

# NEXT GENERATION SEQUENCING TO ASSESS BACTERIAL DIVERSITY



## Summary

# Extracting DNA



# Remember what happens in PCR?

## Polymerase chain reaction - PCR

original DNA  
to be replicated

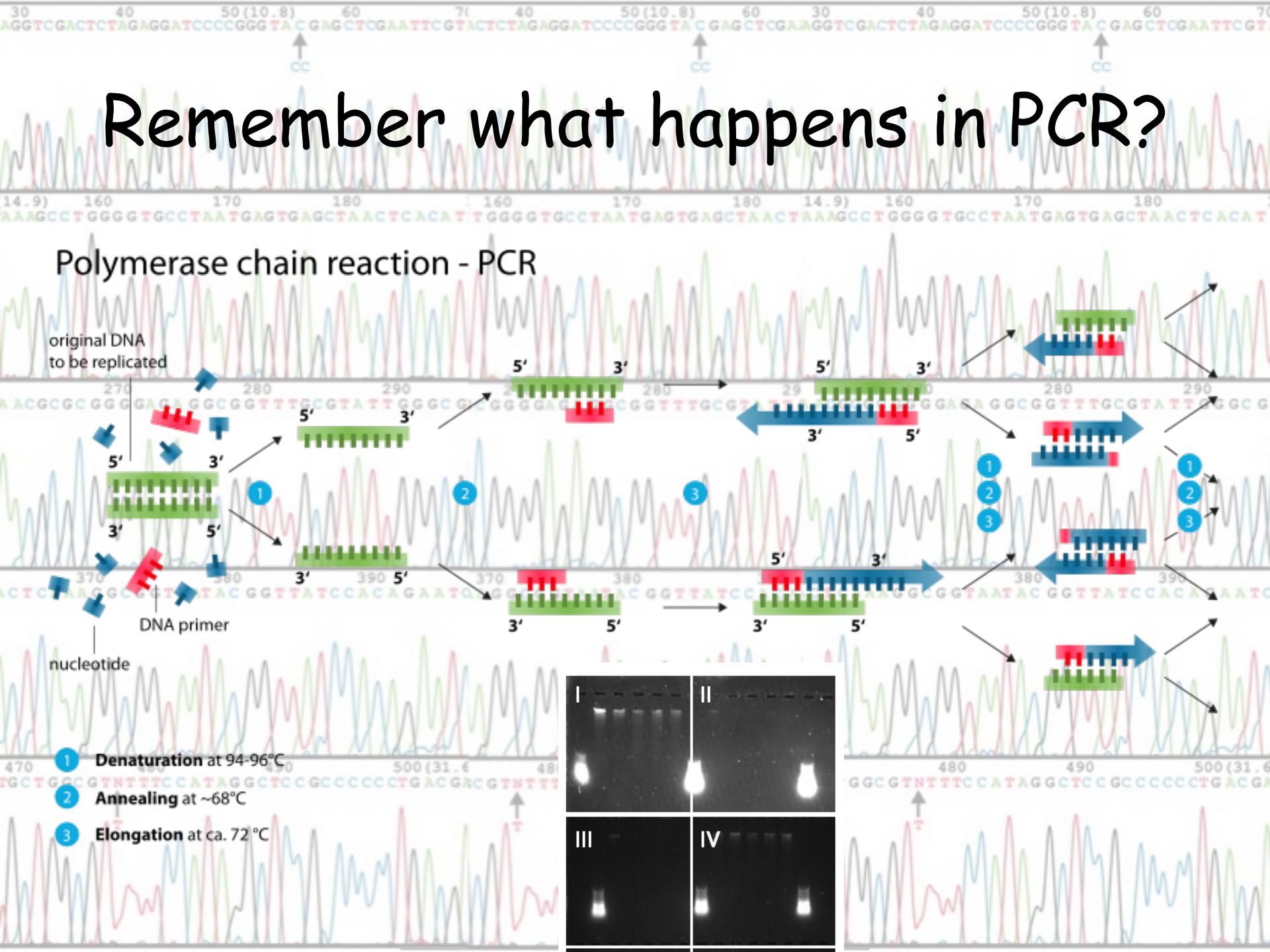
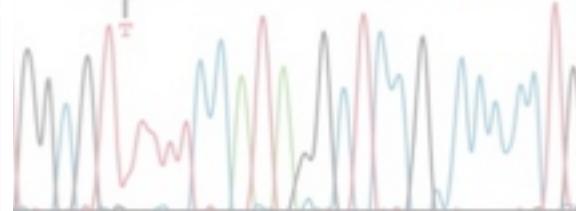
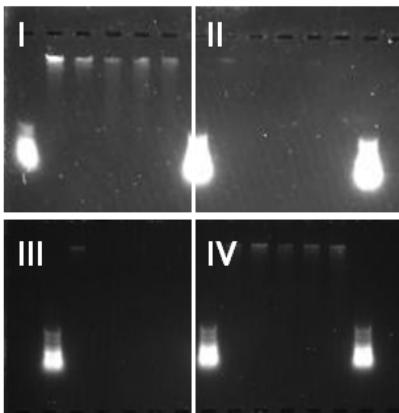
270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 (31.6)

5' 3'

DNA primer

nucleotide

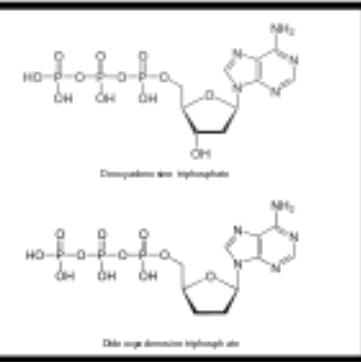
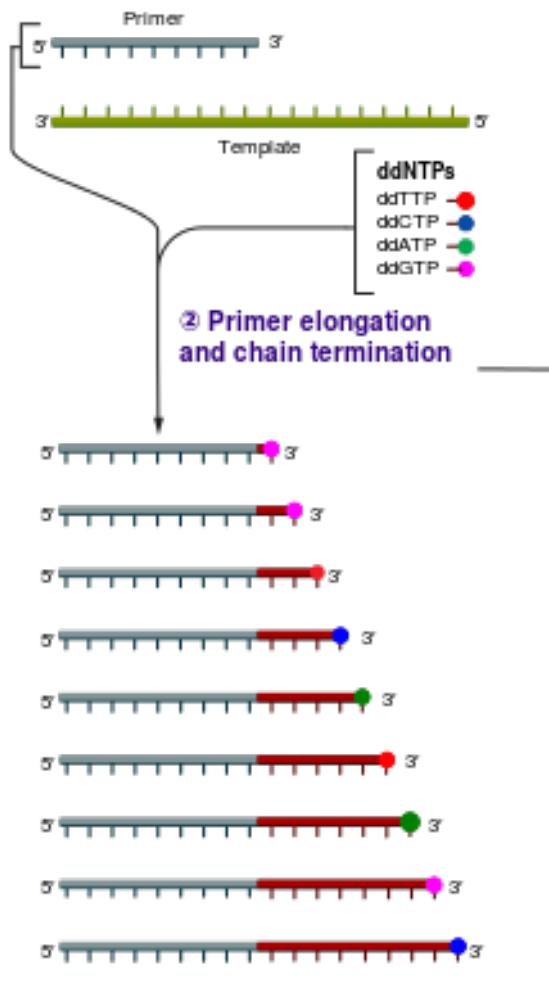
- 1 Denaturation at 94-96°C
- 2 Annealing at ~68°C
- 3 Elongation at ca. 72 °C



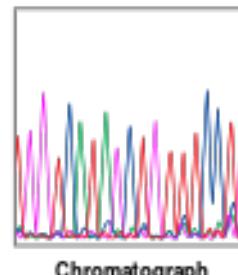
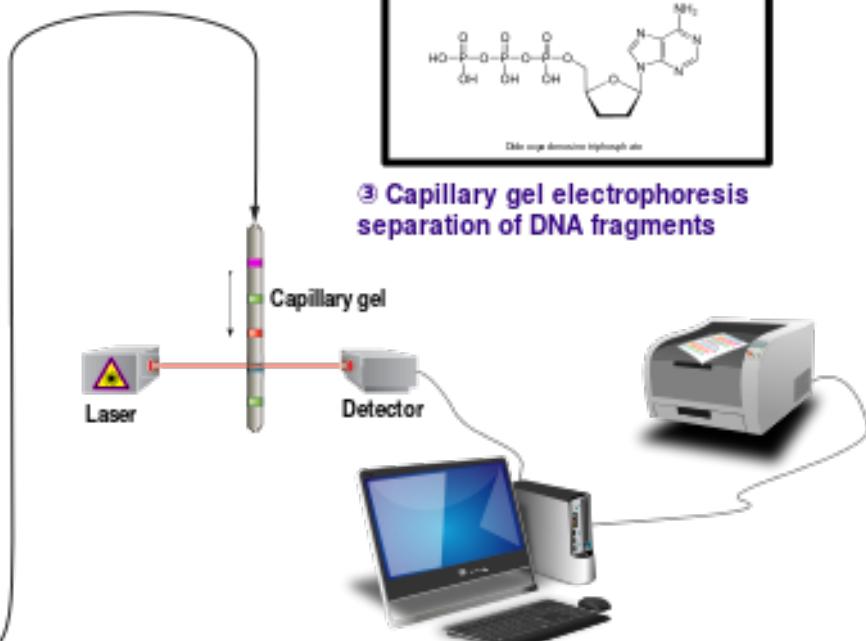
# Sanger Sequencing

## ① Reaction mixture

- Primer and DNA template → DNA polymerase
- ddNTPs with flourochromes → dNTPs (dATP, dCTP, dGTP, and dTTP)



## ③ Capillary gel electrophoresis separation of DNA fragments



## ④ Laser detection of flourochromes and computational sequence analysis

# Why do we use NGS?

Community Amplicon sequencing

Metagenomes

Whole genome sequencing

Transcriptome sequencing

# Community Amplicon Sequencing

Which amplicon did you use?

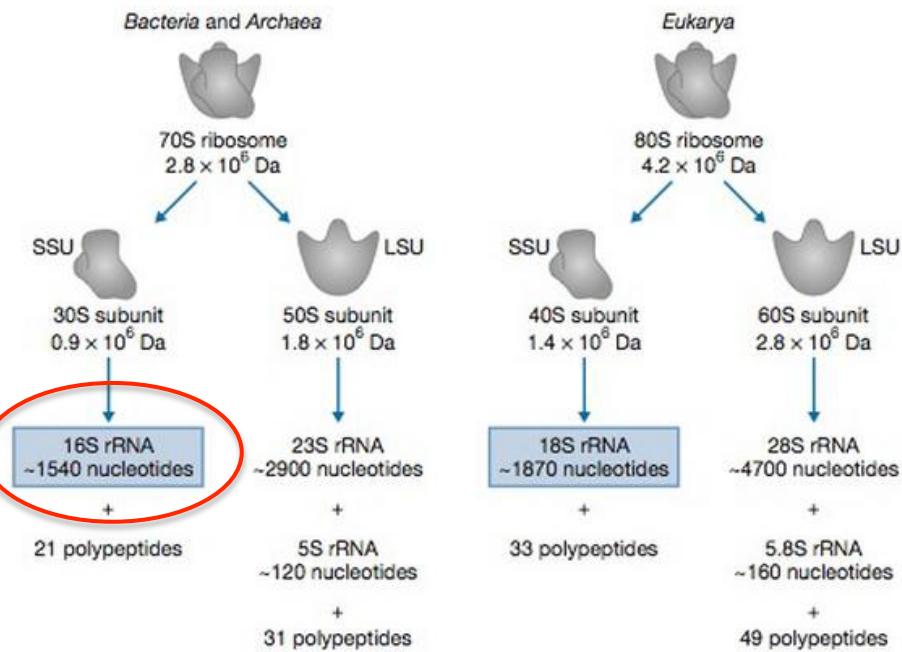
Sequencing of the 16S ribosomal RNA gene which encodes a small subunit RNA

Direct sequencing of DNA amplified by PCR

Similarity of sequences compared in free databases

Why use rRNA genes?

- All organisms contain ribosomes
- Some parts of the rRNA genes are highly conserved, others change more quickly (hypervariable regions)



# NGS Platforms

Ion Torrent PGM



454 Life Sciences

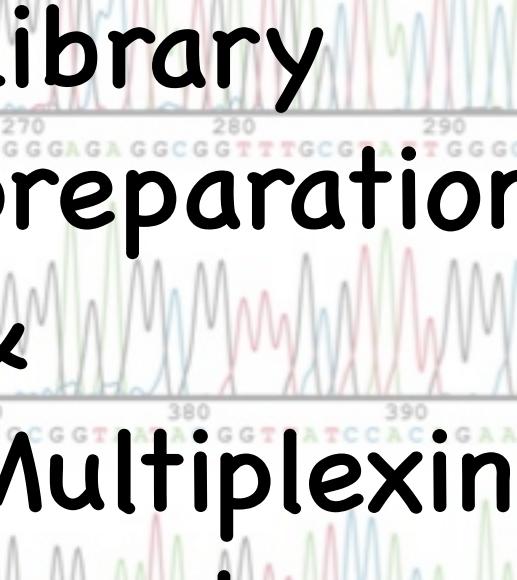
Illuminaria MiSeq

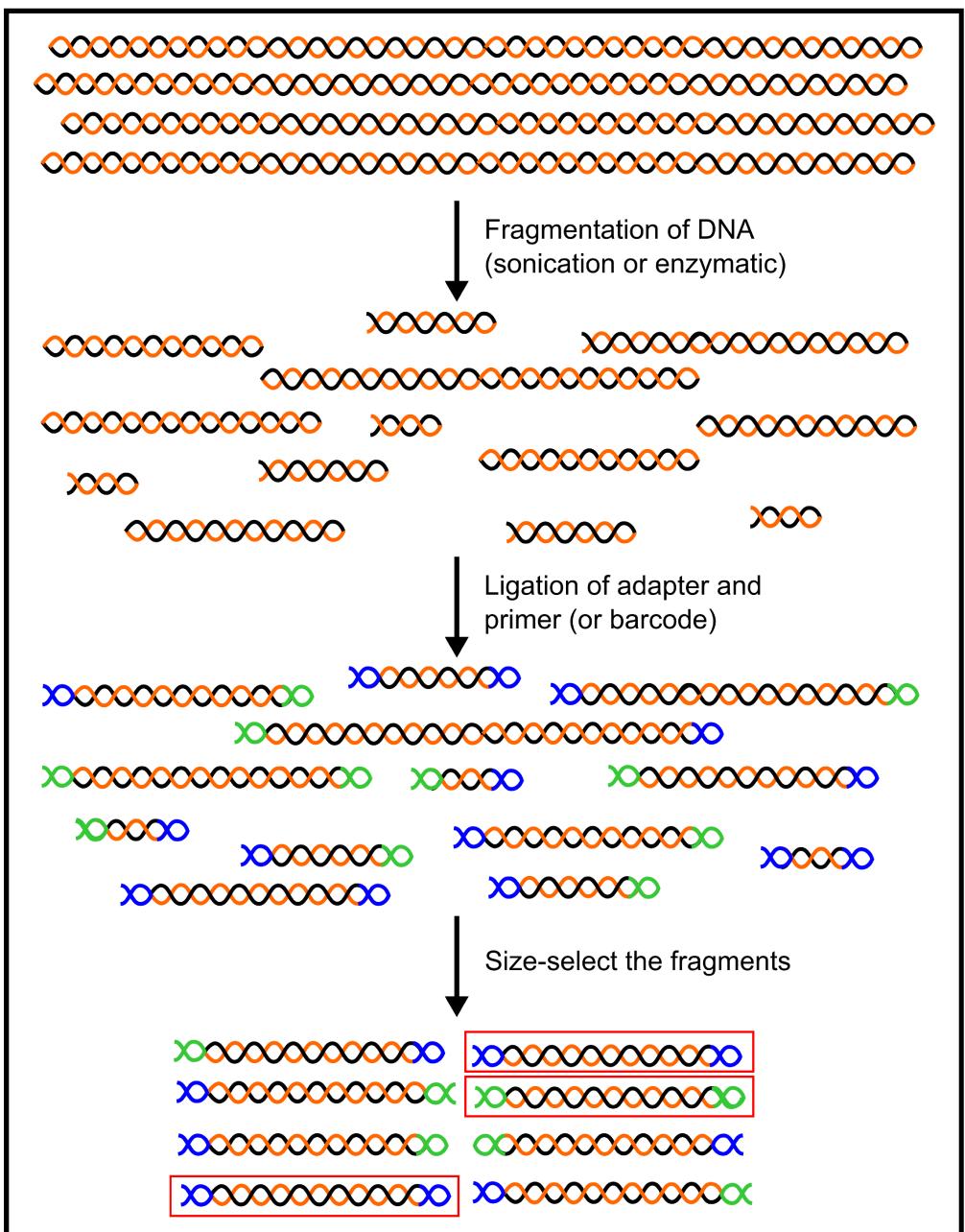


# Which one?

Platform	Capacity	Speed	Read Length	Homopolymers	Cost/ run
454 Roche	450 - 700 Mb	10-23 hours	400-700bp	-	5000 euros
Illumina	15 Gb - 850 MB	5.5 - 65 hours	100-250bp	+	15000 euros
Ion Torrent	1.2-2 Gb	4-5 hours	200-400bp	-	1500 euros

# Workflow: Library preparation & Multiplexing samples





For 16S we multiplexed during PCR..

Illumina 5' Adapter + Golay Barcode + Forward Primer Pad + Linker + Primer + DNA

AATGATAACGGCGACCACCGAGATCTACAGCT | AGCCTTCGTCGC | TATGGTAATT | GT | GTGYCAGCMGCCGCGGTAA -

# Workflow

454 Life Sciences  
Library preparation



emPCR



Pyrosequencing

Illumina  
Library preparation



Bridge Amplification



Reversible  
Termination  
Sequencing

Ion Torrent PGM  
Library preparation



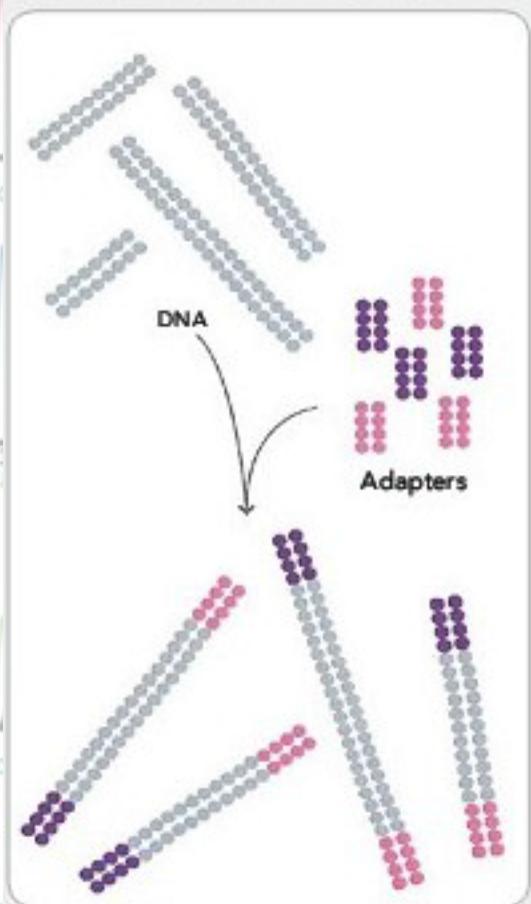
emPCR



Semiconductor  
Sequencing

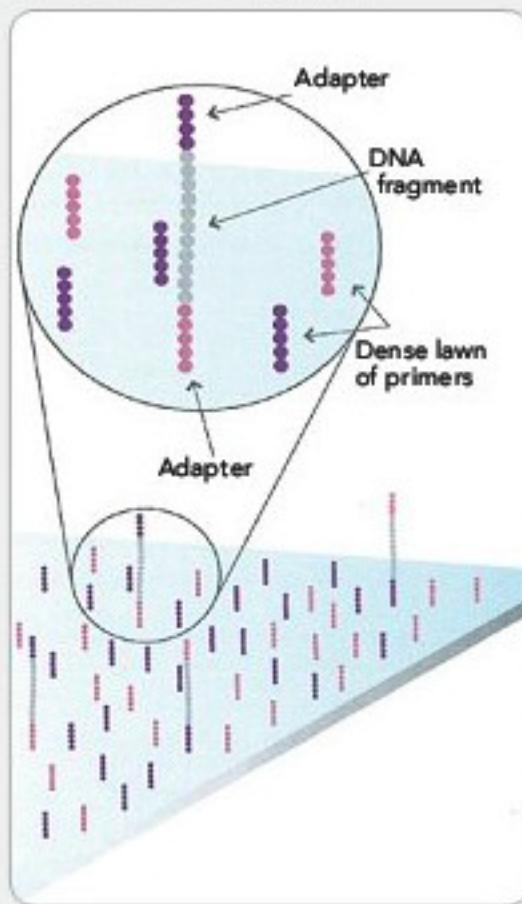
# Workflow: Illumina Bridge Amplification

## 1. PREPARE GENOMIC DNA SAMPLE



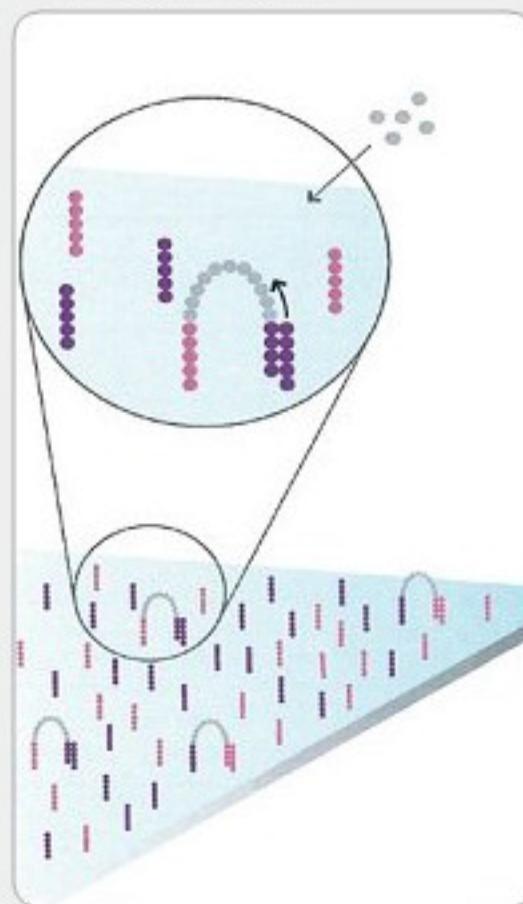
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

## 2. ATTACH DNA TO SURFACE



Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

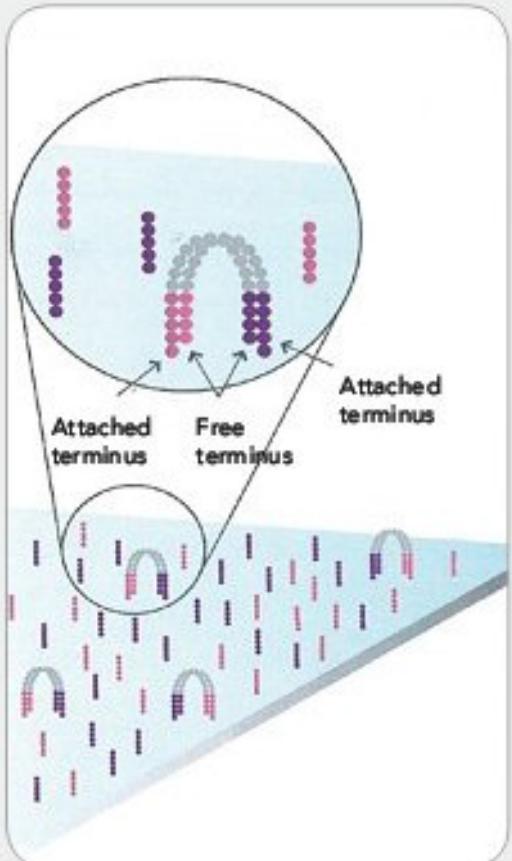
## 3. BRIDGE AMPLIFICATION



Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

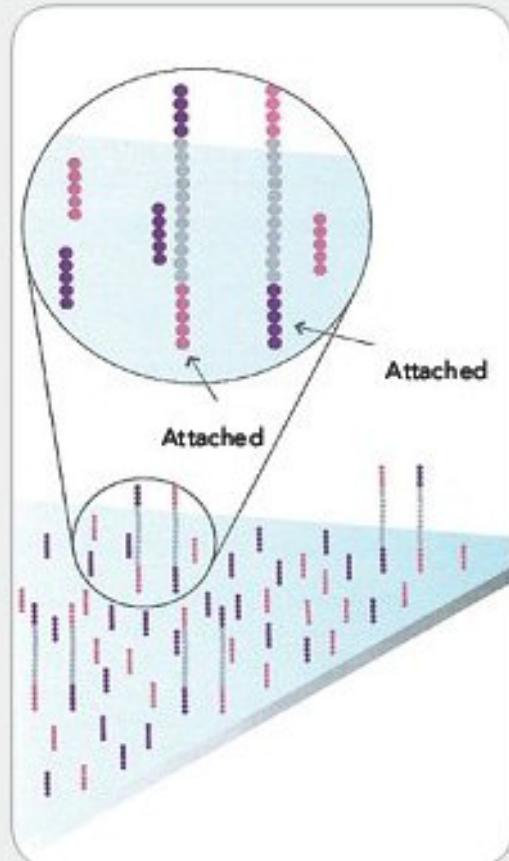
# Workflow: Illumina Bridge Amplification

## 4. FRAGMENTS BECOME DOUBLE STRANDED



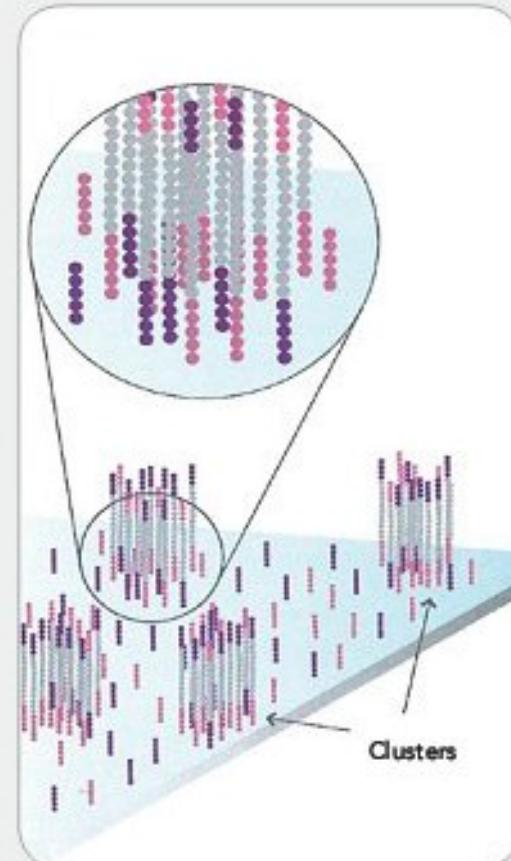
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

## 5. DENATURE THE DOUBLE-STRANDED MOLECULES



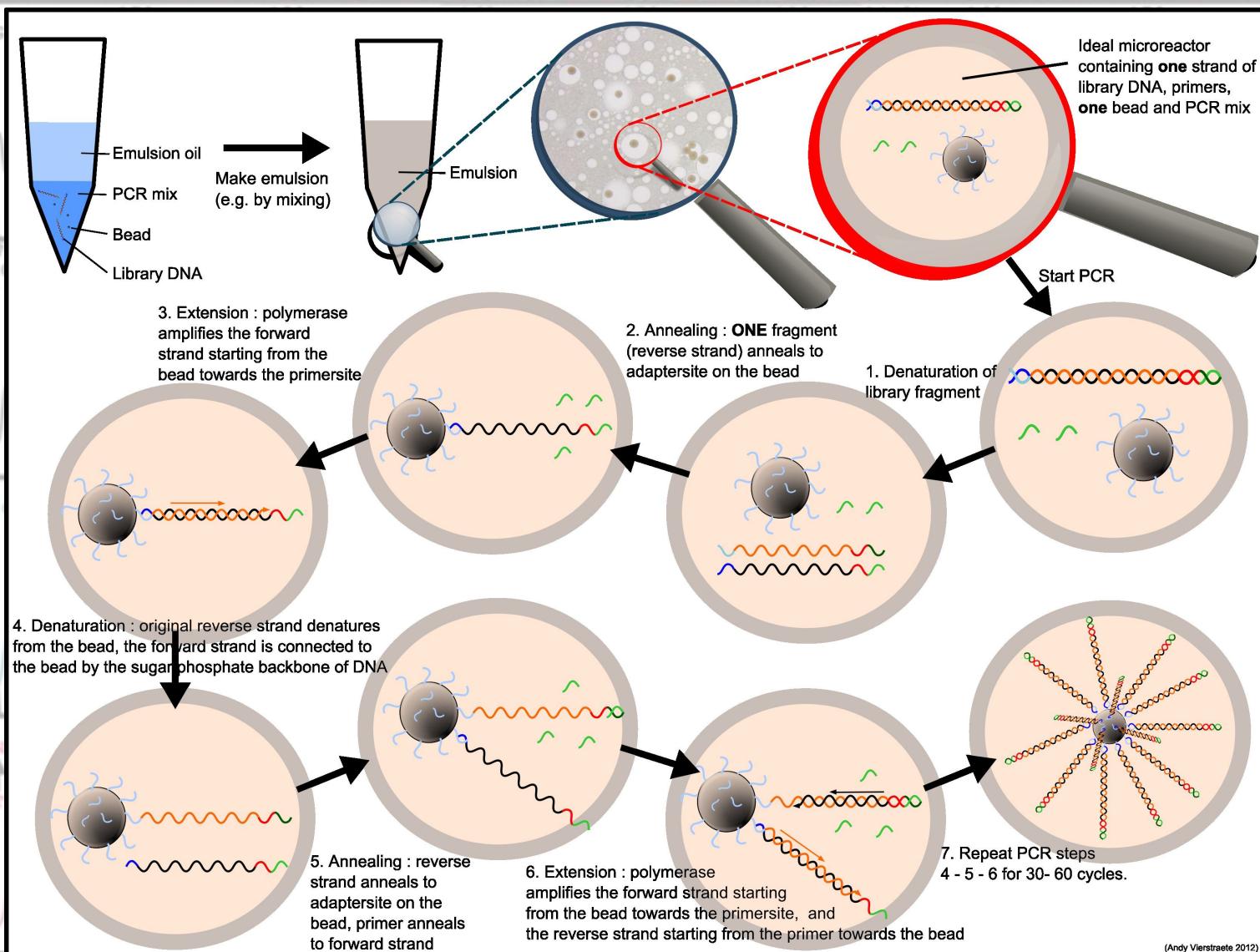
Denaturation leaves single-stranded templates anchored to the substrate.

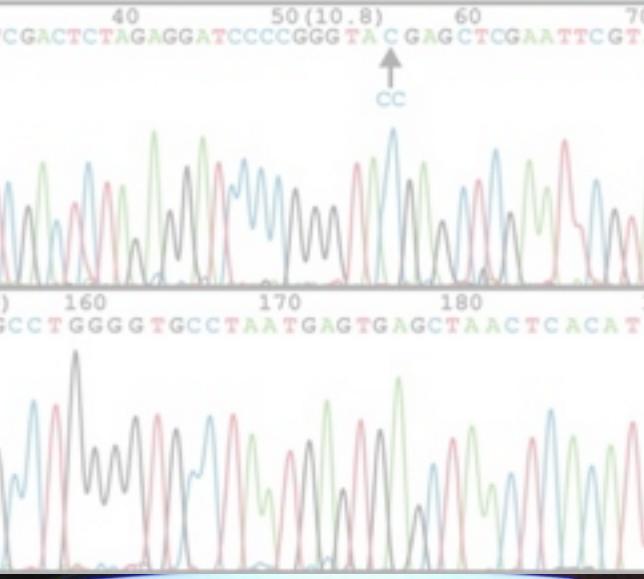
## 6. COMPLETE AMPLIFICATION



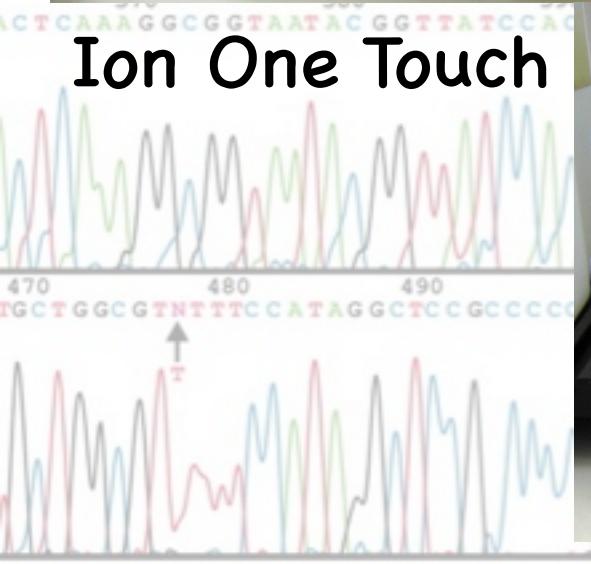
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

# Workflow: Ion Torrent Emulsion PCR



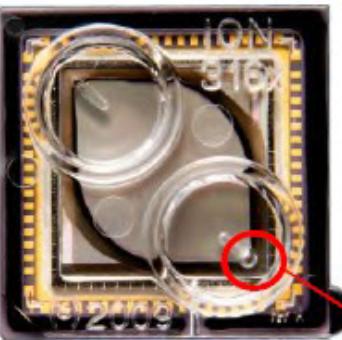


## Ion One Touch

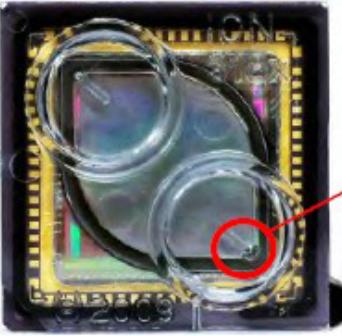


# Workflow: Loading the chip

Ion 316™ Chip:



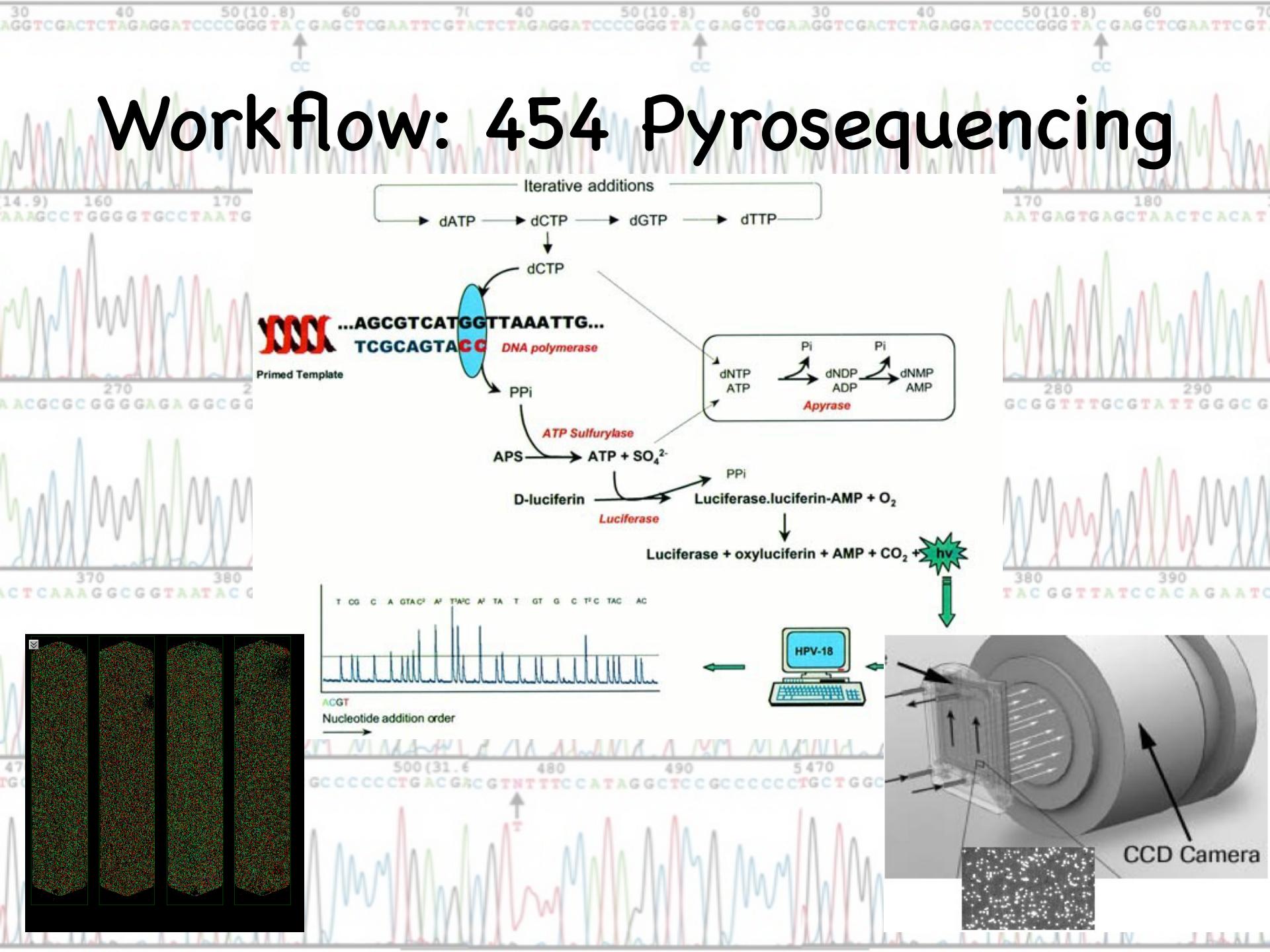
Ion 318™ Chip:



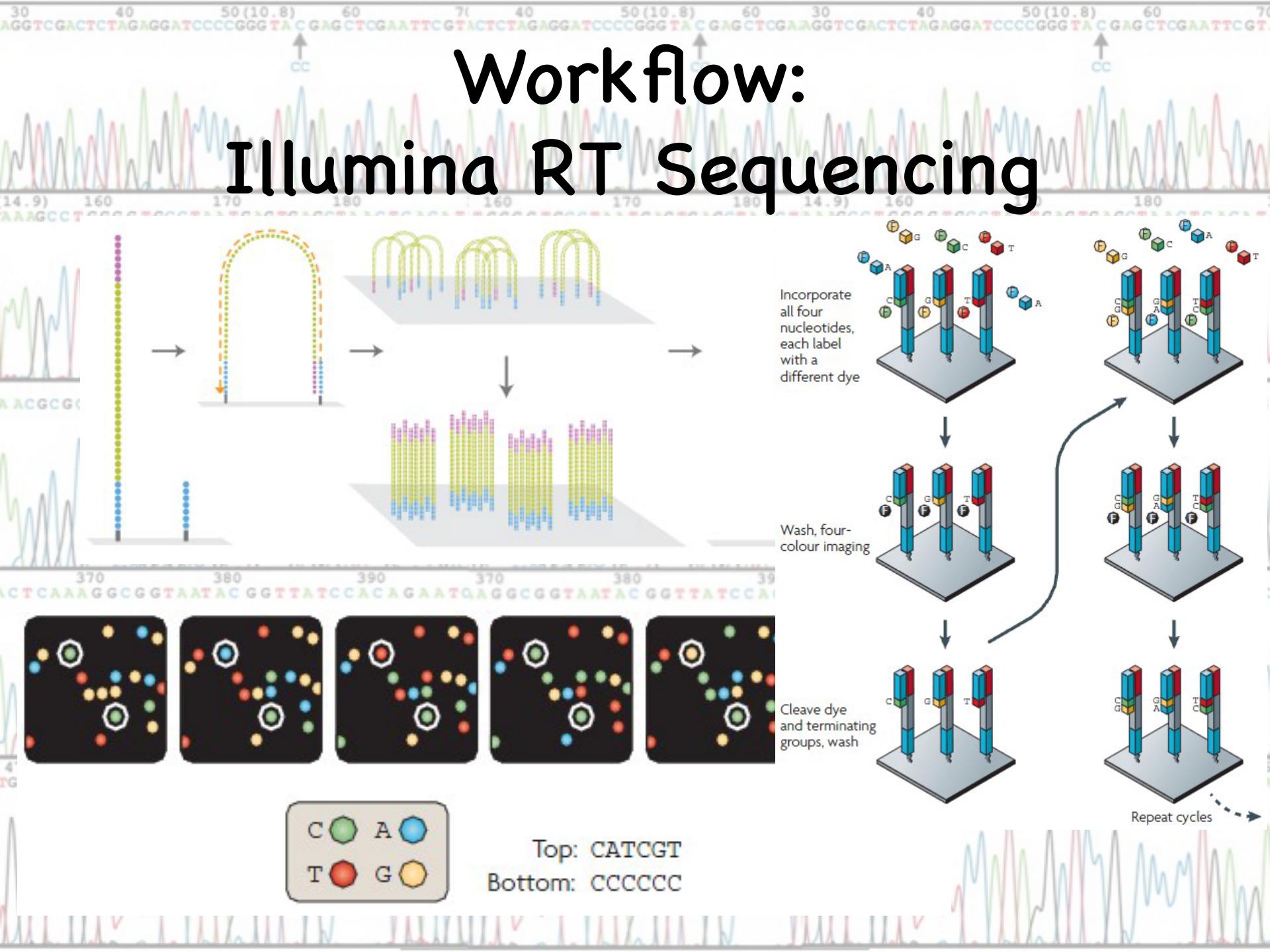
Centrifuge  
adapter  
bucket



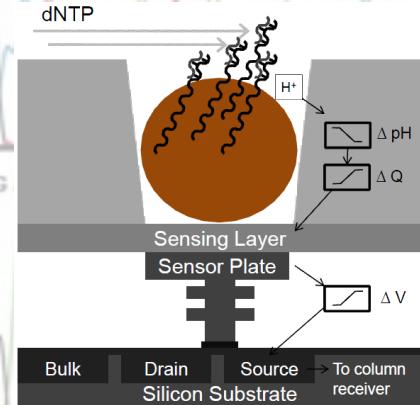
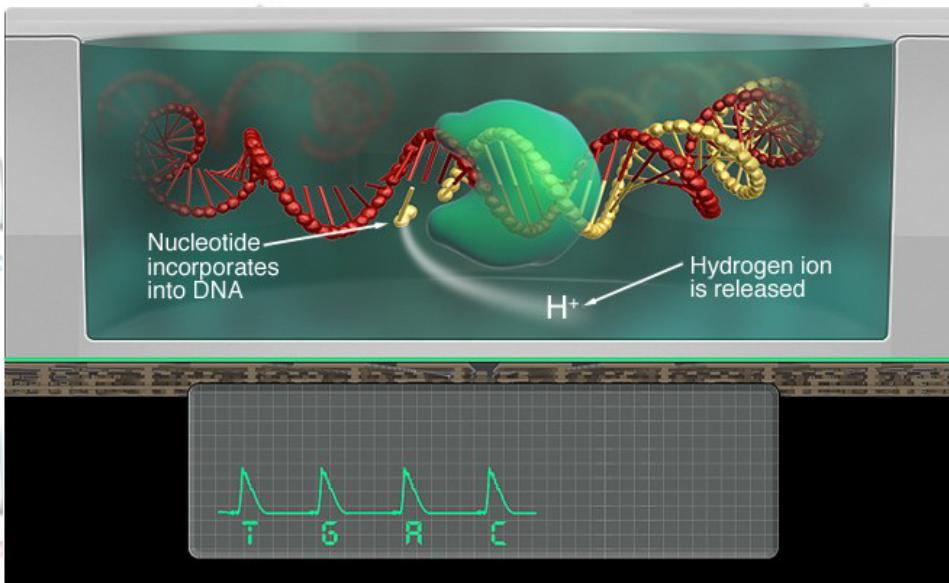
# Workflow: 454 Pyrosequencing



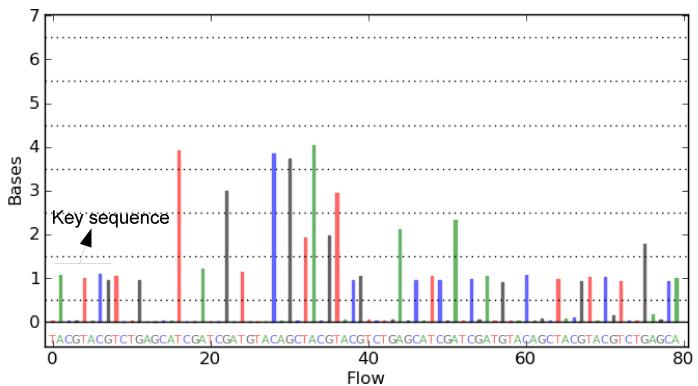
# Workflow: Illumina RT Sequencing



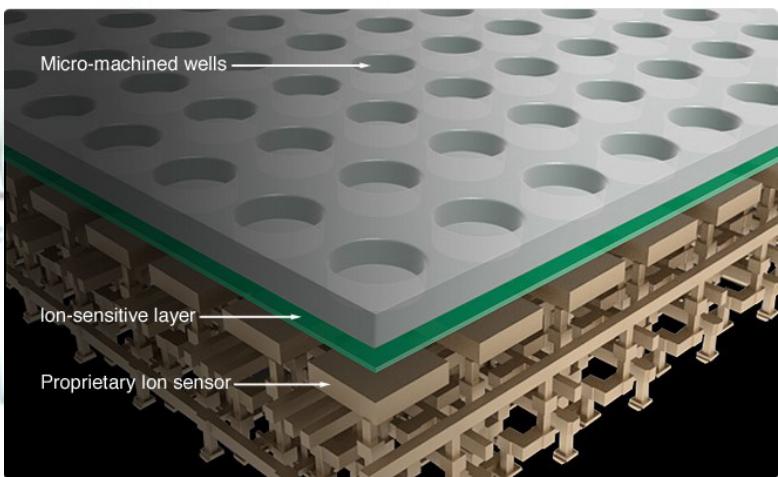
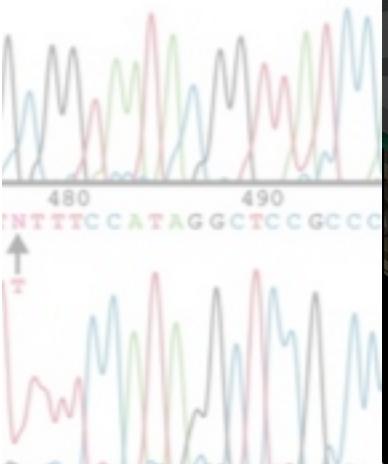
# Workflow: Ion Torrent Semiconductor Sequencing



Average Corrected Ionogram



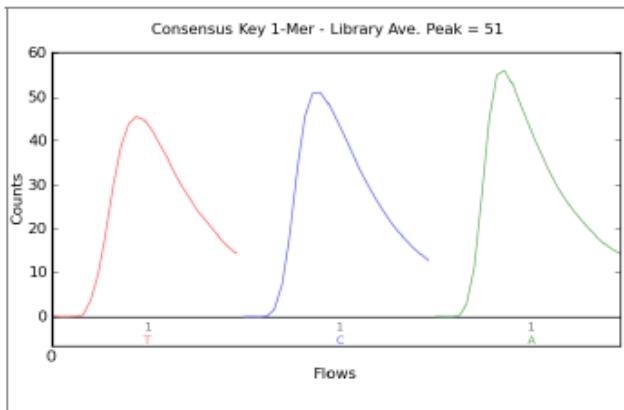
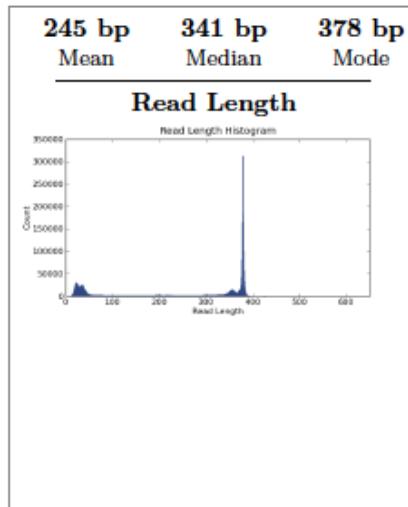
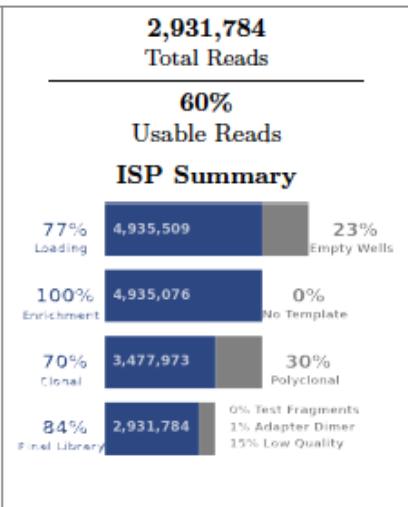
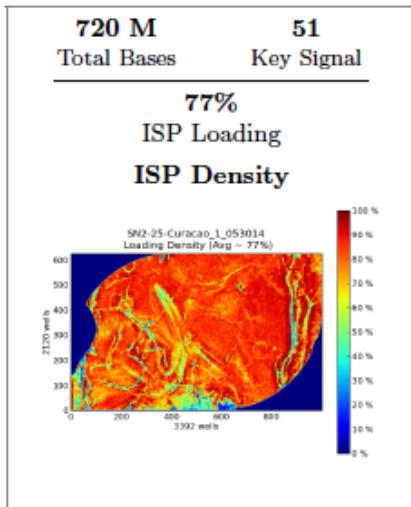
ATCGTACGTCTGAGCATCGATCGATGTACAGCTACGTACGCTCTGAGCATCGATCGATCGATGTACAGCTACGTACGCTCTGAGCA...



# What you get..

## Run Report for Auto\_user\_SN2-25-Curacao\_1\_053014\_52

### Run Summary



**Addressable Wells 6,391,166**

With ISPs	4,935,509	77.2%
Live	4,935,076	100.0%
Test Fragment	8,293	00.2%
Library	4,926,783	99.8%

**Library ISPs 4,926,783**

Filtered: Polyclonal	1,457,103	29.6%
Filtered: Low Quality	519,430	10.5%
Filtered: Primer Dimer	18,466	00.4%
<b>Final Library ISPs</b>	<b>2,931,784</b>	<b>59.5%</b>

Notes: 23 percent enrichment. 1.22 ul of 3 and 7 at 50 pmol. 0.52 ul of 11 at 50 pmol.

# Good data... Bad data?

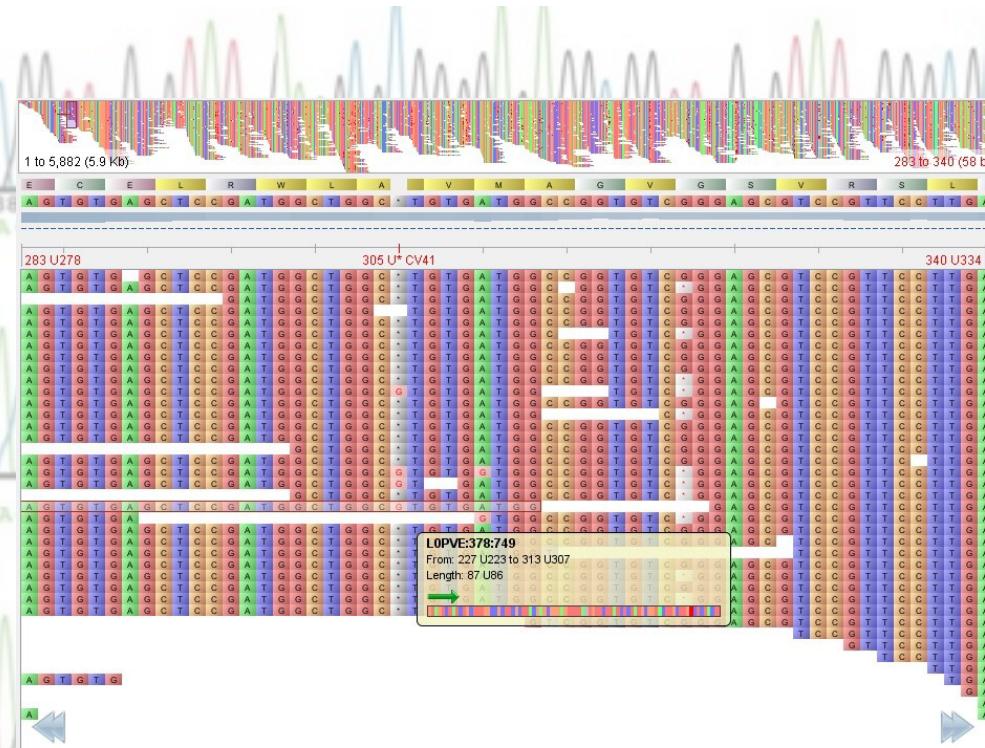
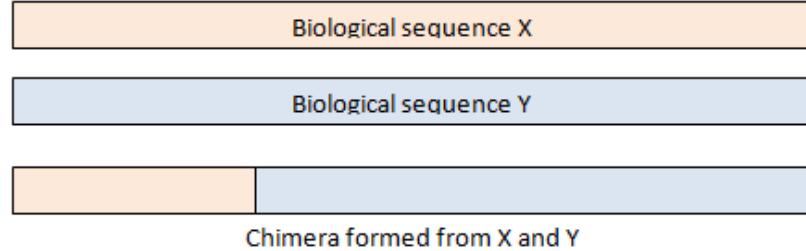
Polyclonality

Short reads

Homopolymer reads

Chimeras

Enough coverage?



# Basic Background to Bioinformatics



Raw sequences

Trim and align sequences

Detection of poor quality sequences

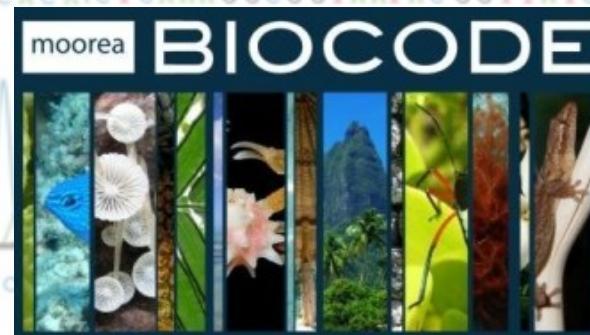
Removal of chimeric sequences

Cluster sequences into OTUs

Search for representative sequences in databases

Use probability based software to increase identities

A curated database is essential



<http://mooreabiocode.org>