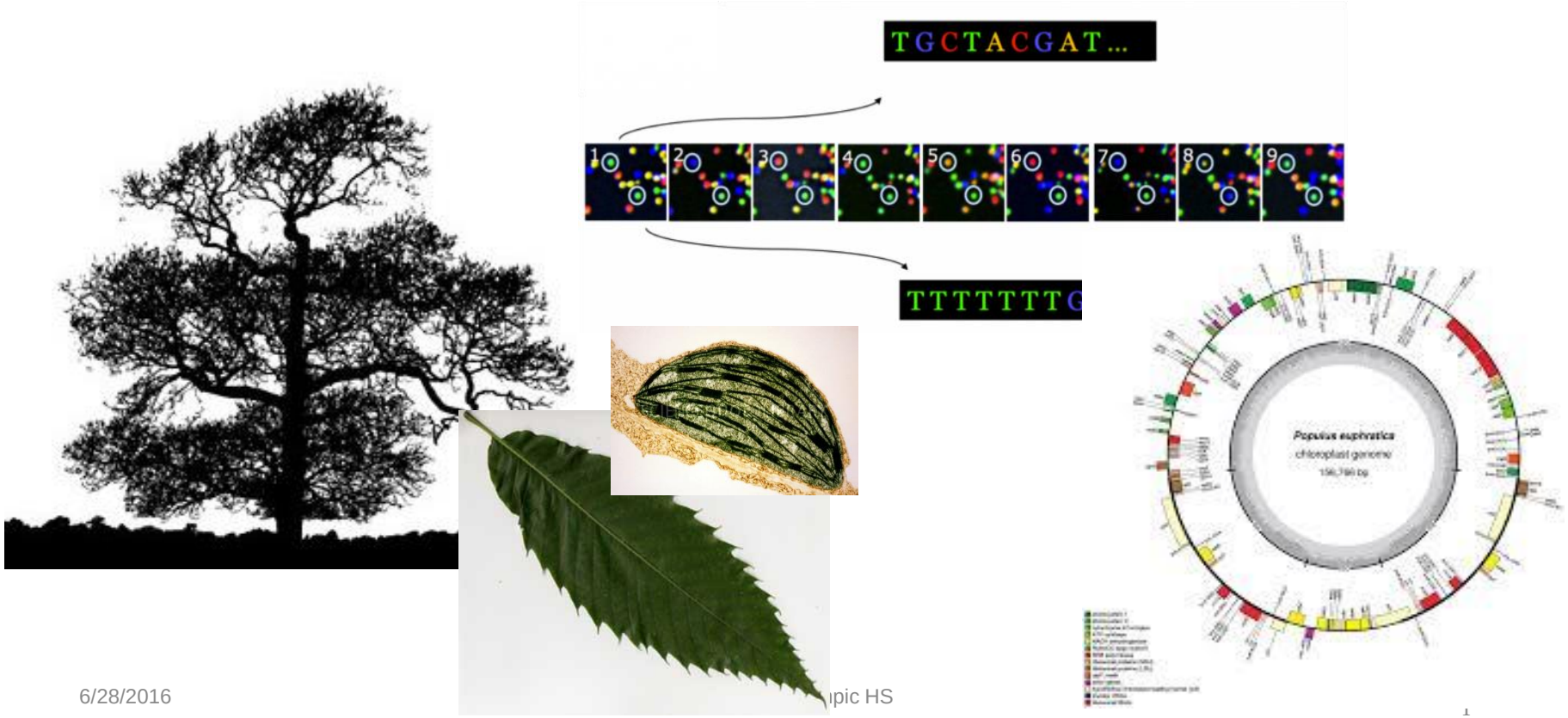
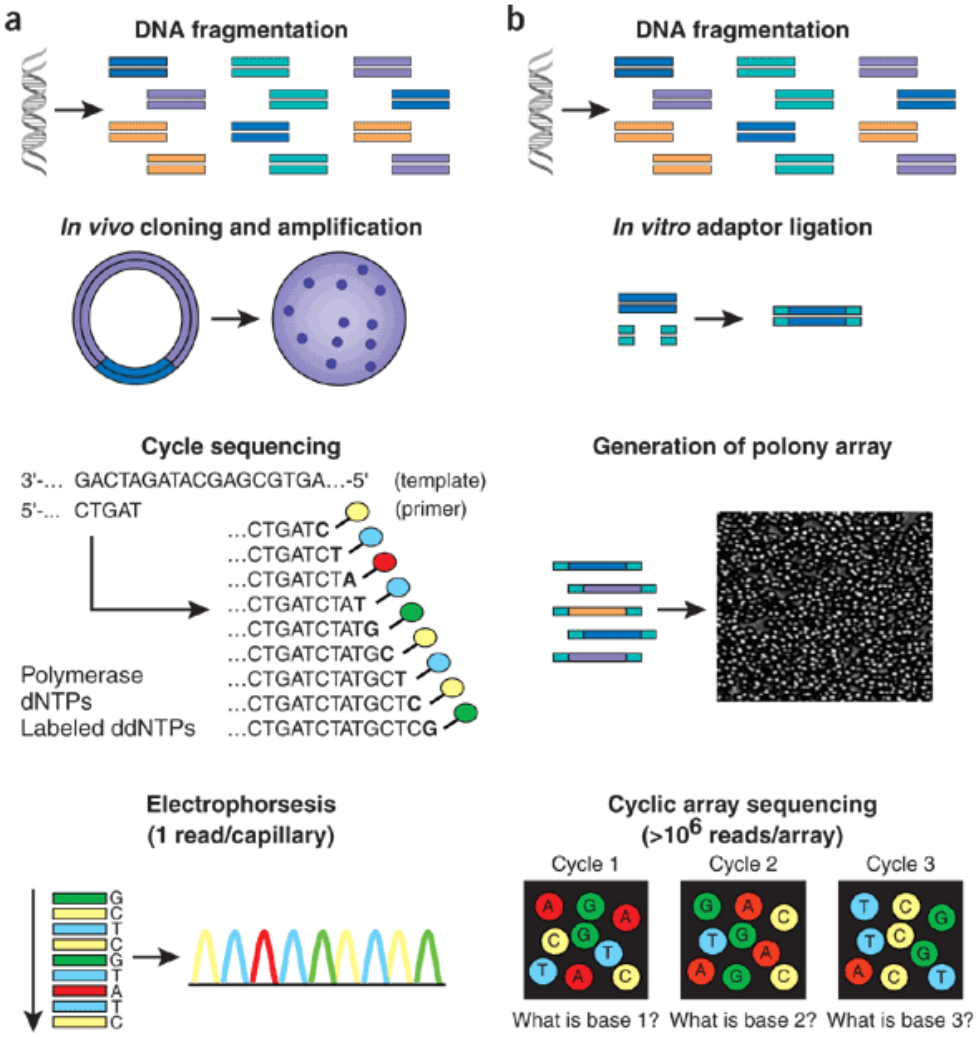
Dr. Jennifer Weller
June 2016
Sequencing Technologies

B3- Summer Science Camp

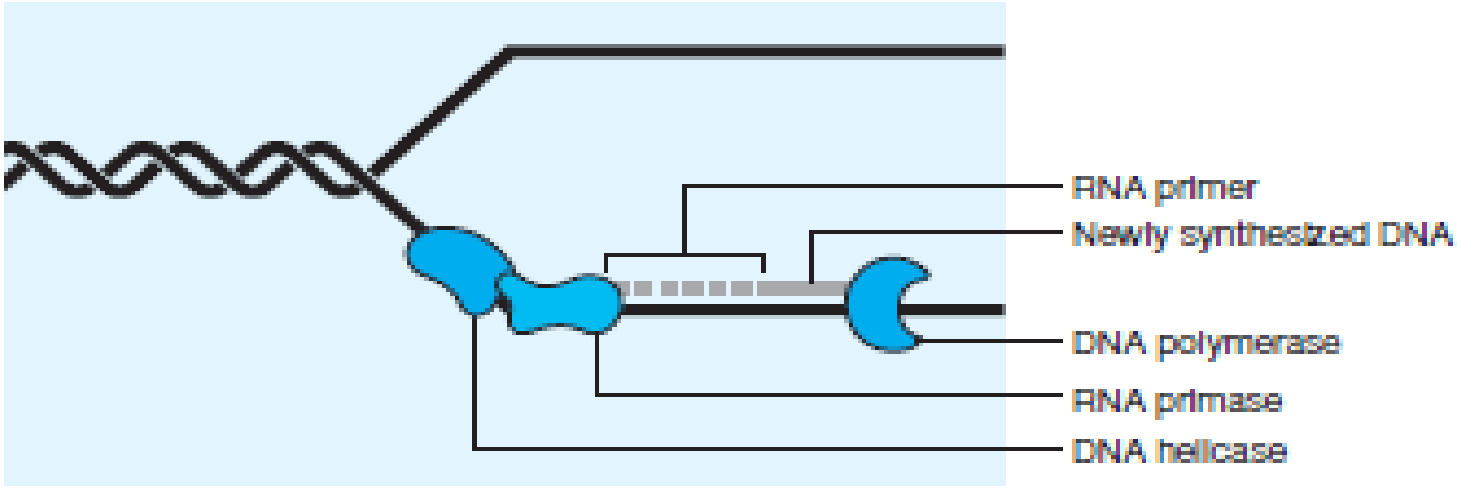


Topics

- Fundamentals of ‘Sequencing by Synthesis’
- Sequencing Technologies or ‘Platforms’
  - Sanger/Capillary sequencing
  - High-throughput or Next-Generation Sequencing Technologies
- Part 2: Signal recognition, errors, base calls and quality scores with Assembly



Synthesis = Organic chemistry OR DNA replication

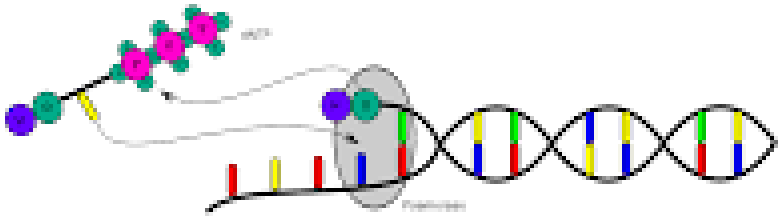


Cells make new copies of DNA using replication proteins

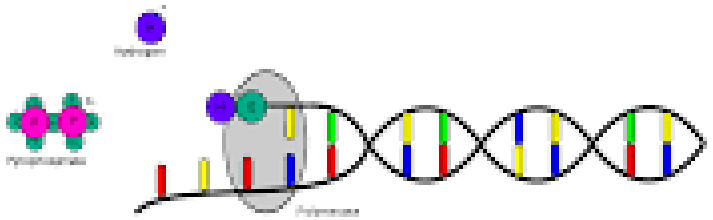
- 1 short bit of complementary primer (short piece of nucleic acid, usually RNA)
- Building blocks = deoxyribonucleotide subunits (dNTPs.)
- 3 enzymes: DNA helicase, DNA polymerase, RNA primase
  - DNA helicase unwinds the 2 strands – they become templates.
  - RNA primase introduces the primer at the junction
  - DNA polymerase then binds to primer-template and starts adding nucleotides to the 3′-hydroxyl end of the primer so the new strand grows in the 5' → 3' direction.

In vitro biosynthesis

- How can you separate the strands?
  - The H-bonds can be broken using heat or base while the covalent backbone stays intact.
- What can you do to provide the primer?
  - Organic chemists learned how to synthesize dNTPs and use them to make short pieces with a known sequence
- What can you do to provide the polymerase?
  - Purify the protein from a source that has a lot, like E coli
  - Clone a gene, put it into E coli or yeast, purify the protein



Polymerase integrates a nucleotide.

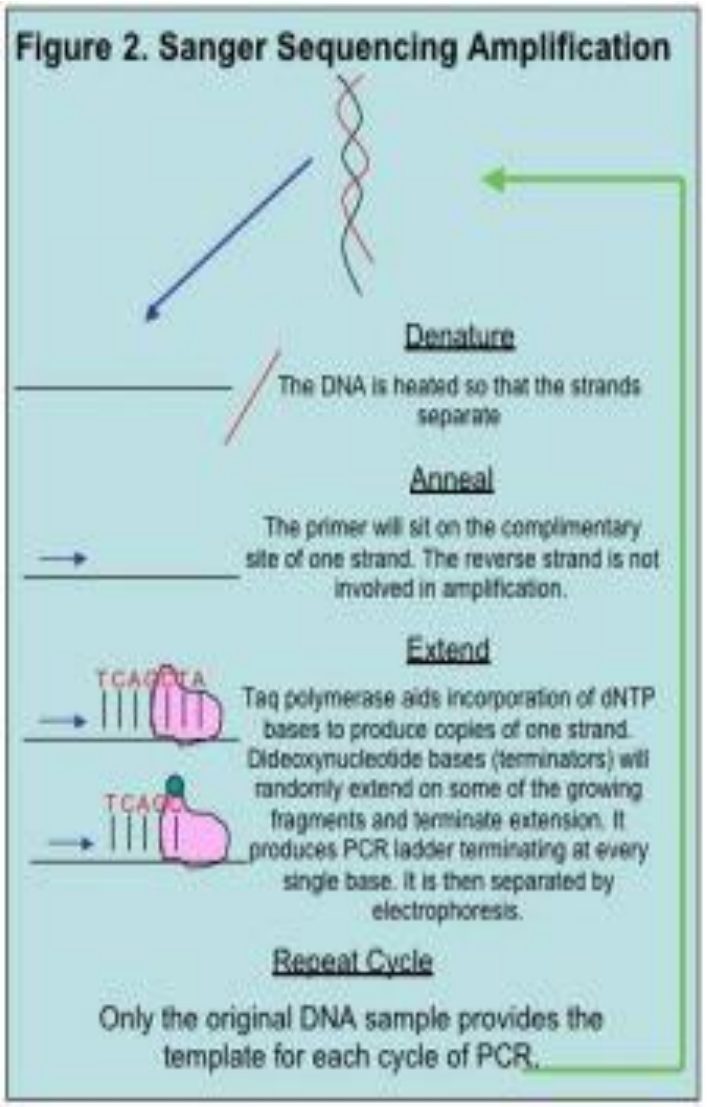


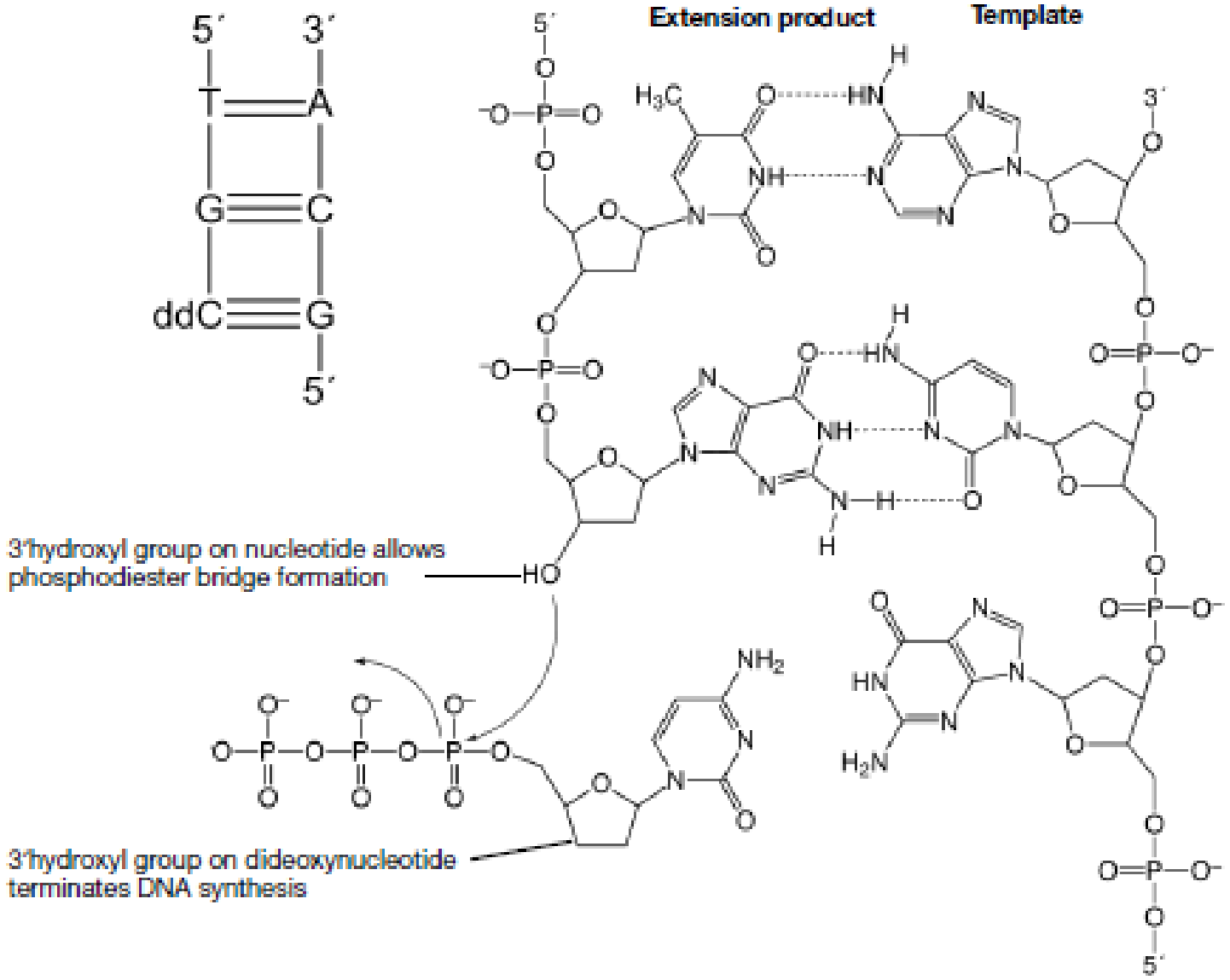
Hydrogen and pyrophosphate are released.

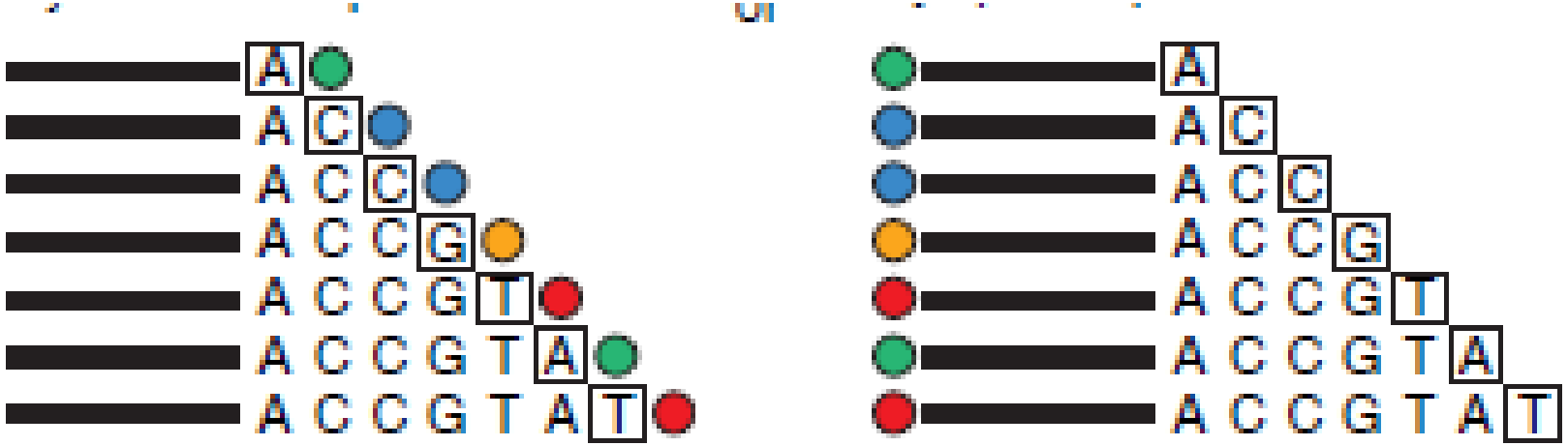
The polymerase moves REALLY fast.

Sequencing by Synthesis – Generation Zero

- Clever Chemist Wins Nobel Prize: Fred Sanger
- We need to
  - Slow the polymerase down
  - Include something you can detect.
- Strategy: include some synthetic modified nucleotides that have a defect: no more nucleotides can be added = stop means you have slowed down a lot).
- Label the phosphate group with radioactive P^{32}, which will cause film to darken.



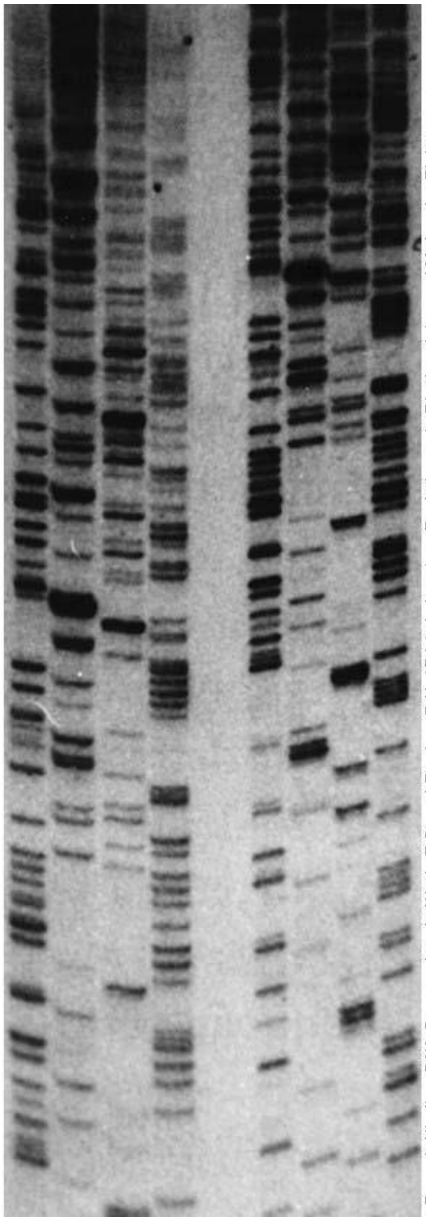

5′
3′
T
A
G
C
ddC
G
5′
Extension product
Template
3′hydroxyl group on nucleotide allows phosphodiester bridge formation
3′hydroxyl group on dideoxynucleotide terminates DNA synthesis



- Add a ratio of about 1 ddNTP* per 10 dNTP so most of the DNA strands keep growing. Run 4 reactions side by side.
  - The end product is a one-sided nested set of complementary fragments.
- Separate the products by length using gel electrophoresis.
- Put a piece of film on top of the gel to expose it.

Fragment separation

- Shortest fragments are at the bottom, closest to the primer
- The order is preserved, but complementary to the template fragment – you are copying it using sequencing by synthesis.
- Bands get closer together towards the top of the gel – this is very hard to ‘read’.



From J. Hindley. DNA Sequencing, in Work, T.S. and Burdon, R.H. (Eds), Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 10, p. 82, Elsevier (1983). Used by permission.

Gen 1: better chemistry, auto-detection and software!

- It would be nice to get rid of the P^{32}. What else do we have sensitive detectors for?
- It would be nice to have automatic, real-time signal detection.

- This is Generation 1 in Sequencing by Synthesis
  - Use fluorescent dyes and a fluorescent signal detector built into the electrophoresis platform.
  - Photon emission data is stored as a function of time and wavelength
  - Software processes the signal and interprets it.

Fluorescent Dyes and Data Output

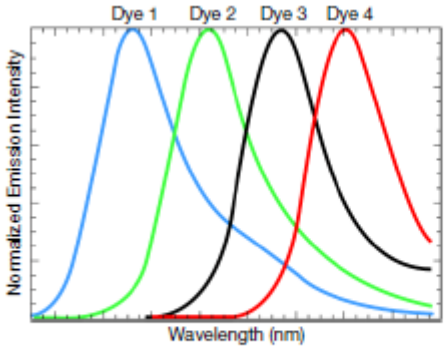
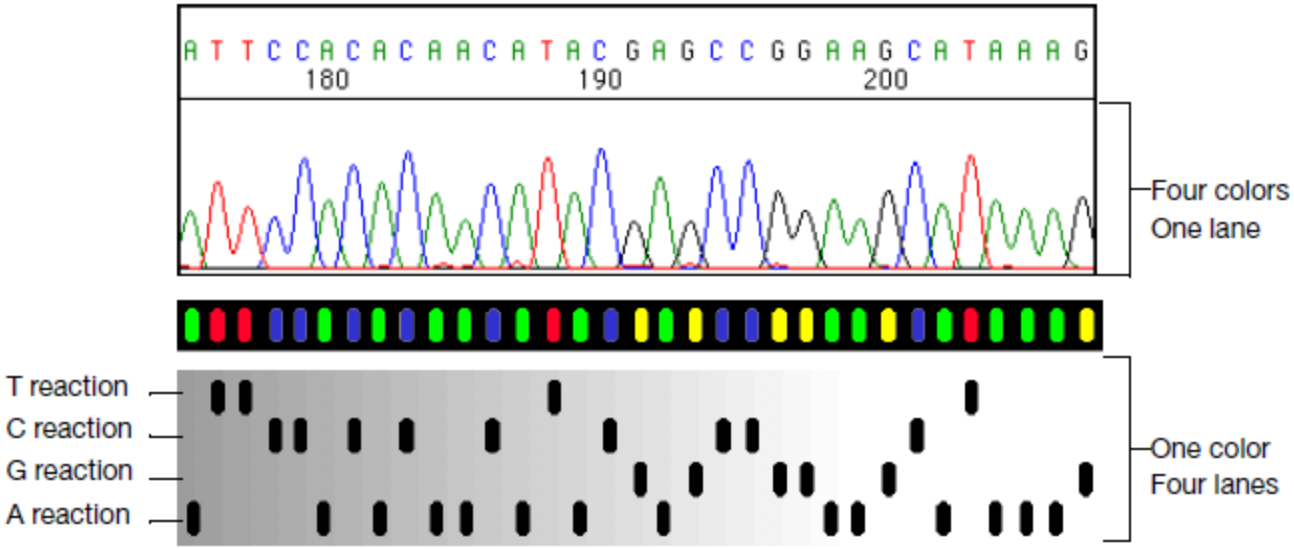


Figure 7 Emission spectra of the four BigDye dyes, where Dye 1 = Big-d110, Dye 2 = R6G, Dye 3 = Big-dTAMRA, and Dye 4 = Big-dROX



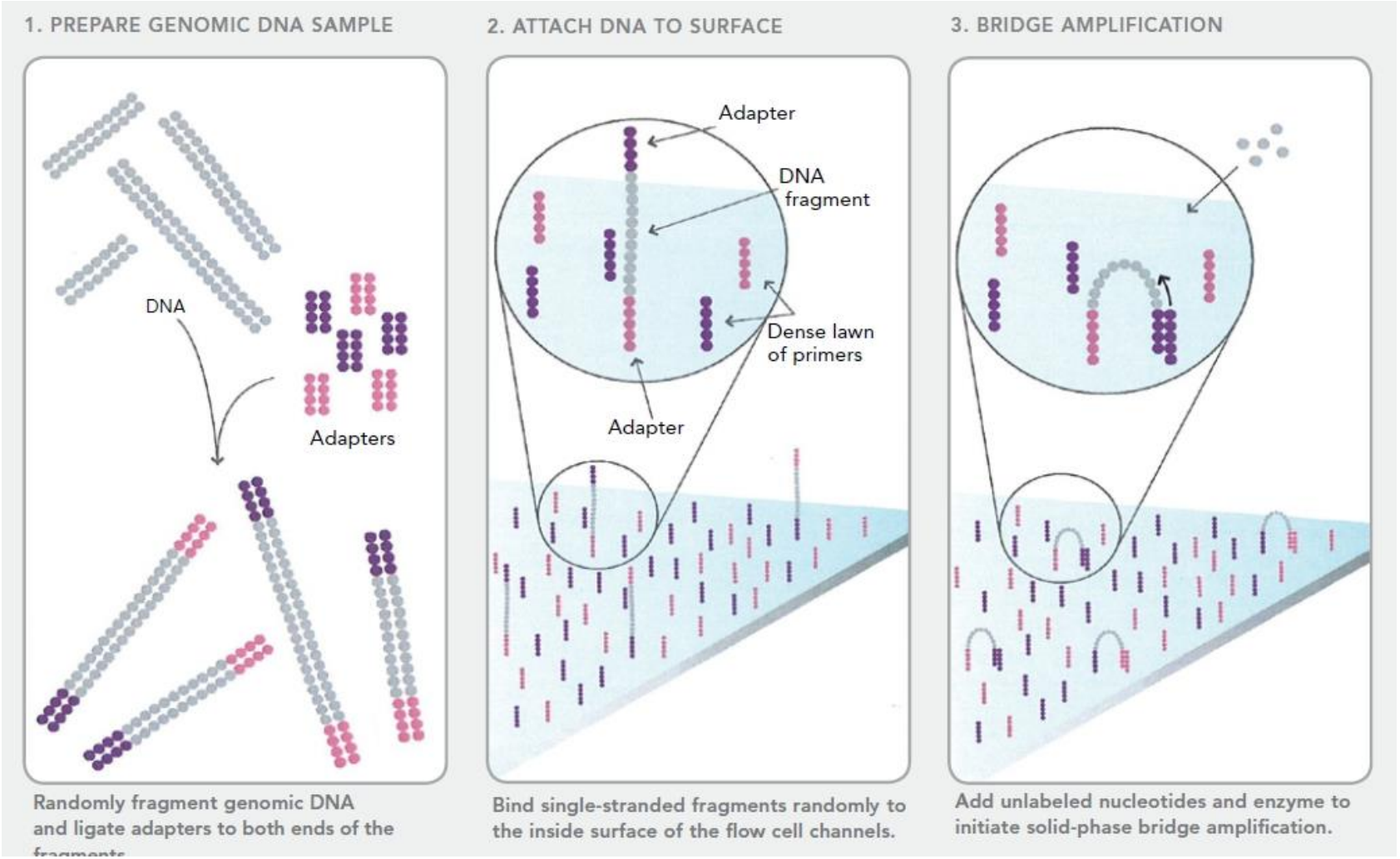
Sequencing by Synthesis Gen2 – High Throughput or Next-Generation Sequencing

- Wish-list: no more electrophoresis and much higher numbers of sequences per run of the instrument.

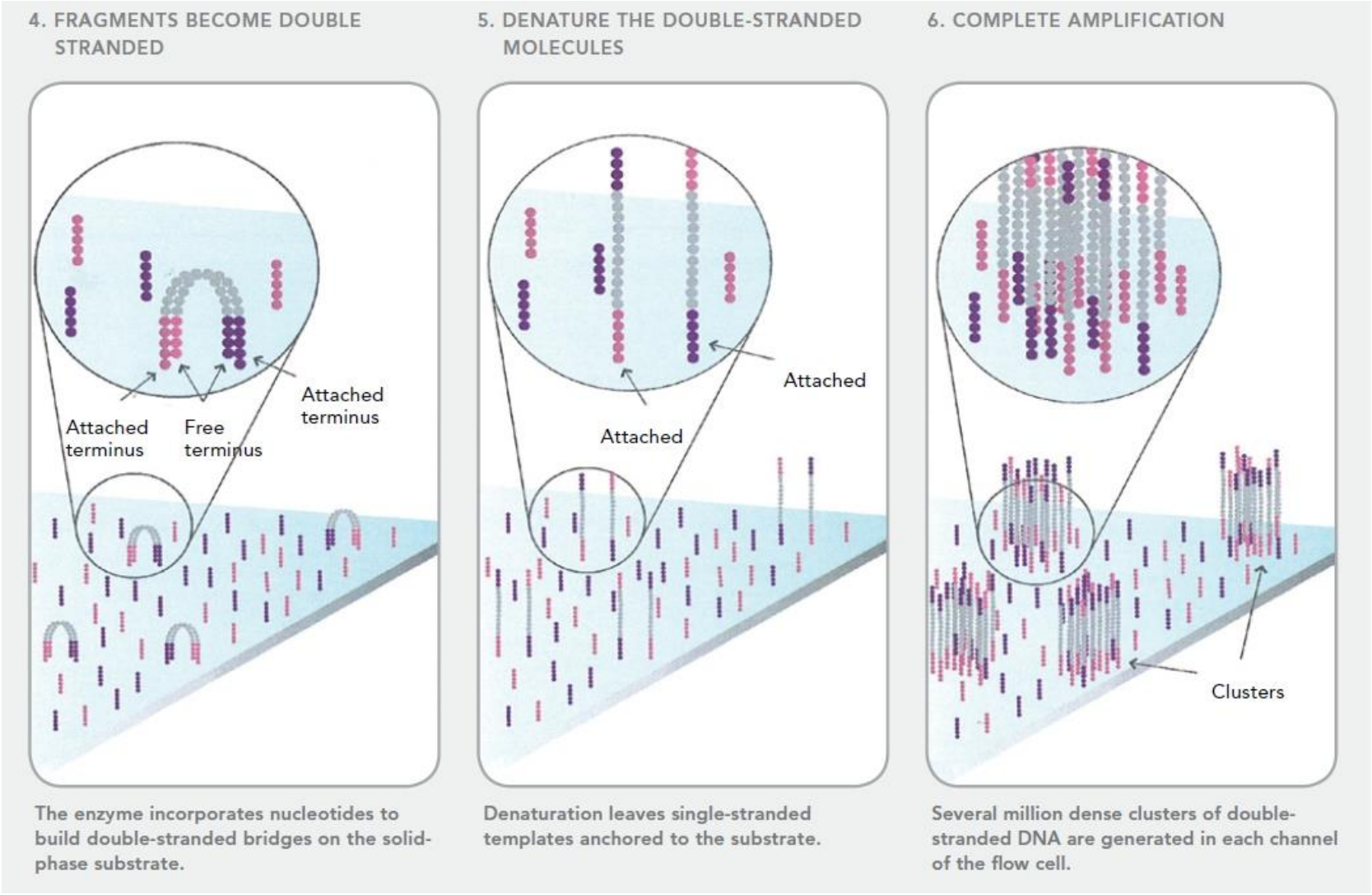
- Illumina strategy – use fluorescent-dye labeled dNTPs, block extension in a reversible way and use much better optical systems

- To get the more reads: use a flat clear surface (flow cell) with short oligonucleotides covalently attached to it at an even density.
- The change in chemistry: each dNTP has its own color, but the block at the 3' –OH can be removed. After each polymerase step the extra dNTPs are washed away, an image is captured, then the block is removed so each strand can be extended again.
- Signal detection: A flow cell is imaged (4 times, once to capture each color) after each flow of dNTPs
  - The flow cell is imaged in 100 tiles each time in 4 colors.
  - 37 cycles gives 118, 400 images
  - Resolution: one pixel is $0.14um^2$ with 9 pixels per cluster

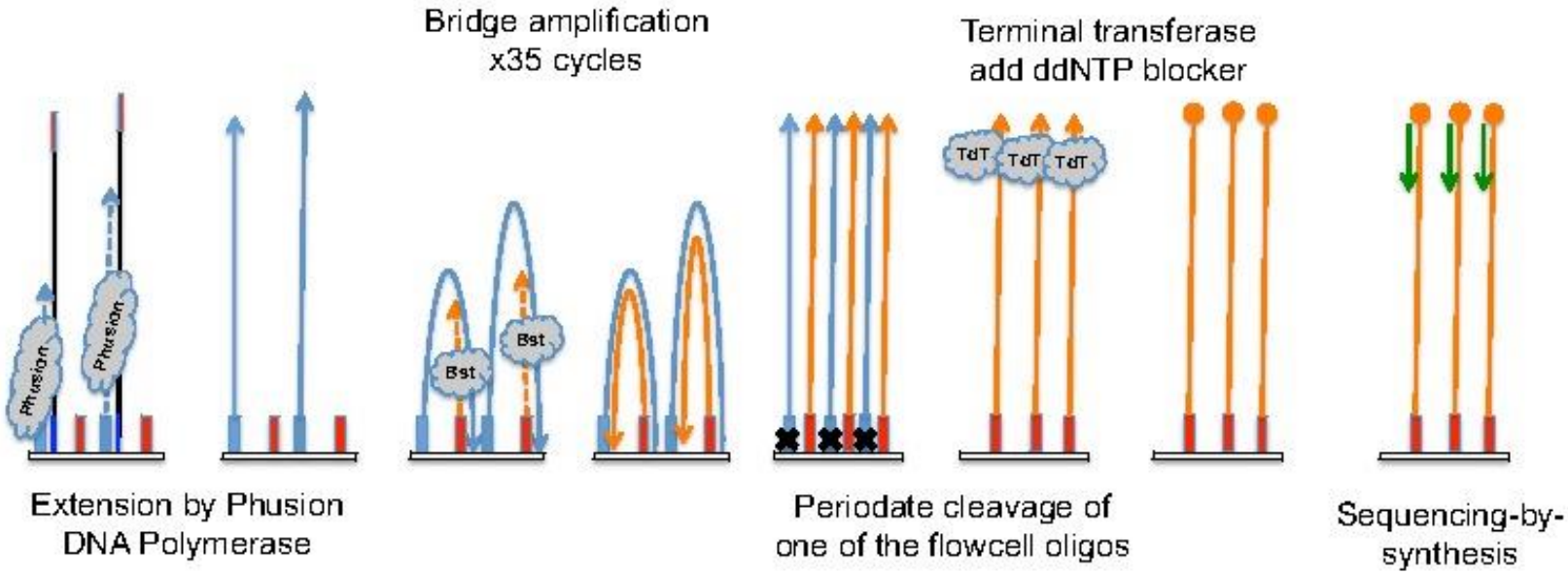
Illumina - MiSeq and HiSeq



PCR colony = Polony creation



Bridge amplification and Sequencing-by-synthesis



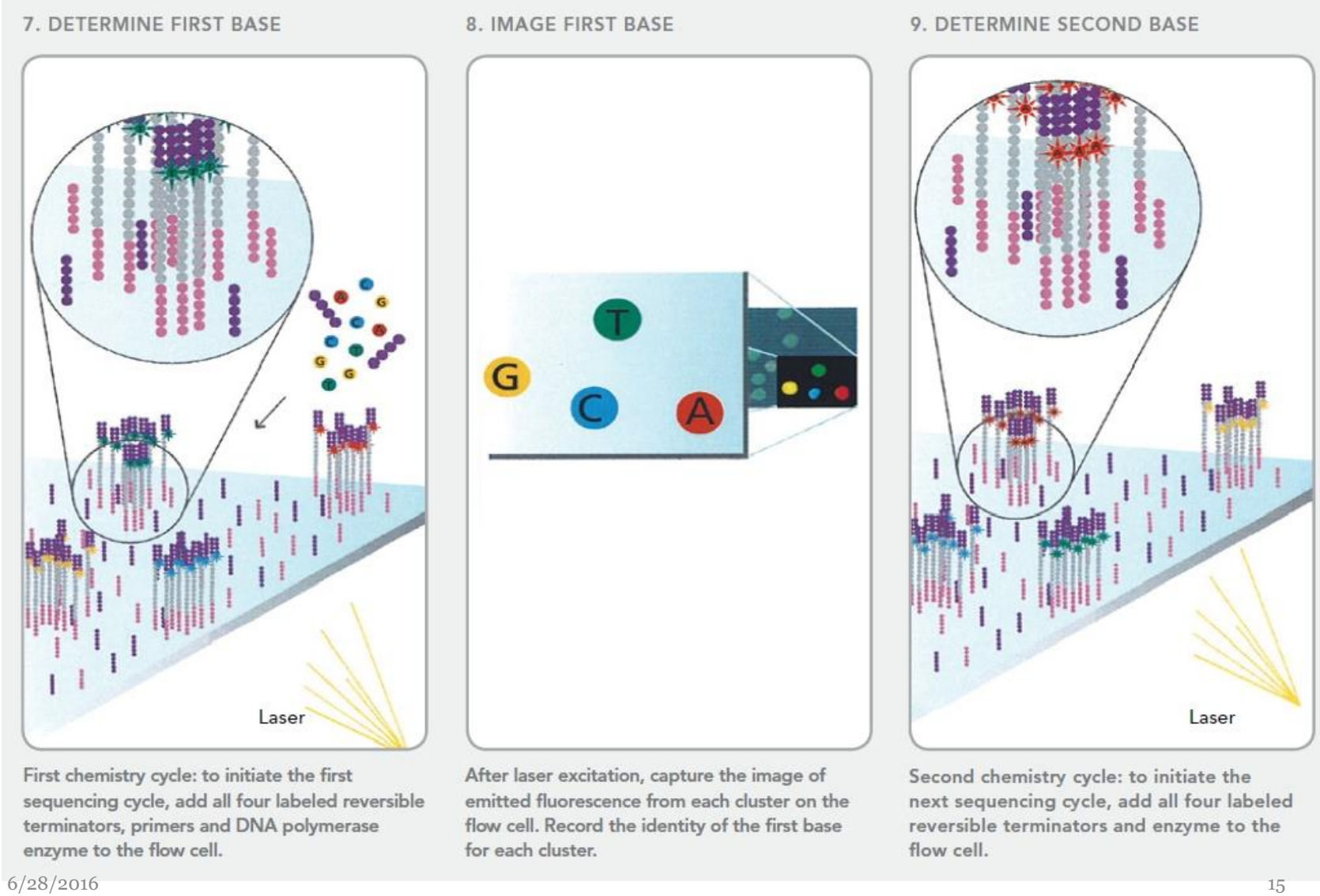
Single-stranded DNA anneals to one of the flowcell oligos (primer) → synthesize copy.

The new strand flops over and finds the complement on its other end on the surface (primer) → synthesize copy.

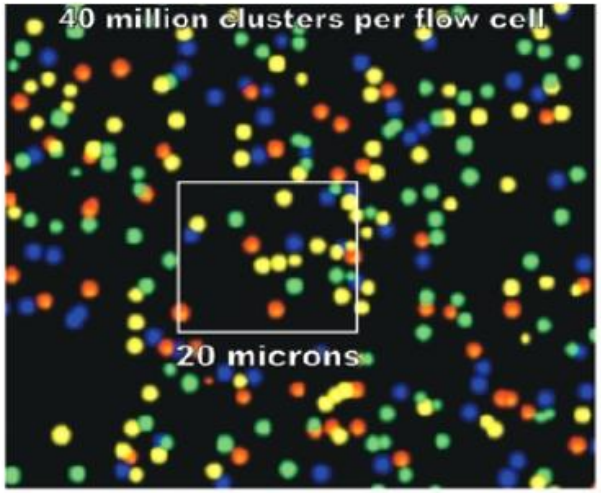
New strands are very close together (the length of the DNA template) → colony

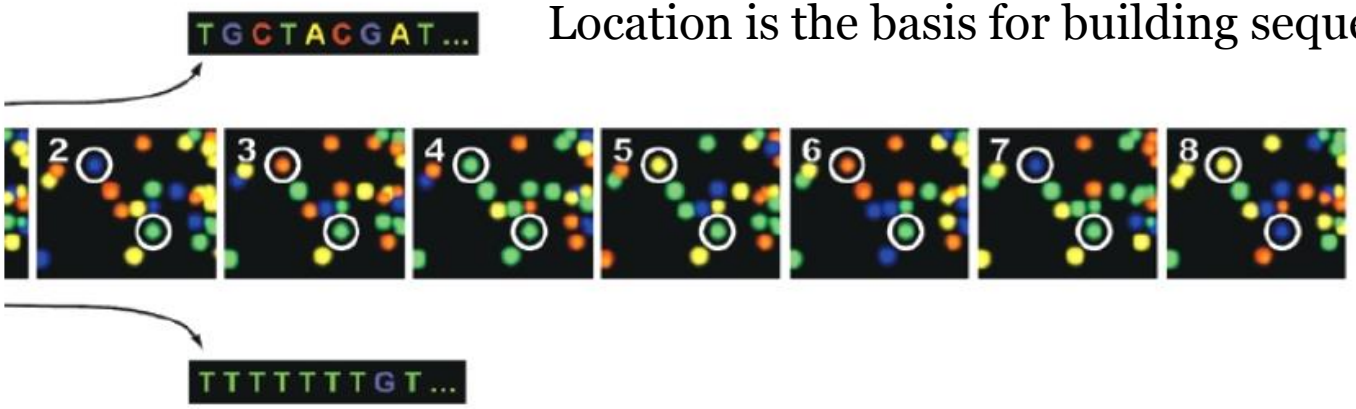
Add sequencing primer and dye-labeled 3' blocked dNTPS.

Reagent Cycling and Data Capture



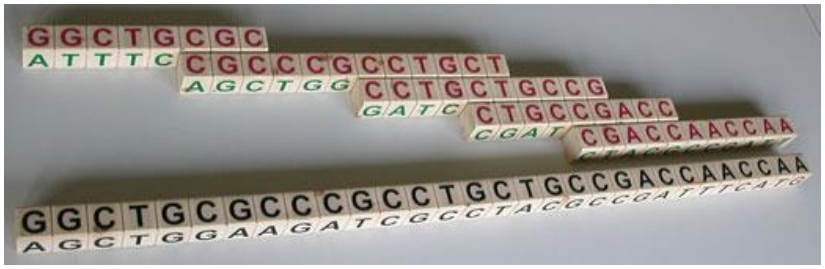
Examples of images





Location is the basis for building sequence

Assembly Rules



- How much overlap (the k-mer where k is the overlap length) should you require?
  - Are there ways you could estimate this length for a known genome?
- Should the quality score matter?
- Should you trim the sequence?
- Should you filter out sequences with low-quality bases in the middle?
- Is it useful to keep track of how many times a nucleotide has been incorporated (through individual reads) in the longer contig?