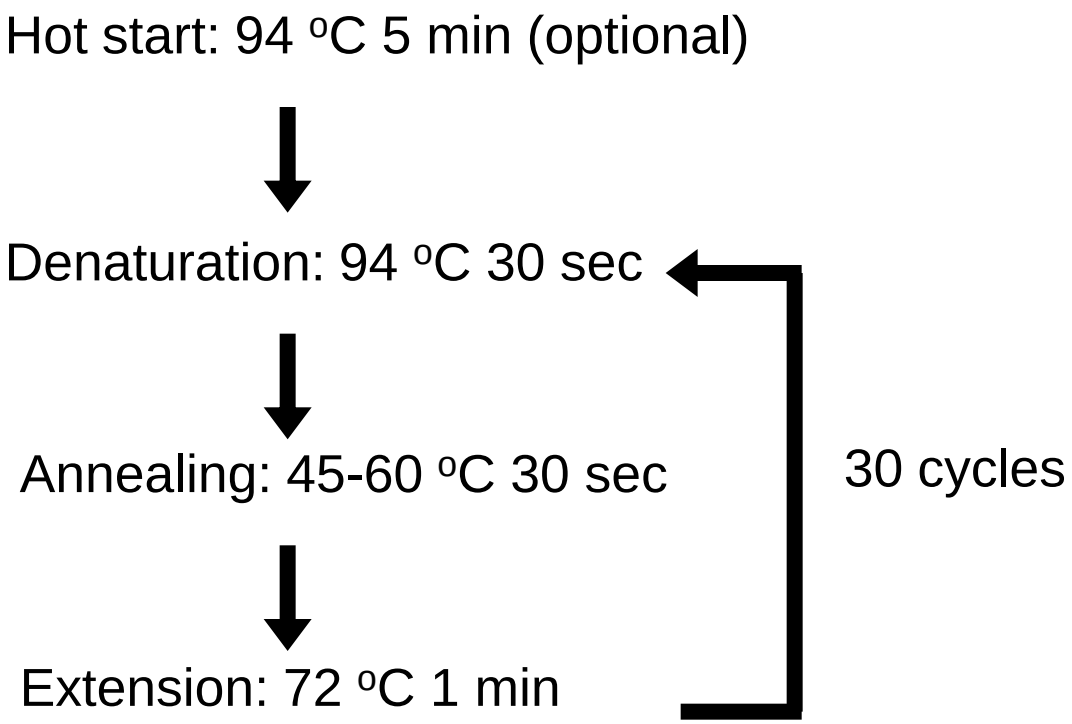


PCR: Basic conditions

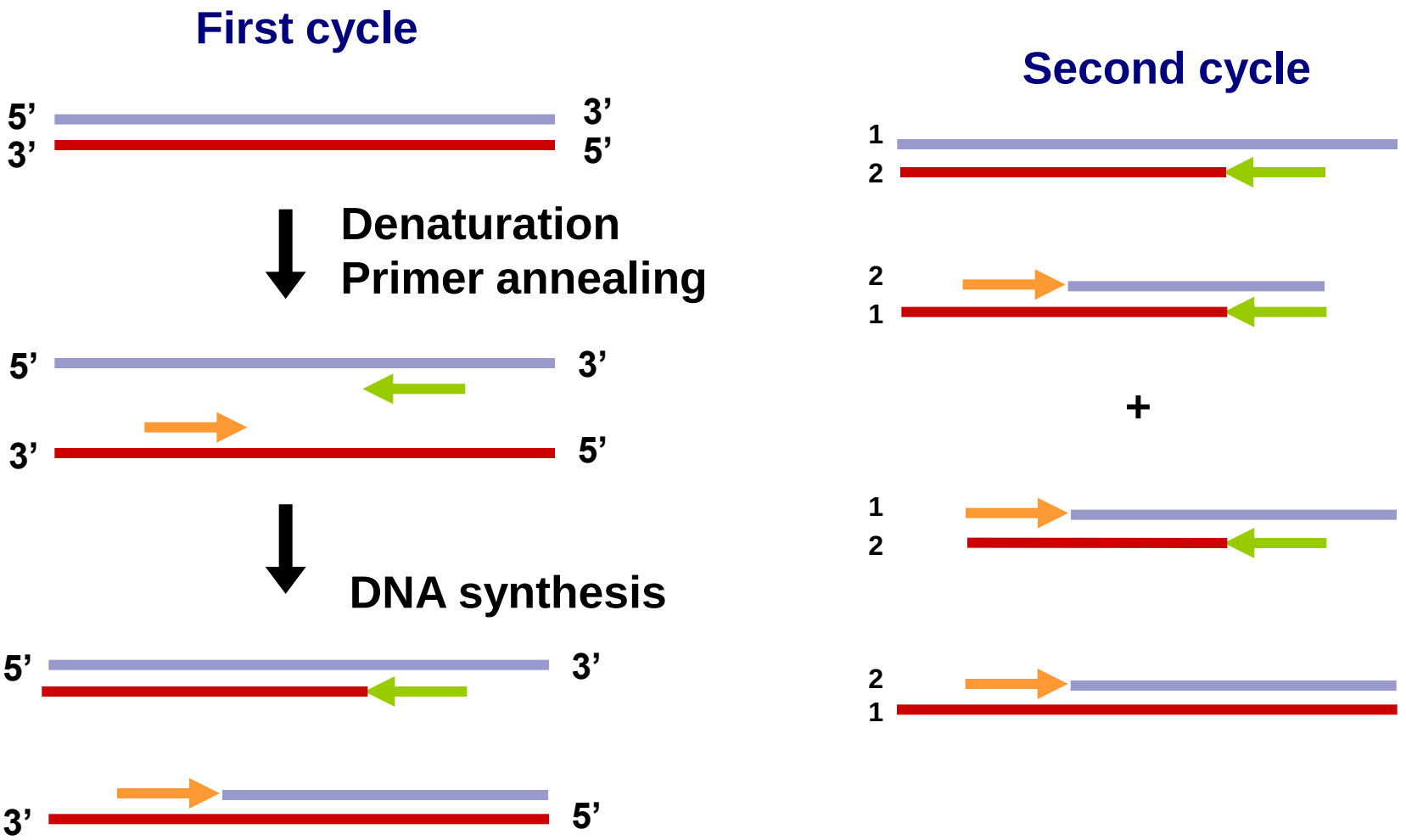
Reaction Mixture

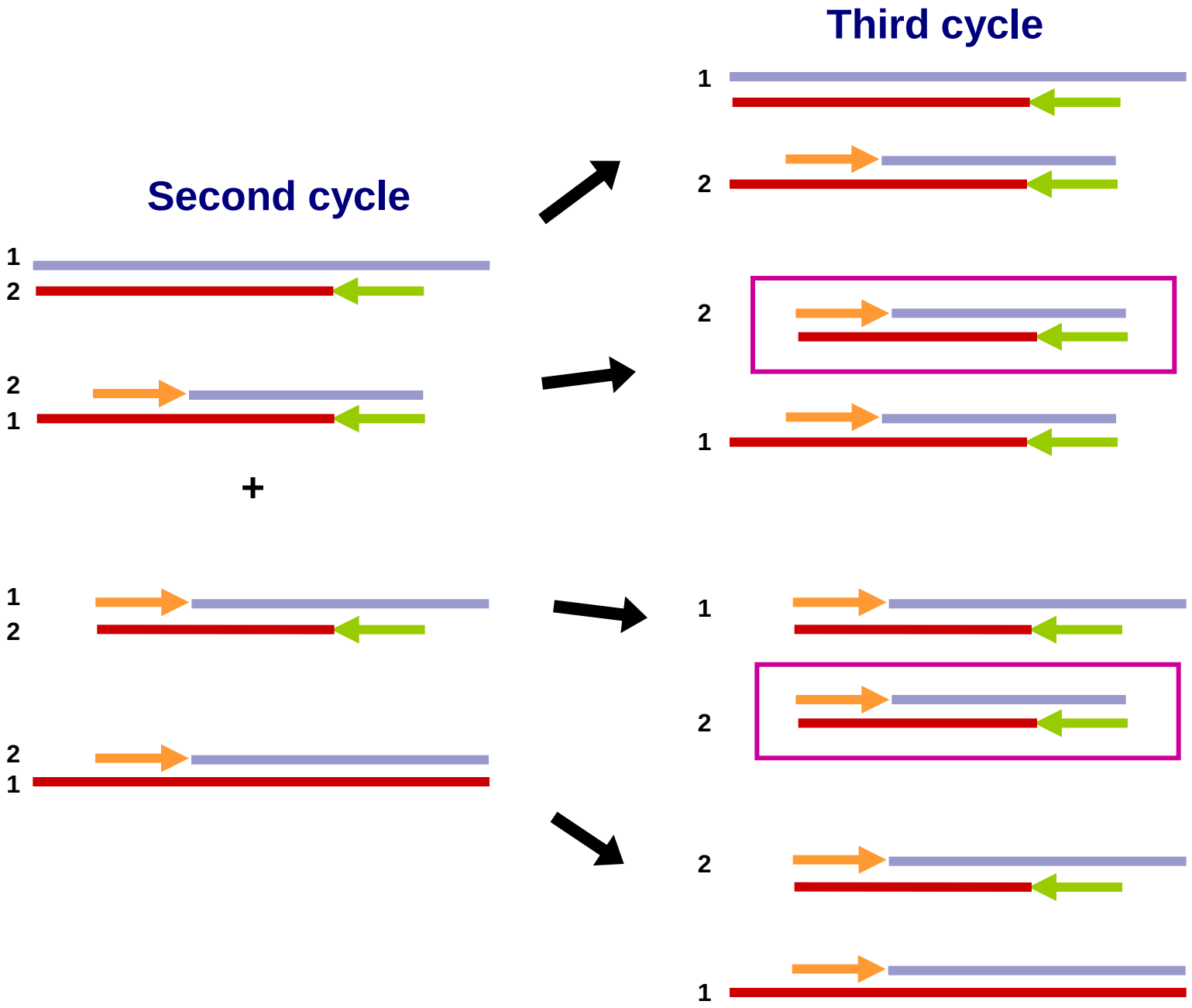
- Template DNA
- Primers
- dNTPs
- DNA polymerase
- Reaction buffer
- $MgCl_2$

Reaction



PCR


First cycle
5’
3’
3’
5’
Denaturation
Primer annealing
5’
3’
3’
5’
DNA synthesis
5’
3’
3’
5’
Second cycle
1
2
2
1
+
1
2
2
1


Second cycle
Third cycle
1
2
2
1
+
1
2
2
1
1
2
2
1
2
1
2
1

PCR: Components

Primers

- 17 to 30 nt
- About 50% GC content
- Avoid long stretch of single nt
- Avoid secondary structure formation
- Avoid complementarity between the two primers

DNA polymerase

Taq (Thermus aquaticus): fast, low fidelity, adds A to 3' end

Pfu (Pyrococcus furiosus): slow, high fidelity (3' to 5' exonuclease activity), blunt end

Gene Synthesis by PCR