Introduction to short read NGS:

Library construction, UCE capture and ddRADseq

The Natural History Museum, London

Autumn 2021

Instructor: Jeff Streicher

j.streicher@nhm.ac.uk

Herpetology Evolution

Integrative Biology

Litoria iris, Papua New Guinea

Course overview

- Unit 1: An introduction to short read high-throughput DNA sequencing and library preparation
- Unit 2: Illumina libraries: *de novo* assembly and reference mapping
- Unit 3: Targeted sequence capture of ultraconserved elements (UCEs)
- Unit 4: Double digest restriction-site associated DNA sequences (ddRADseq)

GitHub

https://github.com/nhm-herpetology/museum-NGS-training

Course overview

- Unit 1: 9th and 10th September
- Unit 2: 16th, 17th and 20th September
- Unit 3: 23rd, 24th and 27th September
- Unit 4: 30th September and 1st and 4th October

GitHub https://github.com/nhm-herpetology/museum-NGS-training

What we **will** be covering

- The Illumina® platform
- Laboratory methods for generating Illumina sequencing libraries
- Practical examples of the different bioinformatic steps needed to analyze Illumina data

What we **won’t** be covering

- Non-Illumina short read platforms
- Long read ‘3rd generation sequencing’ methods (e.g. PacBio, Oxford Nanopore)
- Analyses beyond initial data cleaning and alignment/assembly (e.g. phylogenetic/population genetic analyses)

Unit 1: Introduction to short read sequencing and library preparation

Lecture

GitHub https://github.com/nhm-herpetology/museum-NGS-training

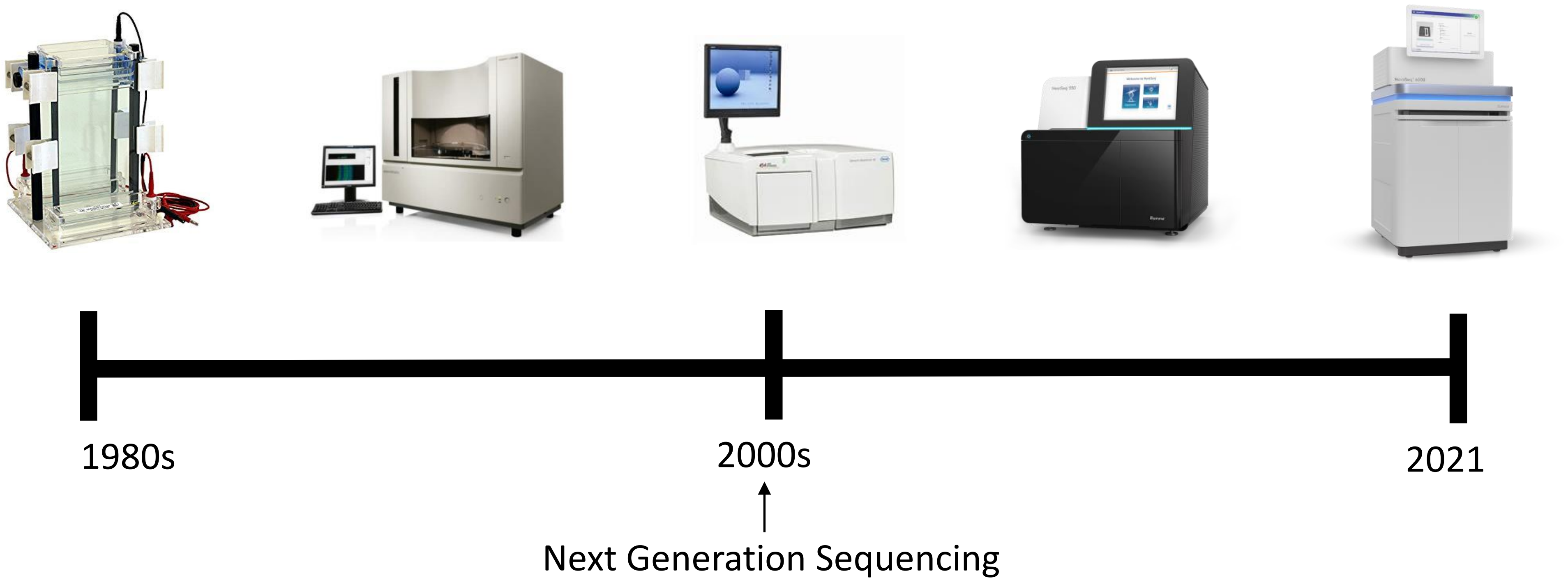
Next Generation Sequencing (NGS)

aka high-throughput sequencing
aka 2nd generation sequencing

New methods for DNA sequencing were developed in the mid to late 1990s and early 2000s. These were dubbed the "next-generation" or "second-generation" sequencing methods, in order to distinguish them from the earlier methods, including Sanger sequencing.

In contrast to the first generation of sequencing, NGS technology is typically characterized by being highly scalable, allowing entire genomes to be sequenced at once.

Text *mostly* from Wikipedia ☺



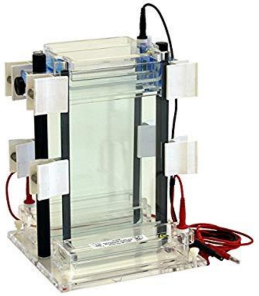
1980s

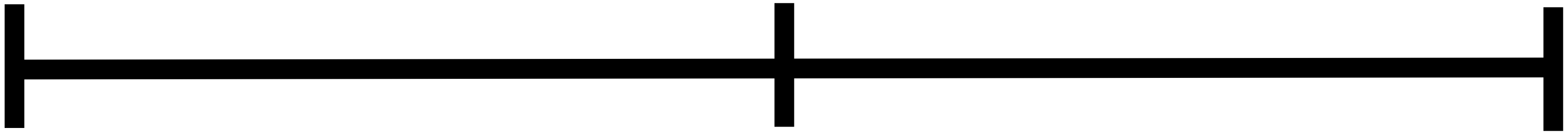
2000s

2021

Next Generation Sequencing

200-1000 bp per sample





1980s

2000s

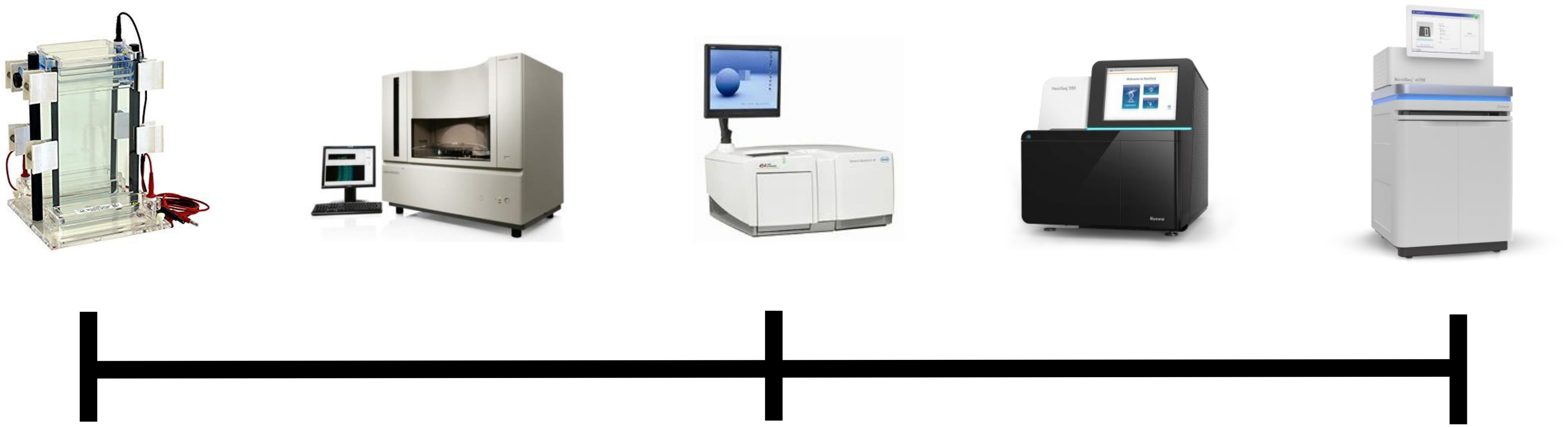
↑

Next Generation Sequencing

2021

200-1000 bp per sample

1 million+ bp per sample



1980s

2000s

↑

Next Generation Sequencing

2021

Next Generation Sequencing (NGS)

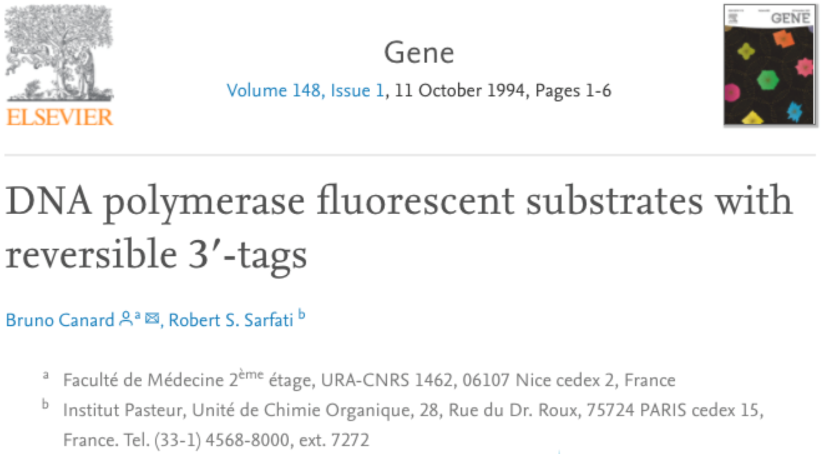
aka high-throughput sequencing
aka 2nd generation sequencing

One of the first NGS methods was based on fragmenting genomes into small pieces, randomly sampling for a fragment, and sequencing ~75 to 300 bp of the fragment. **The resulting small piece of DNA that is sequenced is where the term "short read" originates.**

The technology/platform that now dominates short read sequencing is called Illumina (Solexa) sequencing.

History of Illumina (Solexa) method

- Reversible dye-terminators / Sequencing-by-Synthesis (SBS)
- Initial biochemical reaction description Canard & Sarfati (1994)
- Further development in 1998 by Shankar Balasubramamian and David Klenerman @Cambridge into Solexa method
- Purchased by Illumina in 2007 for $600 million USD



ELSEVIER

Gene

Volume 148, Issue 1, 11 October 1994, Pages 1-6

DNA polymerase fluorescent substrates with reversible 3′-tags

Bruno Canard[a], Robert S. Sarfati[b]

[a] Faculté de Médecine $2^{ème}$ étage, URA-CNRS 1462, 06107 Nice cedex 2, France

[b] Institut Pasteur, Unité de Chimie Organique, 28, Rue du Dr. Roux, 75724 PARIS cedex 15, France. Tel. (33-1) 4568-8000, ext. 7272

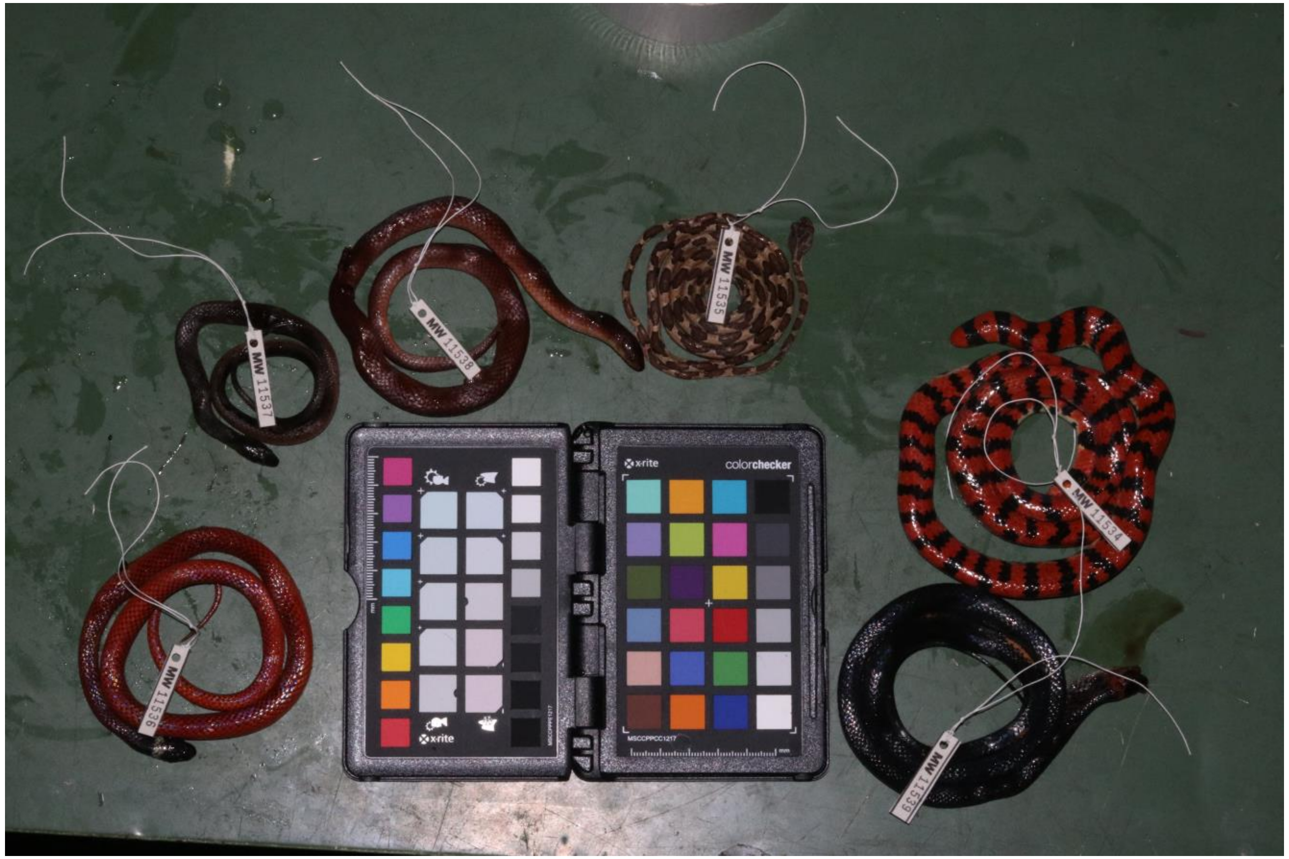
Solexa

illumina®

The Solexa method: How does it work?

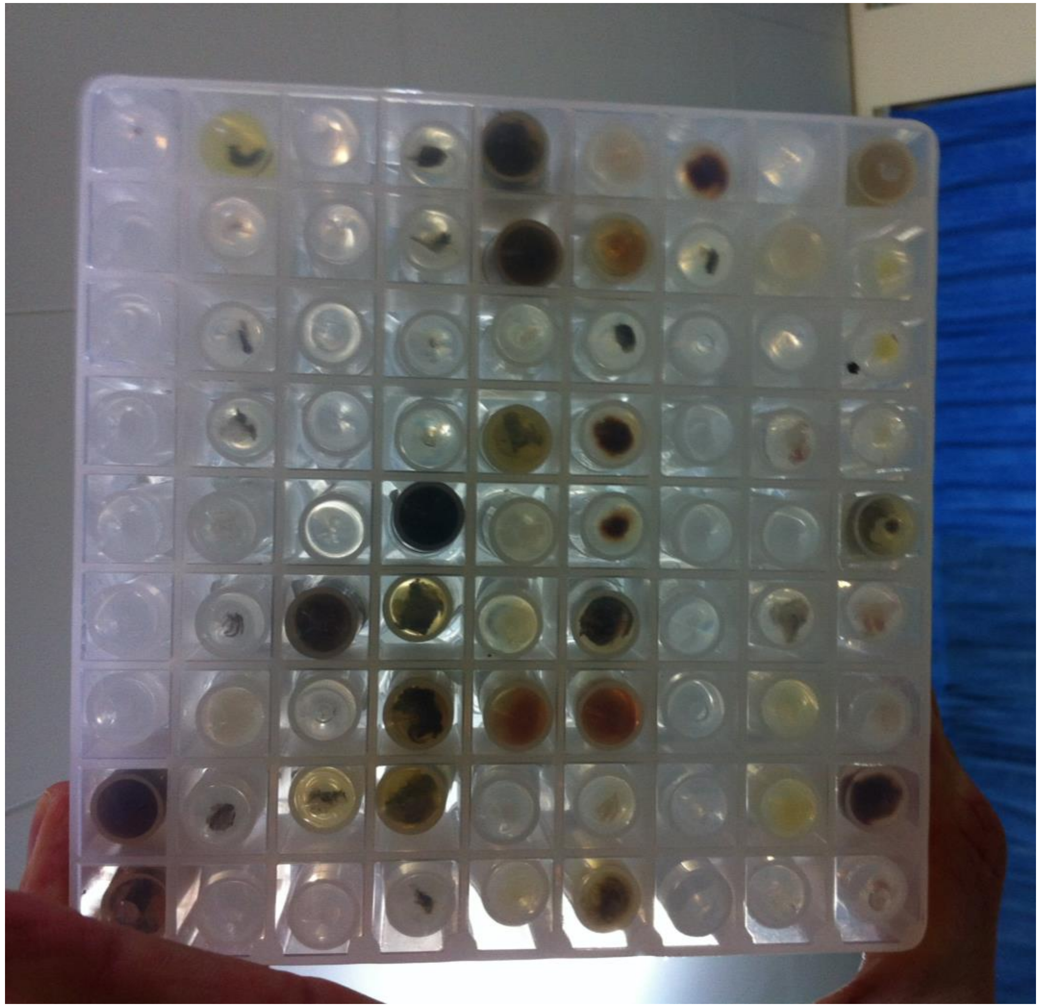
- Fragmentation of genomic DNA
- Construction of DNA sequencing library (containing many ‘reads’)
- Cluster generation on a flow cell (aka bridge amplification)
- Clonal amplification
- ‘Sequencing-by-Synthesis’

First steps for organismal biologists

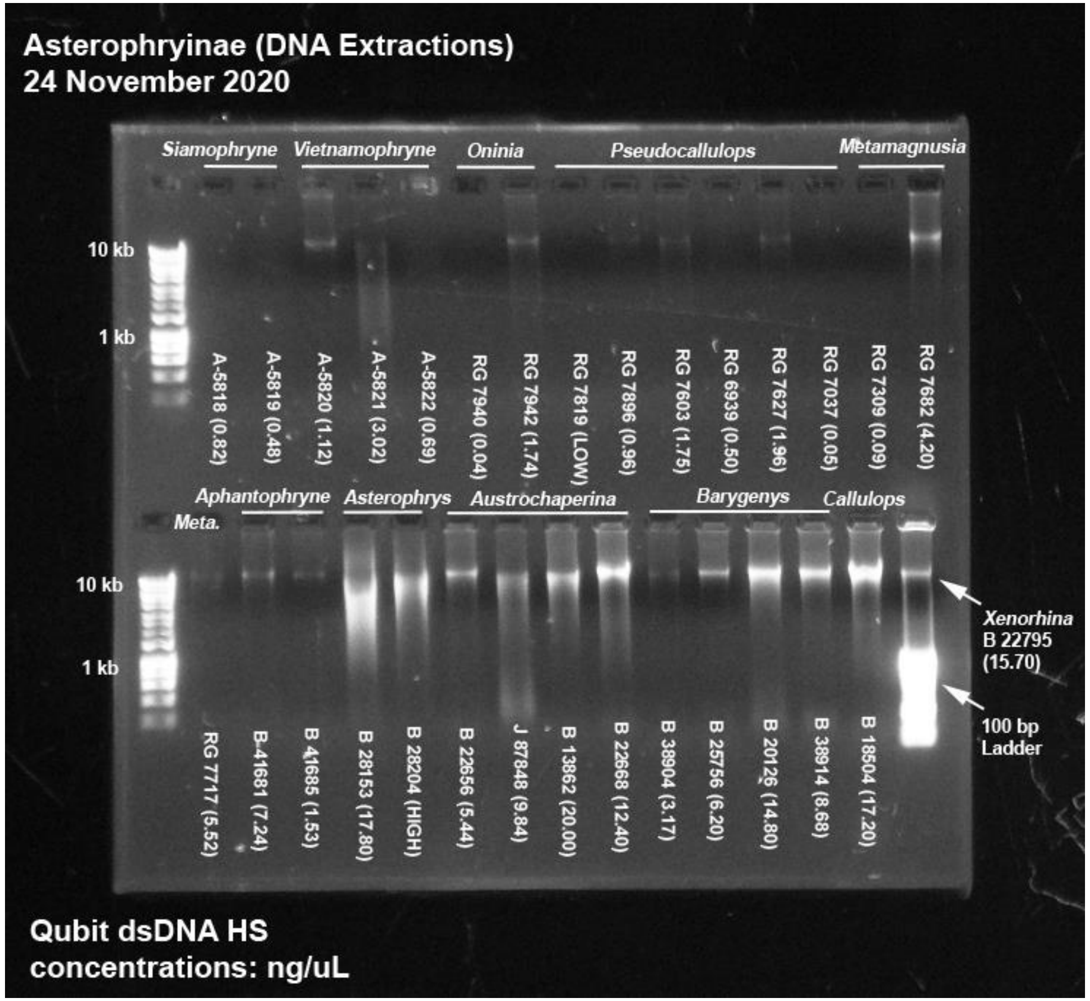


Fieldwork
(Specimen and data collection, photography, preservation, field ID numbers etc.)

First steps for organismal biologists



Tissue sampling
(Muscle, liver, etc.)



Asterophryinae (DNA Extractions)
24 November 2020

Qubit dsDNA HS
concentrations: ng/uL

DNA extraction
(Qiagen kit, Phenol-chloroform, salt extraction etc.)

Fragmentation of genomic DNA

- Many ways to shear DNA…

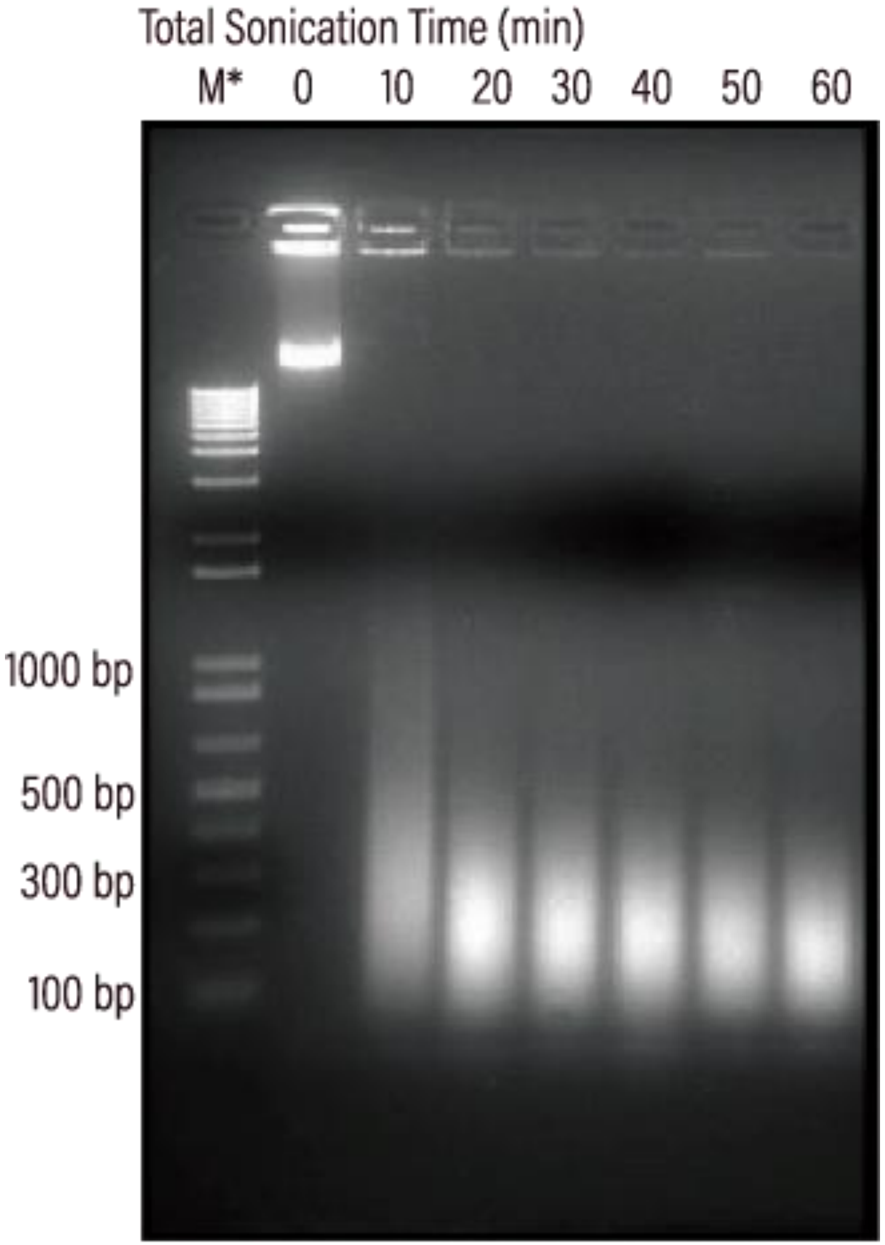
Nebulizer

Sonicator

Hydroshear

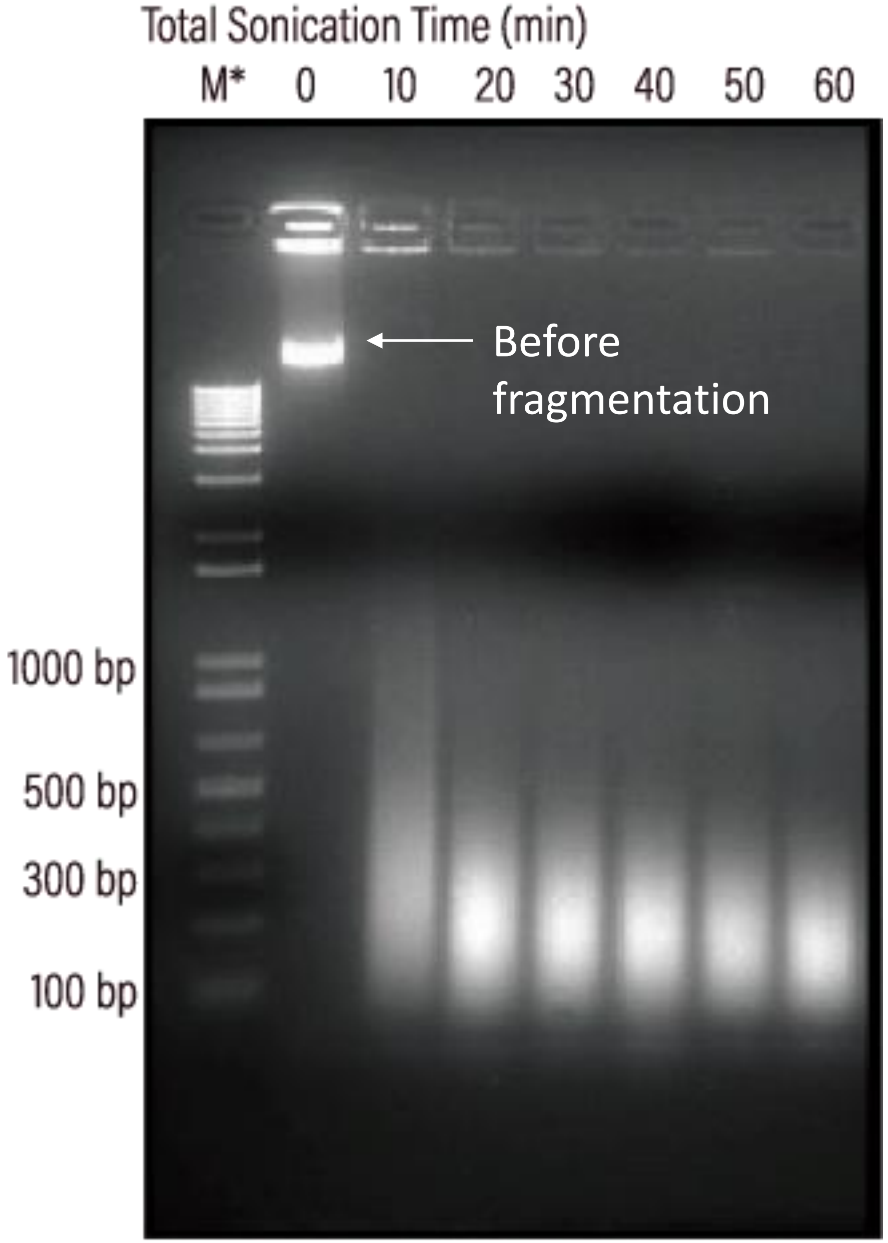
Enzymatically

Time



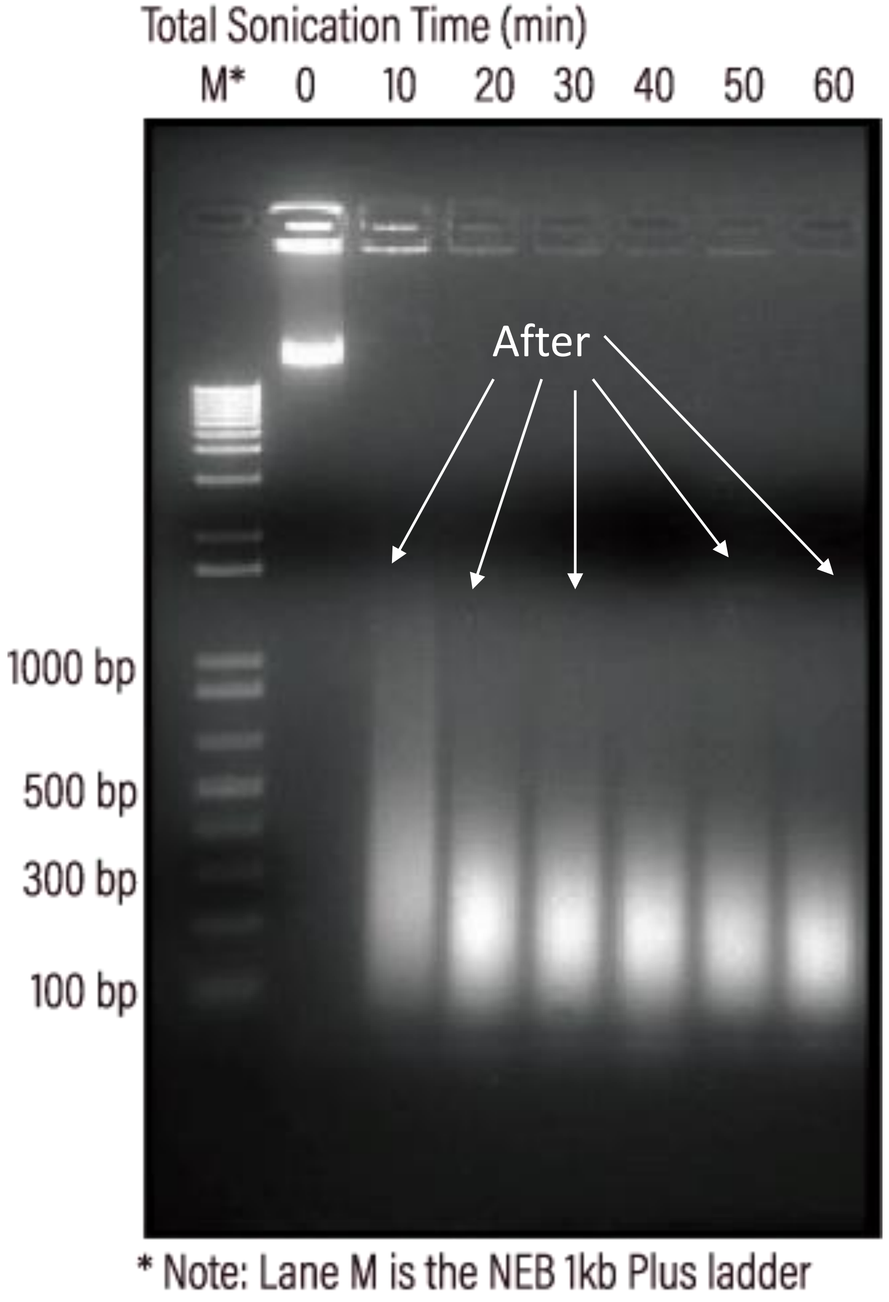
* Note: Lane M is the NEB 1kb Plus ladder

Image from Qsonica



* Note: Lane M is the NEB 1kb Plus ladder

Image from Qsonica

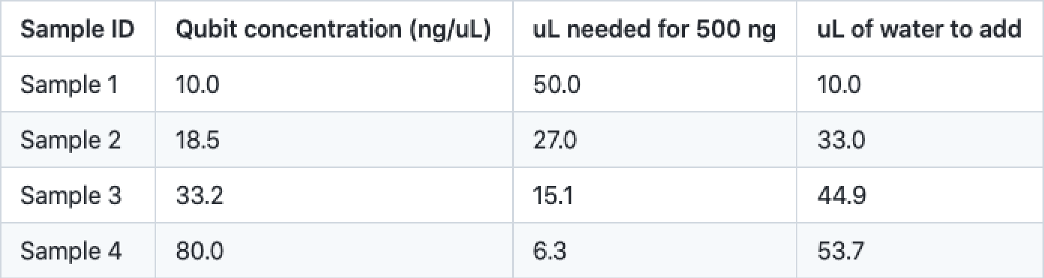


* Note: Lane M is the NEB 1kb Plus ladder

Image from Qsonica

Quantification of fragmented genomic DNA

- Most genomic library construction protocols require specific starting concentrations of fragmented DNA.
- We need to determine the concentration of double-stranded DNA (dsDNA) before or after the fragmentation.
- One of the most effective ways to do this (IMO) is with fluorometry.
- We will cover this during the molecular labs tomorrow and next week.



Sample ID	Qubit concentration (ng/uL)	uL needed for 500 ng	uL of water to add
Sample 1	10.0	50.0	10.0
Sample 2	18.5	27.0	33.0
Sample 3	33.2	15.1	44.9
Sample 4	80.0	6.3	53.7

Table from Unit 2 Molecular Lab Protocol
https://github.com/nhm-herpetology/museum-NGS-training/tree/main/Unit_02/Molecular_Lab

Qubit 2.0 Fluorometer

Genomic library construction

- End-repair of fragmented DNA
- dA-tailing
- Adapter ligation
- Size-selection for optimal fragment lengths
- PCR amplification
- Quantification

End Repair

Illumina genomic library construction

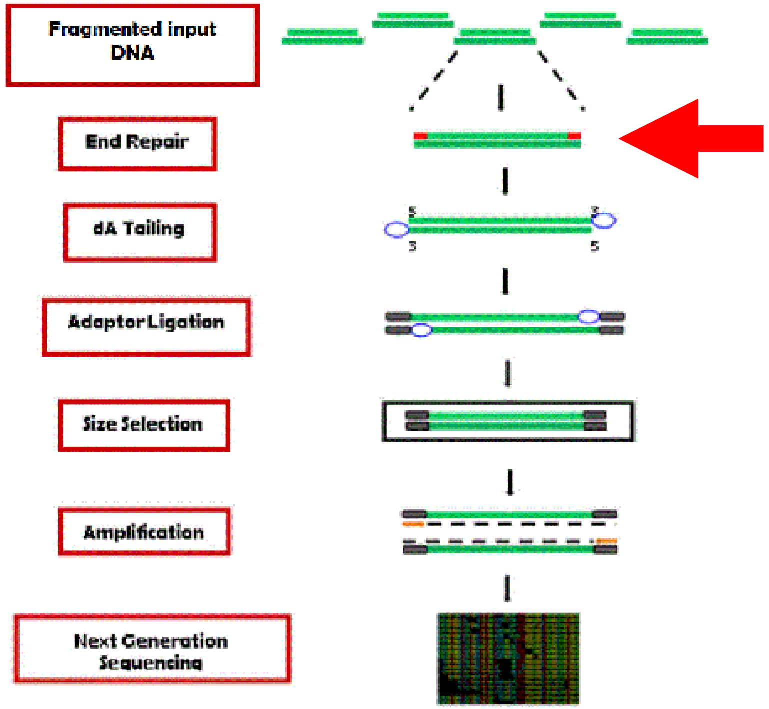


Image from ENZo Life Science

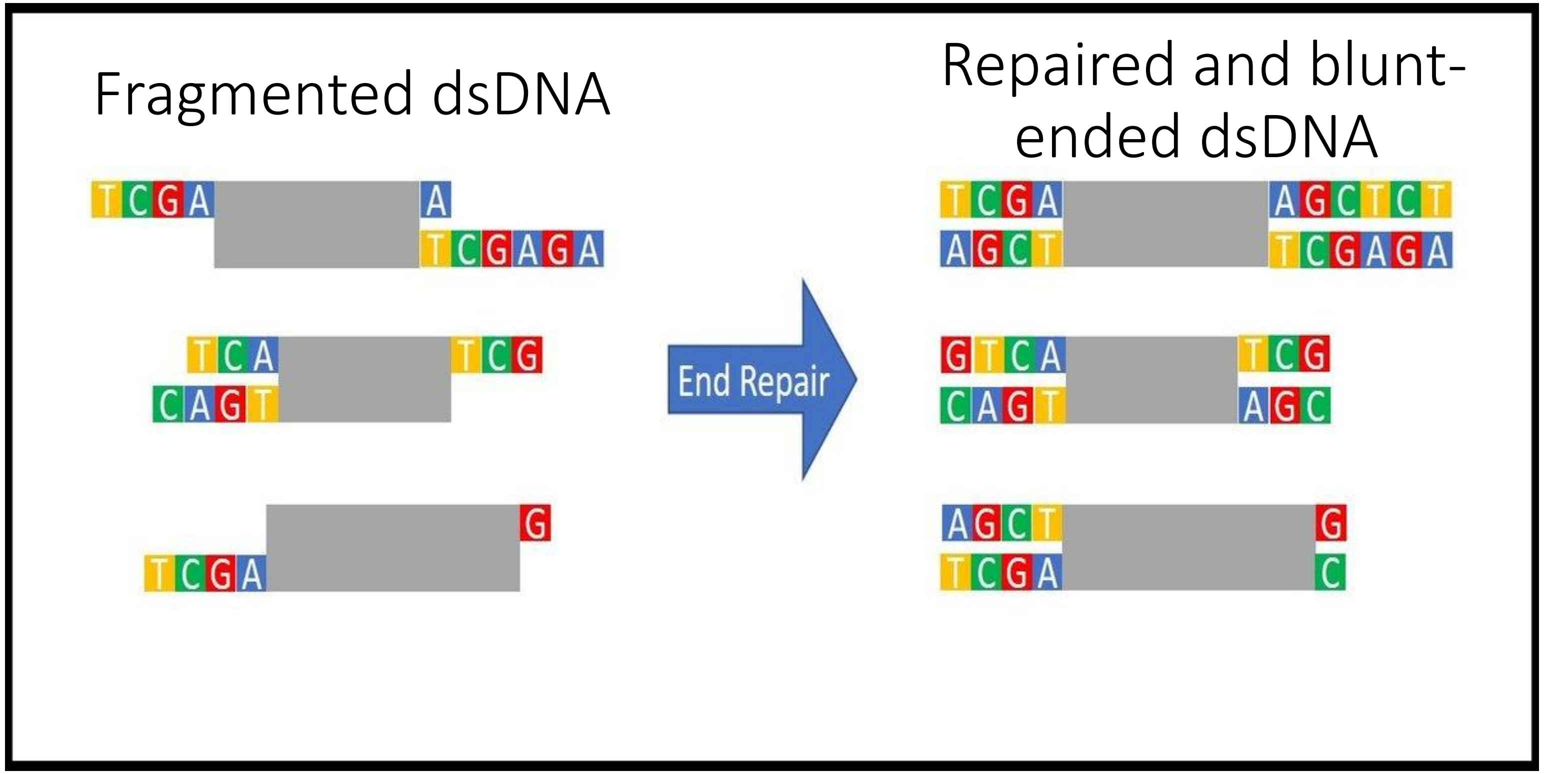


Image from New England BioSciences

Illumina genomic library construction

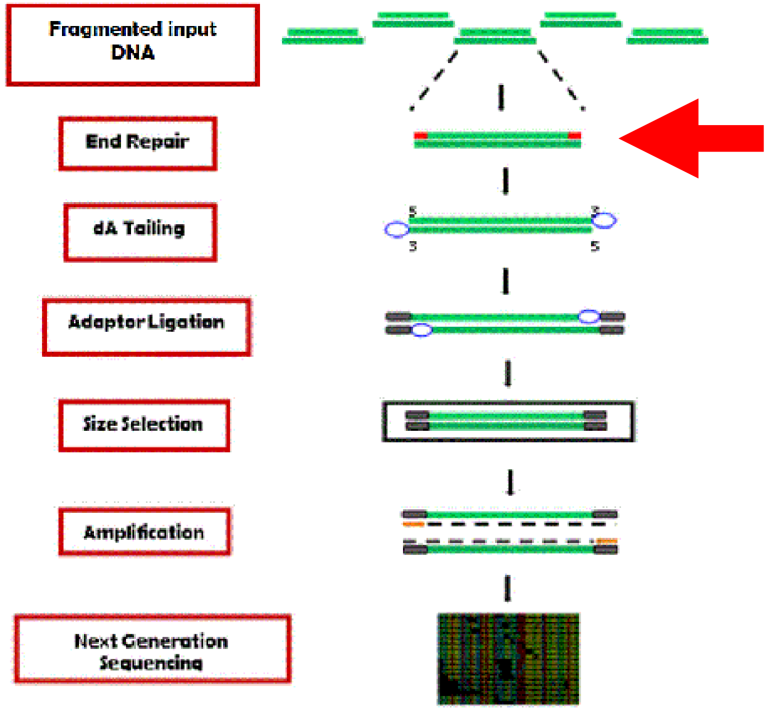


Image from ENZo Life Science

End Repair

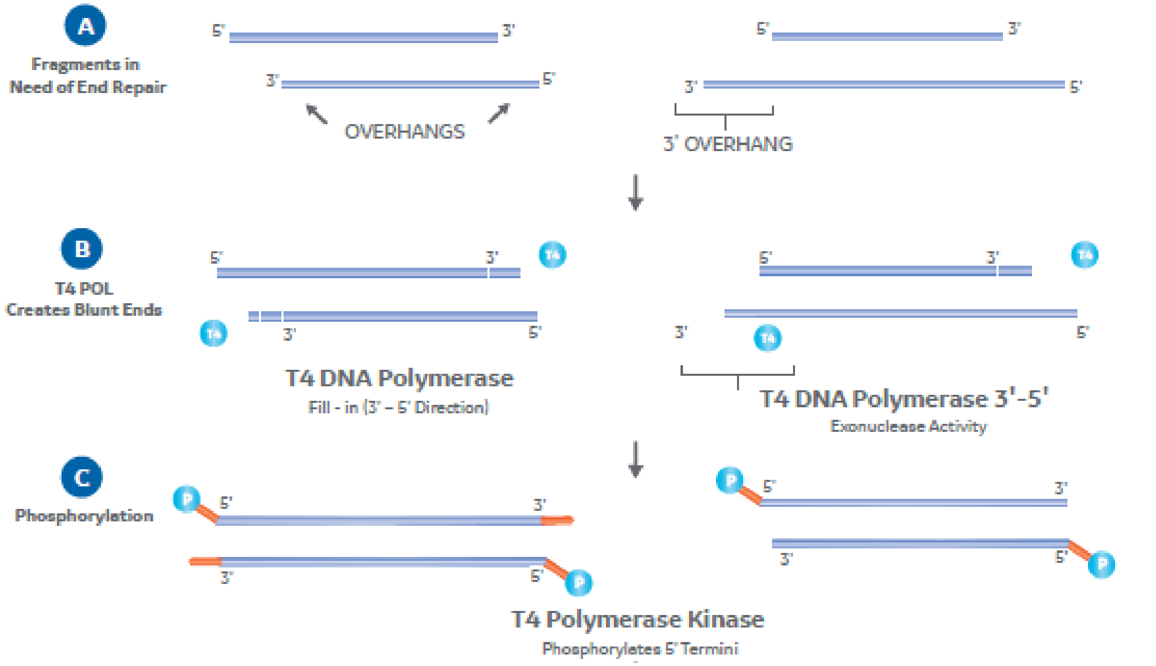


Image from Cytiva

A typical blunting enzyme mix will contain T4 DNA polymerase, dNTPs, and T4 polynucleotide kinase (PNK). T4 DNA polymerase (in the presence of dNTPs) fills-in 5' and overhangs and trims 3' overhangs to generate blunt-ended dsDNA (A-B). The T4 PNK can then phosphorylate the 5' terminal nucleotide (C).

dA Tailing

Illumina genomic library construction

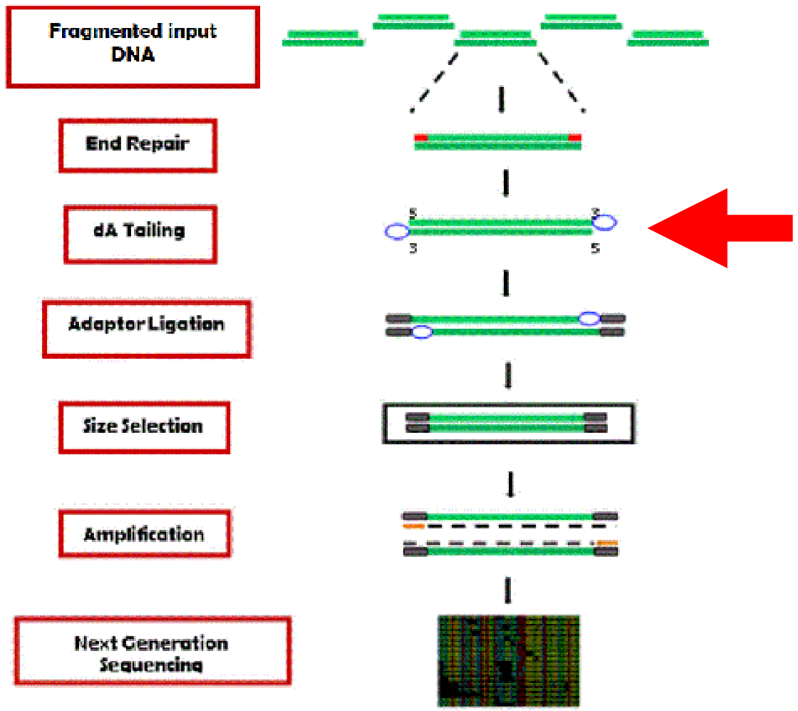


Image from ENZo Life Science

End Repaired DNA

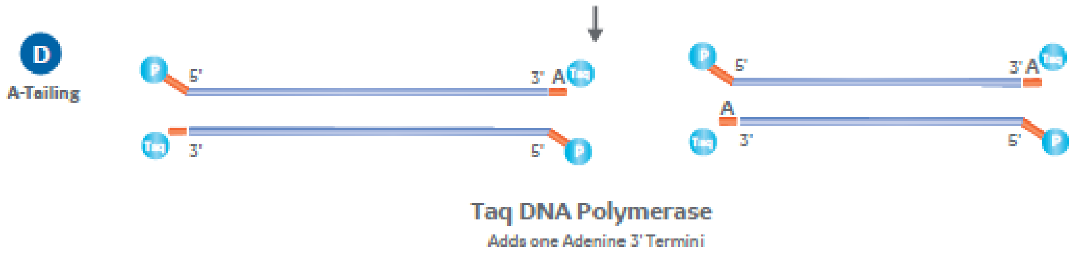


Image from Cytiva

A-tailing also requires a polymerase. Taq DNA polymerase the most common as it has terminal transferase activity and naturally leaves a 3’ terminal adenine (D). **DNA polymerase I Large (Klenow) fragment** is another common option (this is what we will use in Unit 2). Using either of these polymerases leaves A-tailed ends that complement standard Illumina adaptors.

Illumina genomic library construction

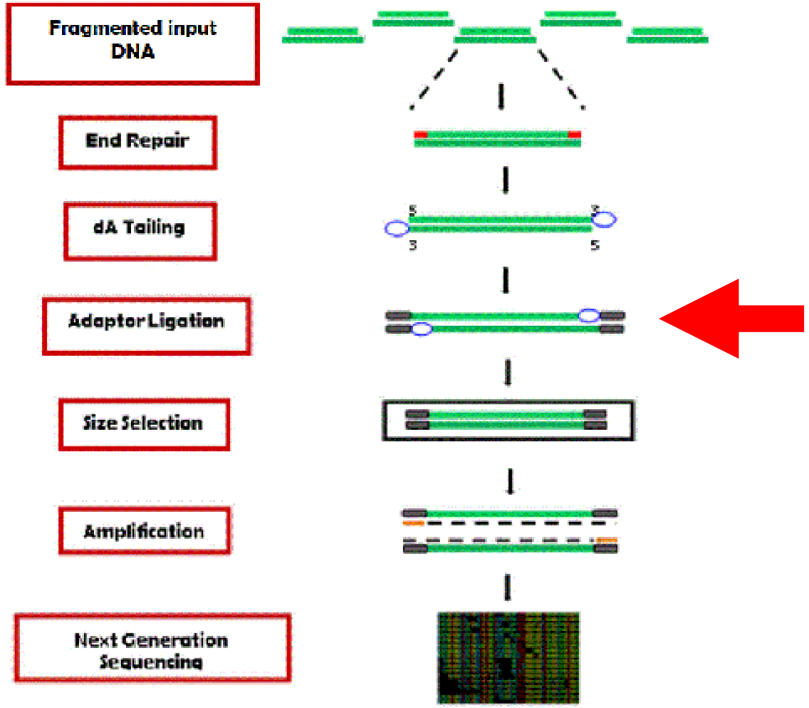


Image from ENZo Life Science

Adaptor Ligation

dA-tailed DNA

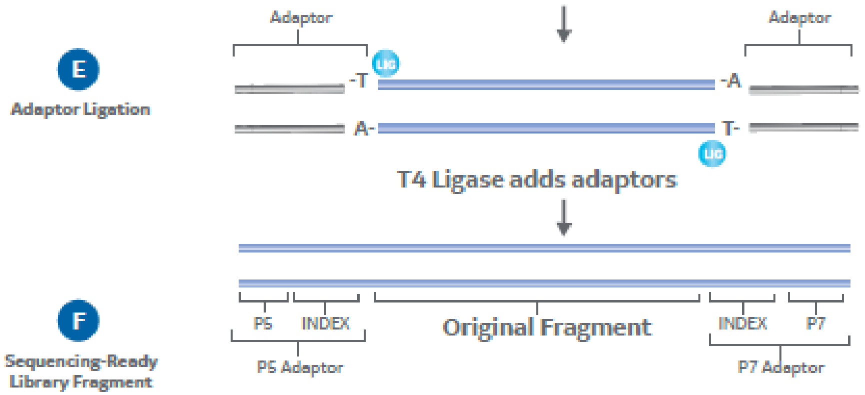
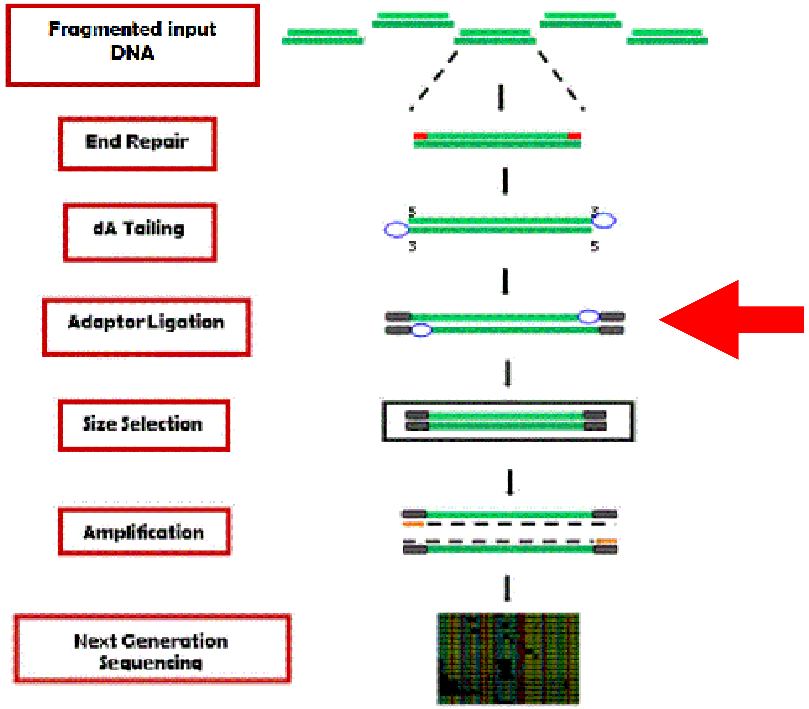


Image from Cytiva

Adding an adaptor at this stage just requires an incubation with T4 DNA ligase . This enzyme will join both blunt and so-called ‘sticky’ ends, in this case catalyzing the formation of a phosphodiester bond between the 5’ and 3’ termini of the end-repaired fragments and sequencing adaptors (E-F).

Illumina genomic library construction



Illumina Adapter Design

“Stubby, Y-Yoked Adapters”

- One oligo with terminal thymine (Required)
- One oligo with phosphorylated terminal nucleotide (Required)
- Illumina P5 and P7 recognition sequences (Required)
- Read 1 and Read 2 priming sequences (Required)
- Unique Index (for multiplexing; Required)
- Second Index (for multiplexing; Optional)
- Unique Molecular Identifier (UMI; Optional)

Illumina Adapter Design

“Stubby, Y-Yoked Adapters”

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATC T

CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATC*PHOS

Illumina Adapter Design

“Stubby, Y-Yoked Adapters”

P5 (i5) Illumina sequence

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC **T**

Read 1 priming sequence

P7 (i7) Illumina sequence

CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC***PHOS**

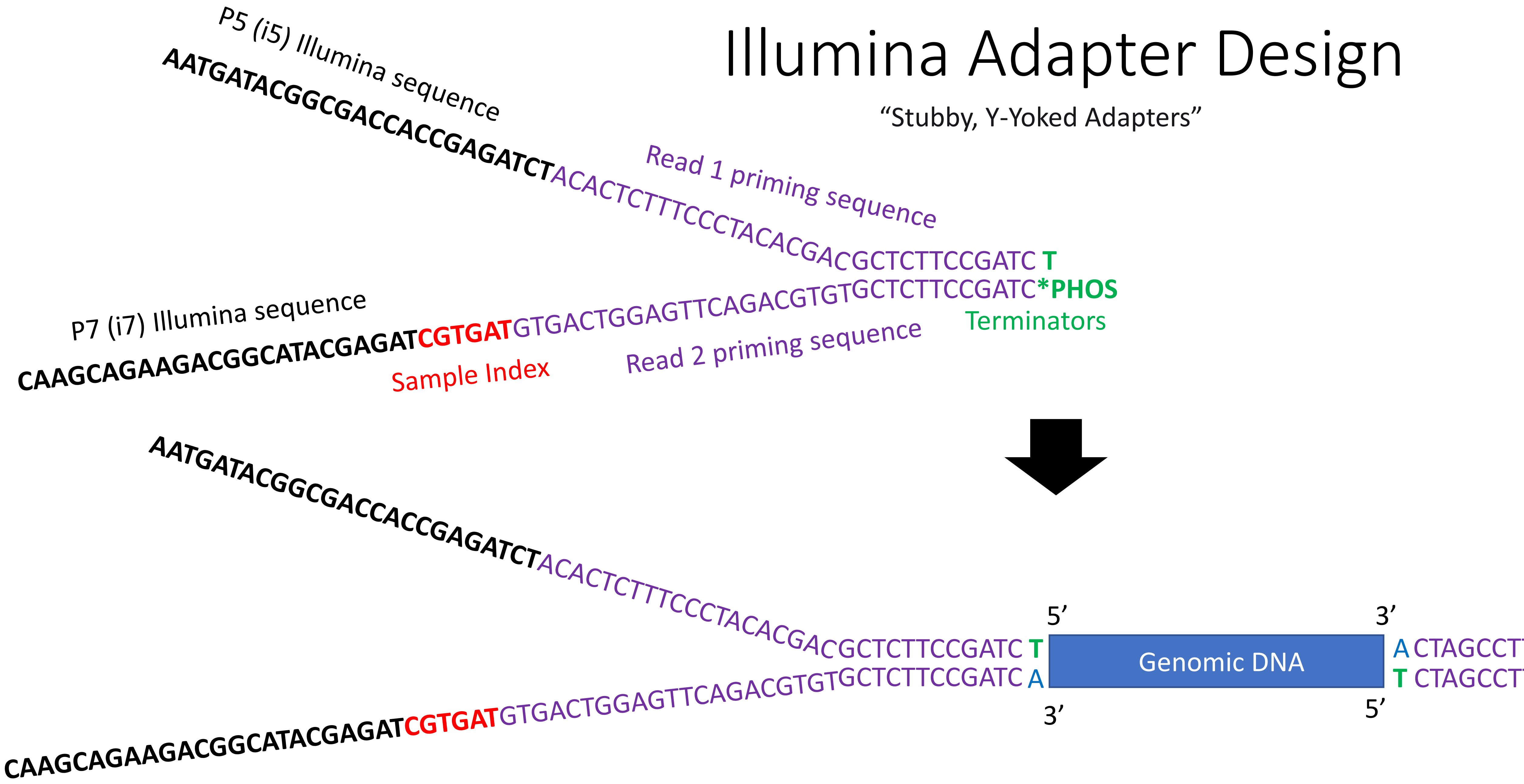
Sample Index

Read 2 priming sequence

Terminators

Illumina Adapter Design

“Stubby, Y-Yoked Adapters”



P5 (i5) Illumina sequence

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC **T**

Read 1 priming sequence

P7 (i7) Illumina sequence

CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC***PHOS**

Sample Index

Read 2 priming sequence

Terminators

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC **T**

CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC A

5’

3’

Genomic DNA

3’

5’

A CTAGCCT

T CTAGCCT

Adaptor Ligation

Illumina genomic library construction

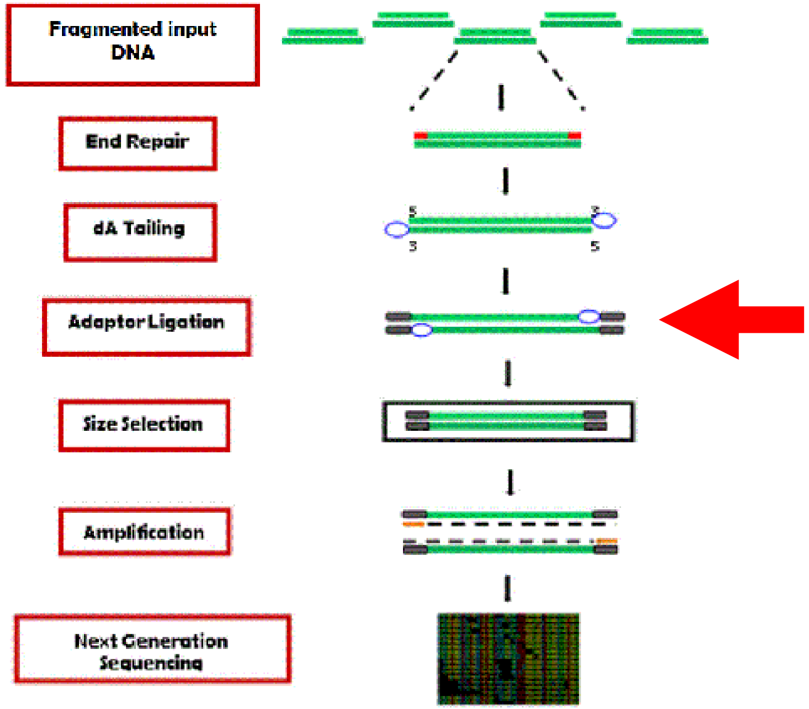
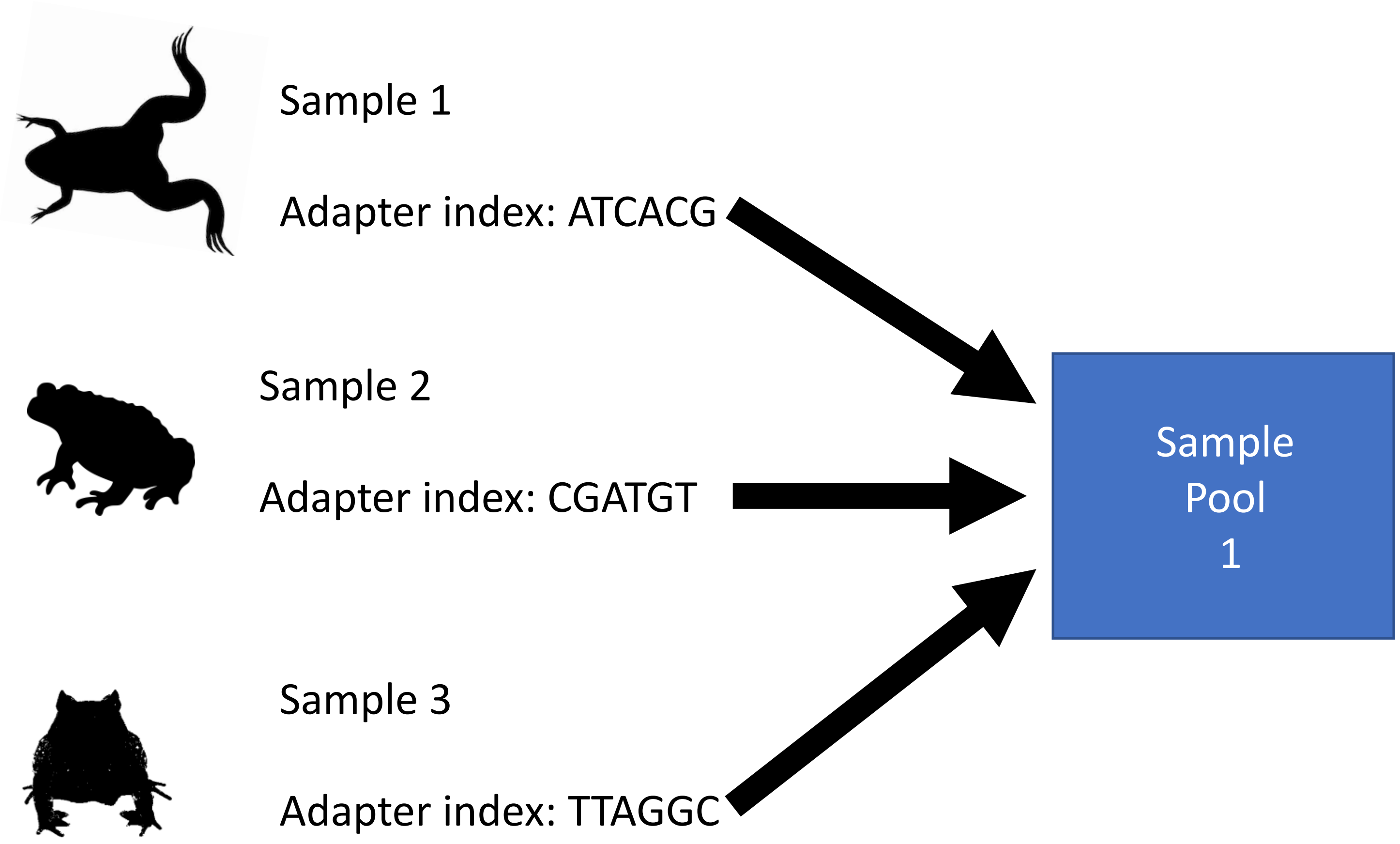


Image from ENZo Life Science

Once adapters have been ligated to the genomic fragments, different samples can be combined because you will bioinformatically sort out the data following sequencing



Sample 1

Adapter index: ATCACG

Sample 2

Adapter index: CGATGT

Sample 3

Adapter index: TTAGGC

Sample Pool 1

Illumina genomic library construction

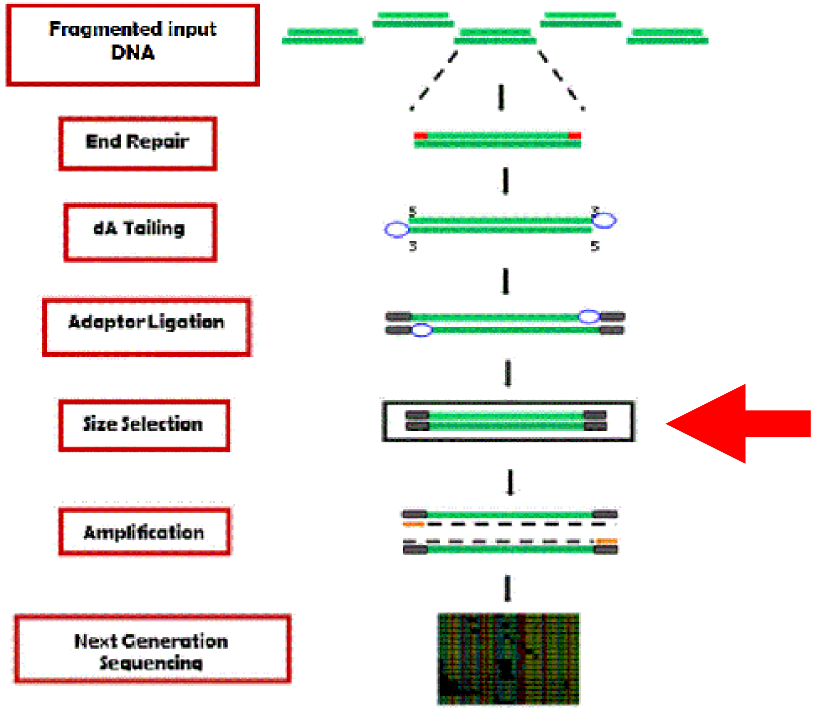


Image from ENZo Life Science

Size Selection

Illumina sequencers can only sequence DNA fragments >600 nucleotides in size, so making sure that the mean size of fragments in your libraries are smaller is critical.

Illumina genomic library construction

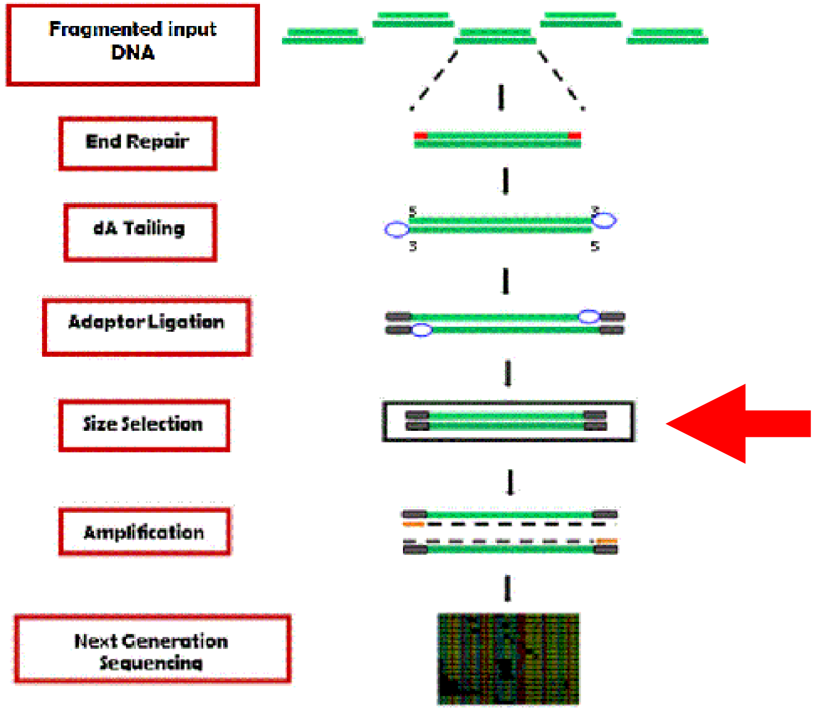


Image from ENZo Life Science

Size Selection

Illumina sequencers can only sequence DNA fragments >600 nucleotides in size, so making sure that the mean size of fragments in your libraries are smaller is critical.

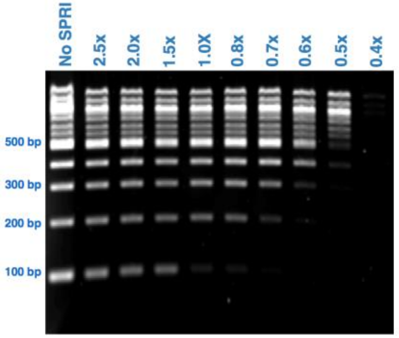


Image from Enseqlopedia

Bead-based size selection

Blue Pippin (Sage Science)

Automated Size Selection

Ideal mean fragment size:
200-500 base pairs

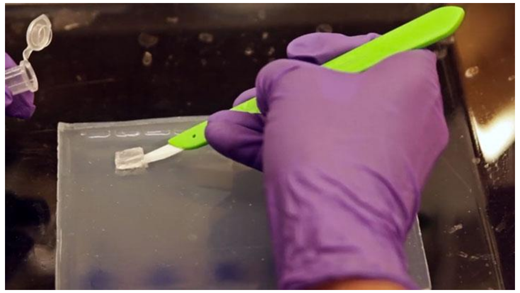


Image from NEB

Gel-extraction size selection

Illumina genomic library construction

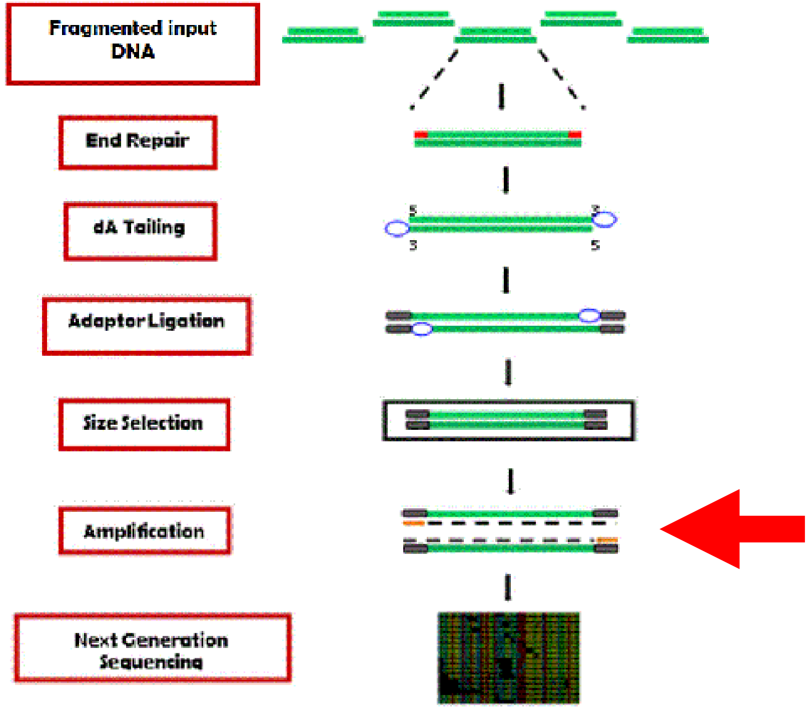


Image from ENZo Life Science

Limited PCR Amplification

PCR usually of 8-12 cycles

PCR Primers

TruSeq P5: AAT GAT ACG GCG ACC ACC GAG A

TruSeq P7: CAA GCA GAA GAC GGC ATA CGA G

Hi-Fidelity Polymerase

Illumina genomic library construction

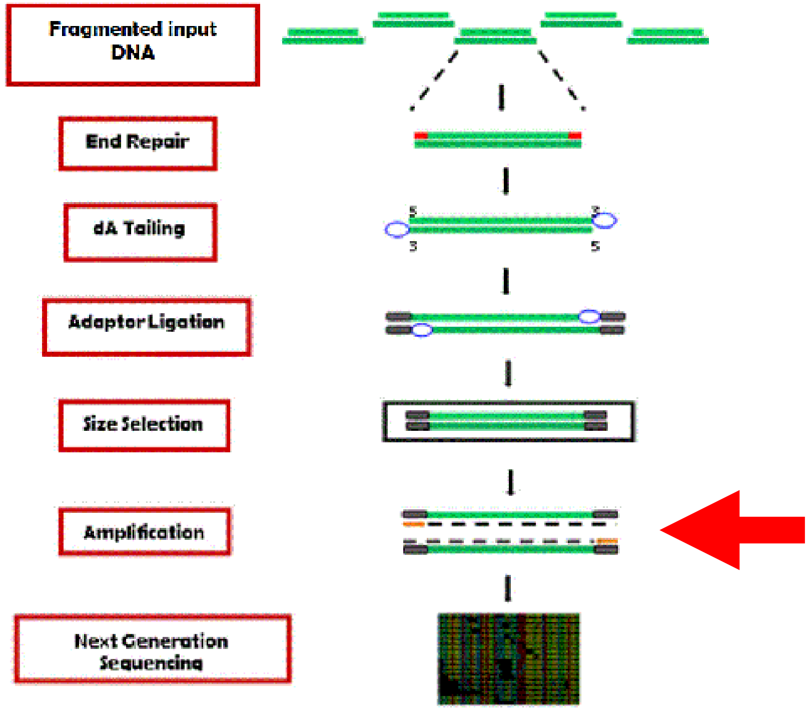


Image from ENZo Life Science

Limited PCR Amplification

PCR usually of 8-12 cycles

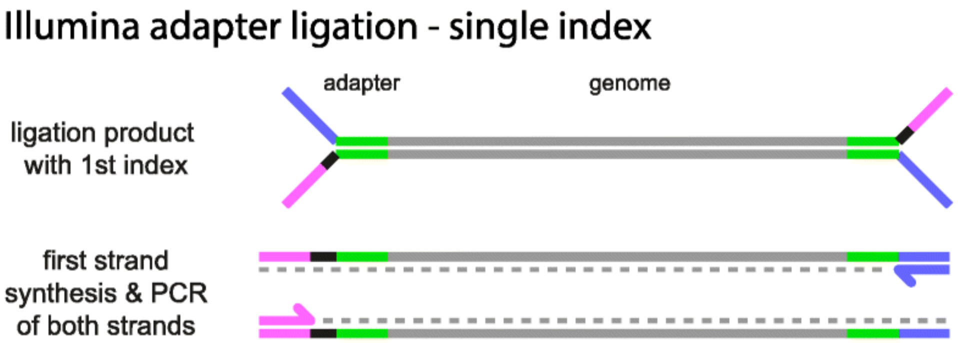

Illumina adapter ligation - single index
adapter
genome
ligation product with 1st index
first strand synthesis & PCR of both strands

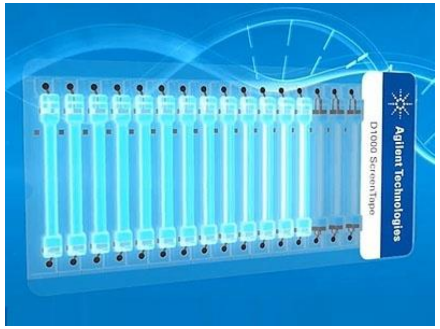
Image from Dawes et al. (2020) Mobile DNA

Now we should have double-stranded, blunt-ended libraries within the size range we selected

LIBRARY CONSTRUCTION COMPLETE!

Quantification of genomic DNA libraries

- Reasonably precise estimates of DNA concentration are needed for Illumina sequencer input

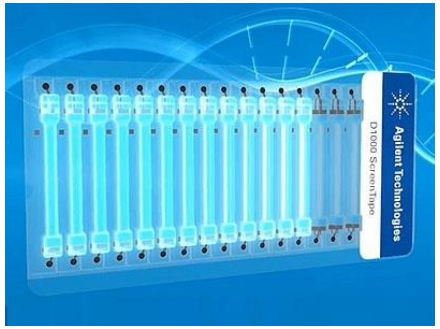


D1000 Screentape (Agilent)

TapeStation 2200 (Agilent)

Quantification of genomic DNA libraries

- Reasonably precise estimates of DNA concentration are needed for Illumina sequencer input



D1000 Screentape (Agilent)

TapeStation 2200 (Agilent)

Lower size standard

Upper size standard

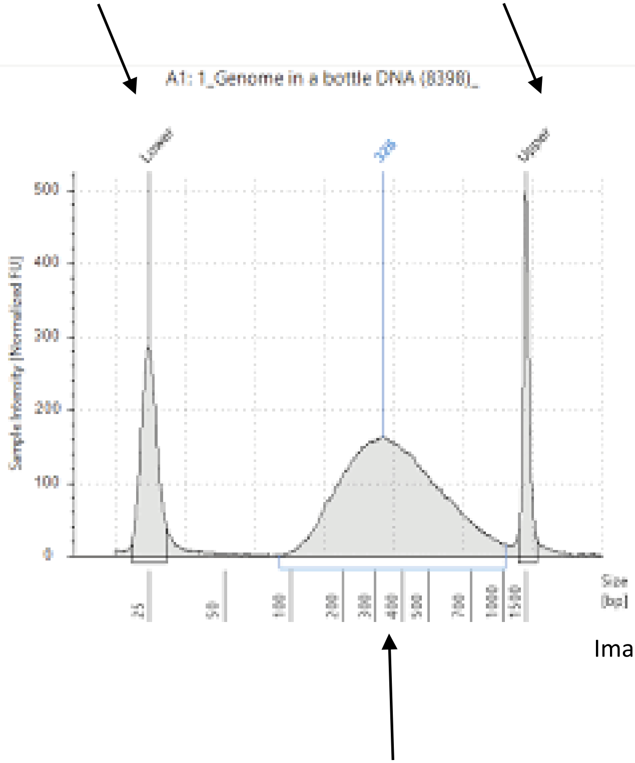
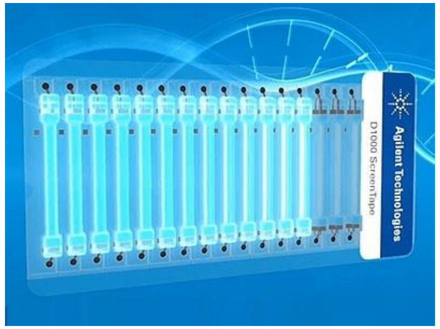


Image from Agilent

Illumina genomic library

Quantification of genomic DNA libraries

- Reasonably precise estimates of DNA concentration are needed for Illumina sequencer input



D1000 Screentape (Agilent)

TapeStation 2200 (Agilent)

Lower size standard

Upper size standard

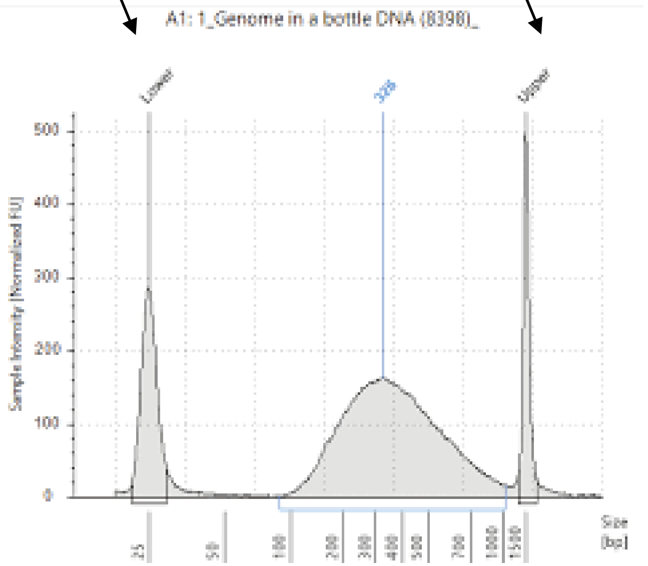
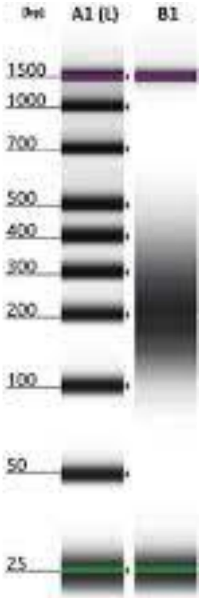


Image from Agilent

Illumina genomic library

"Gel Visualization"



USS

Illumina genomic library

LSS

Image from Agilent

Illumina sequencing

- Load genomic libraries into sequencer
- Cluster generation on a flow cell (aka bridge amplification)
- Clonal amplification
- ‘Sequencing-by-Synthesis’

The Illumina Flow Cell

HiSeq 3000 Flowcell
Image from illumina.com

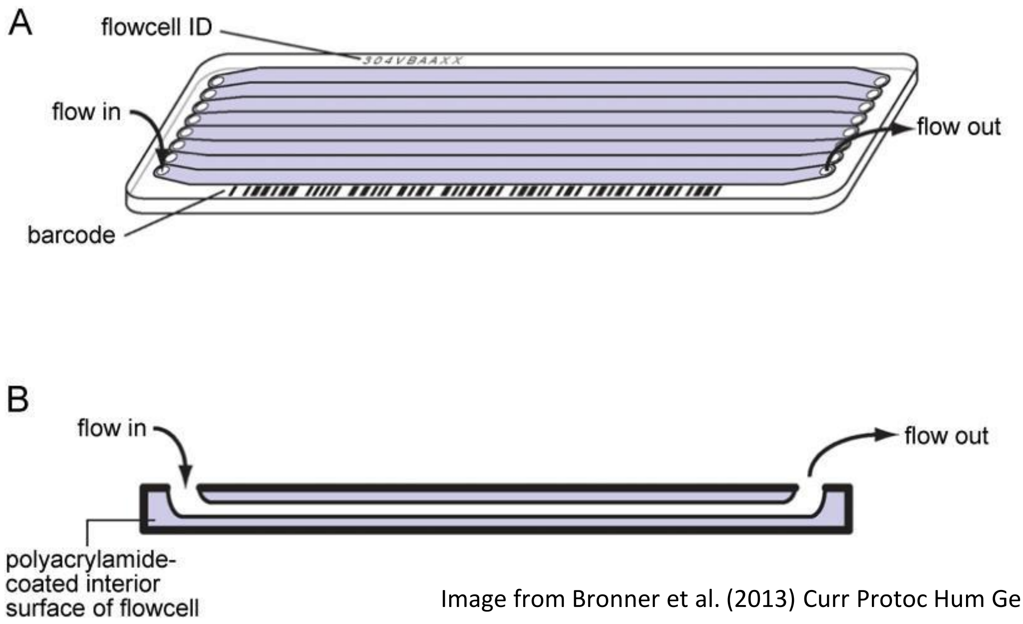


Image from Bronner et al. (2013) Curr Protoc Hum Genet.

The Illumina Flow Cell

HiSeq 3000 Flowcell
Image from illumina.com

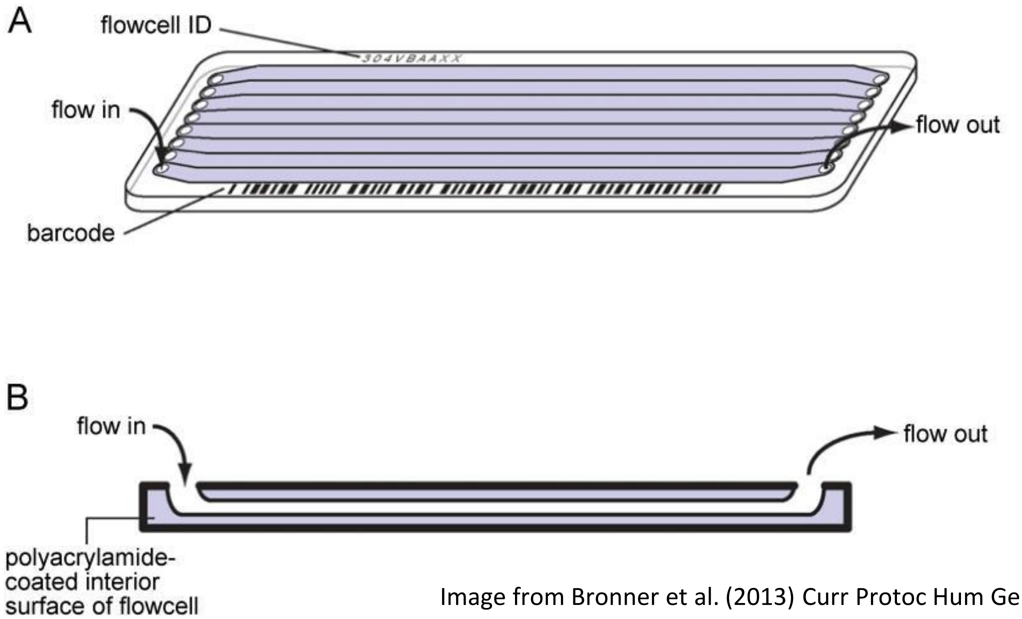


Image from Bronner et al. (2013) Curr Protoc Hum Genet.

🟩 P5 (i5) oligo

🟥 P7 (i7) oligo

P5: AATGATACGGCGACCACCGAGA
P7: CAAGCAGAAGACGGCATACGAG

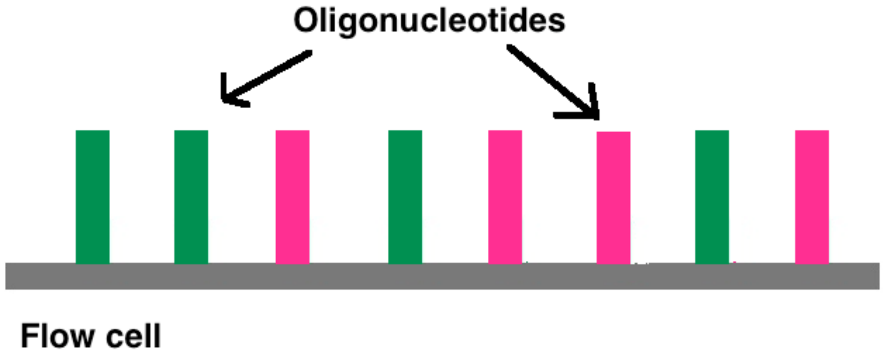
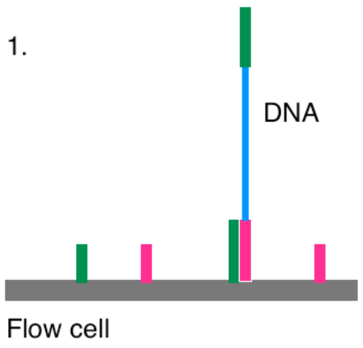


Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

Bridge Amplification + Cluster Generation + Clonal Amplification



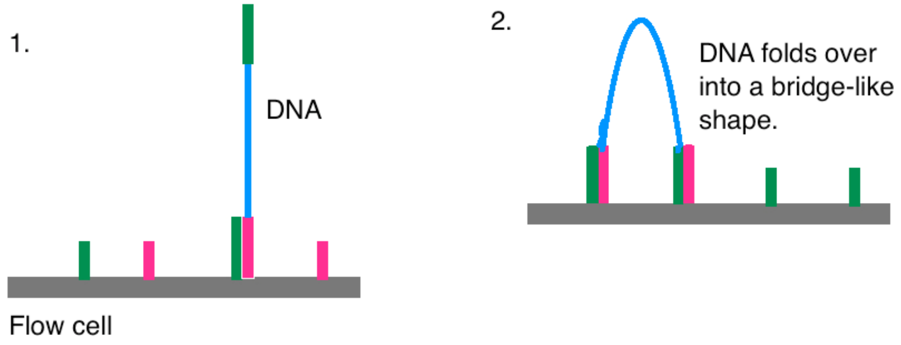
🟩 P5 (i5) oligo

🟥 P7 (i7) oligo

P5: AATGATACGGCGACCACCGAGA
P7: CAAGCAGAAGACGGCATACGAG

Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

Bridge Amplification + Cluster Generation + Clonal Amplification



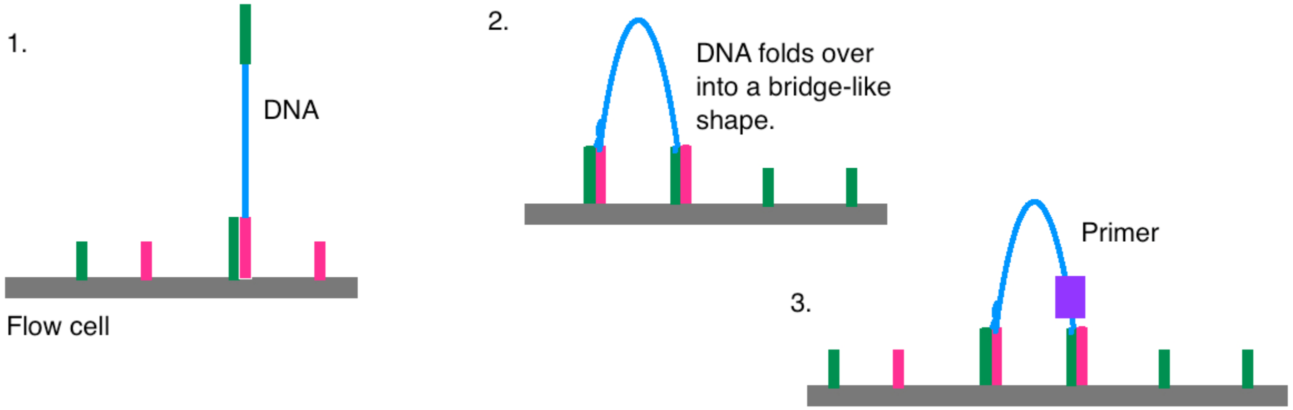
🟩 P5 (i5) oligo

🟥 P7 (i7) oligo

P5: AATGATACGGCGACCACCGAGA
P7: CAAGCAGAAGACGGCATACGAG

Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

Bridge Amplification + Cluster Generation + Clonal Amplification



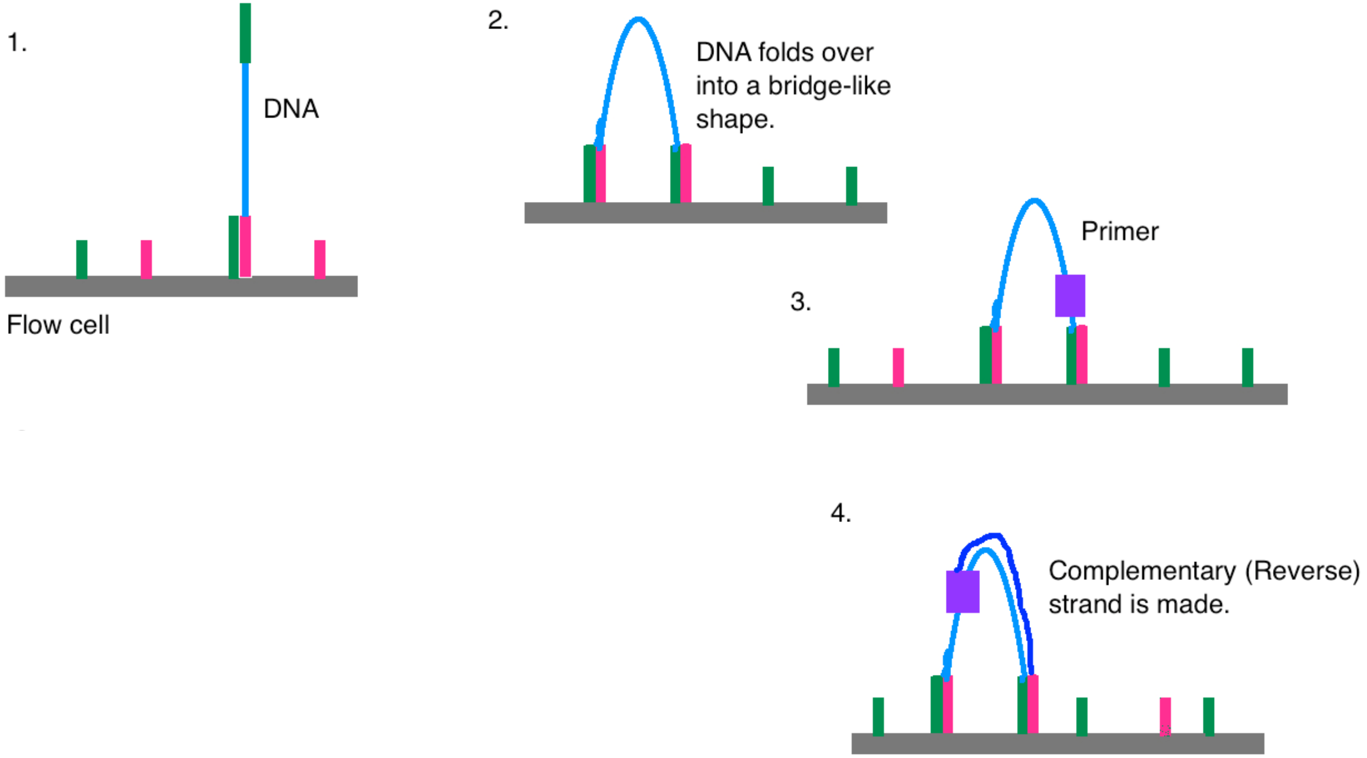
🟩 P5 (i5) oligo

🟥 P7 (i7) oligo

P5: AATGATACGGCGACCACCGAGA
P7: CAAGCAGAAGACGGCATACGAG

Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

Bridge Amplification + Cluster Generation + Clonal Amplification


1.
DNA
Flow cell
2.
DNA folds over into a bridge-like shape.
3.
Primer
4.
Complementary (Reverse) strand is made.

🟩 P5 (i5) oligo

🟥 P7 (i7) oligo

P5: AATGATACGGCGACCACCGAGA

P7: CAAGCAGAAGACGGCATACGAG

Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

Bridge Amplification + Cluster Generation + Clonal Amplification

🟩 P5 (i5) oligo

🟥 P7 (i7) oligo

P5: AATGATACGGCGACCACCGAGA

P7: CAAGCAGAAGACGGCATACGAG

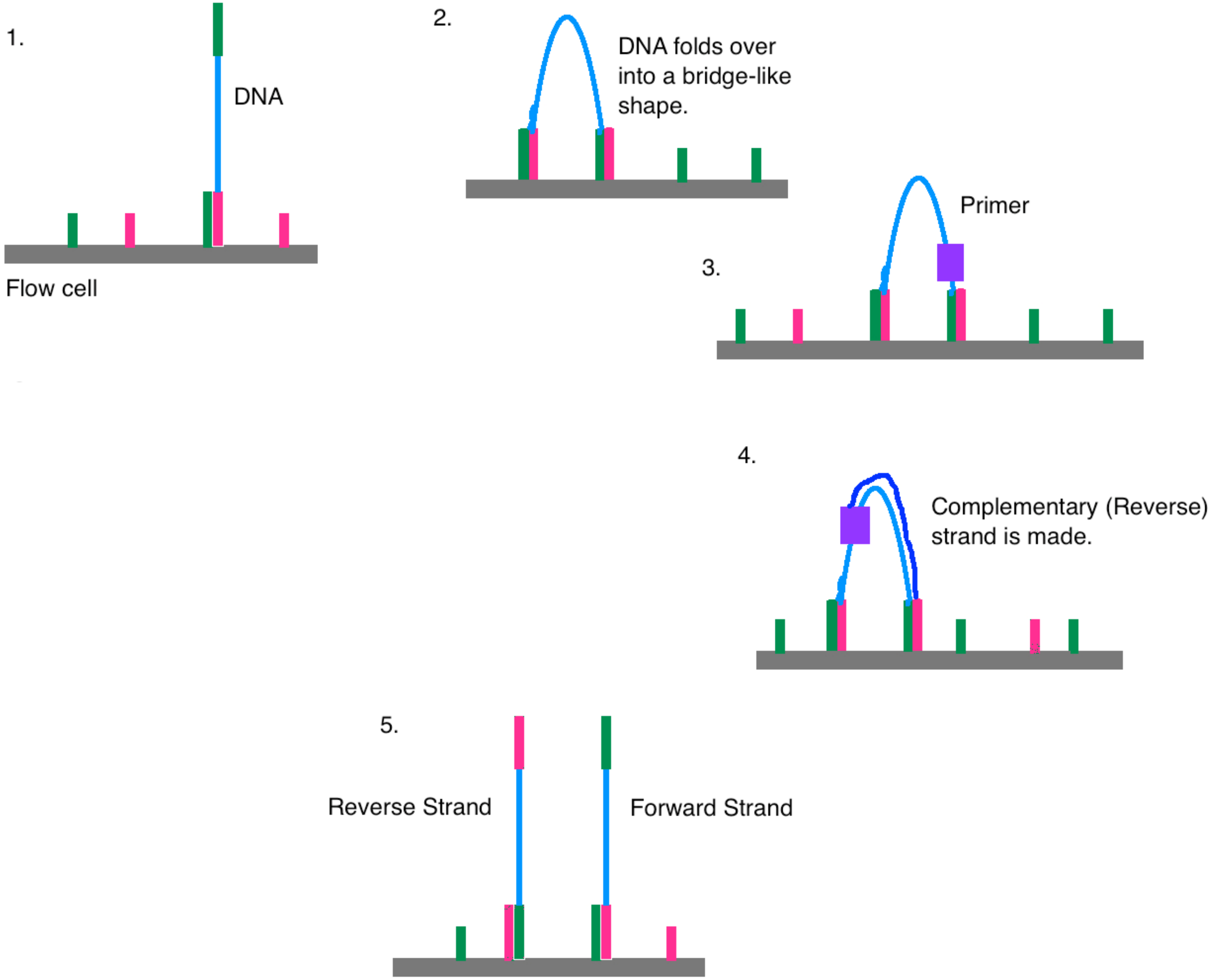
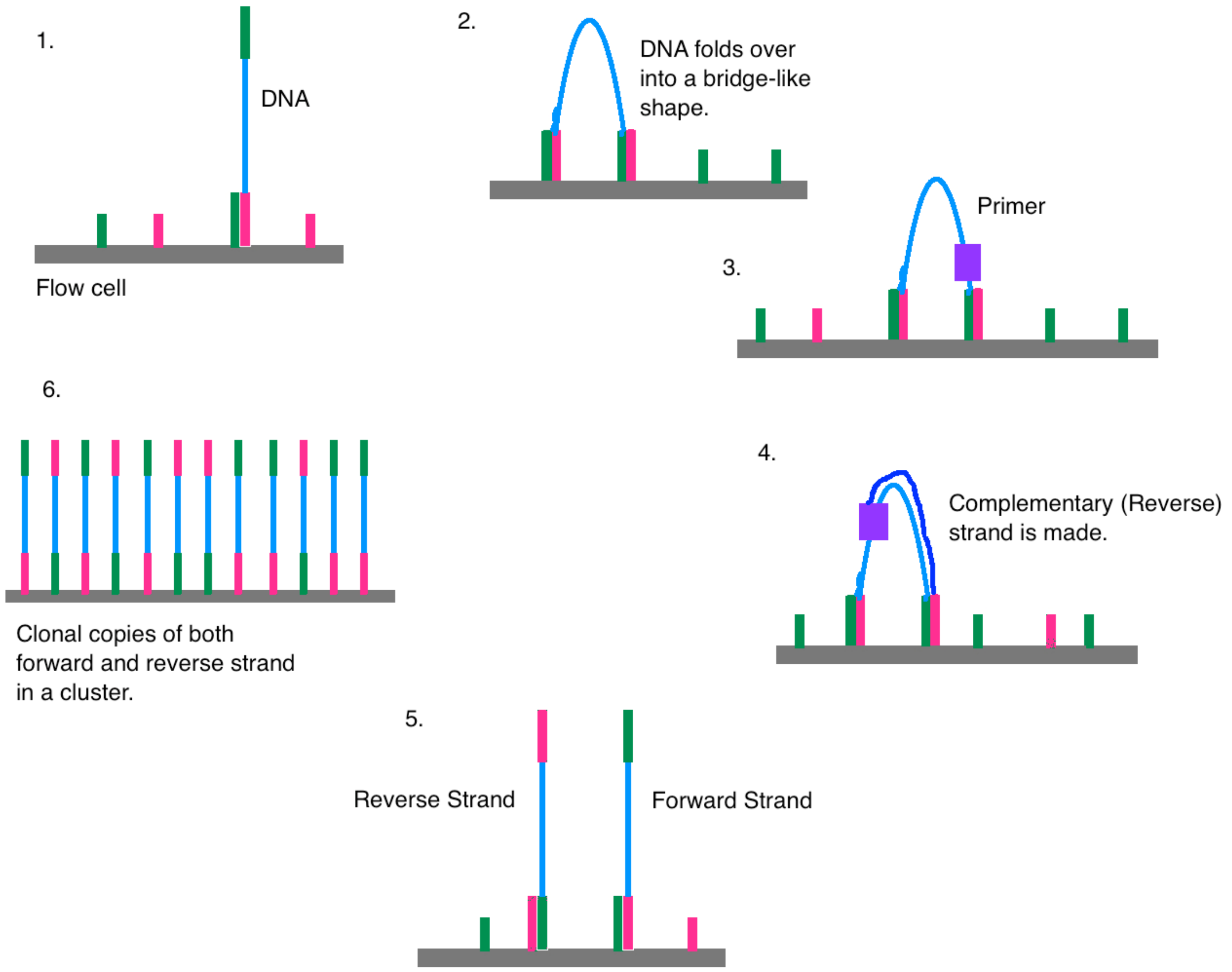

1.
DNA
Flow cell
2.
DNA folds over into a bridge-like shape.
3.
Primer
4.
Complementary (Reverse) strand is made.
5.
Reverse Strand
Forward Strand

Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

Bridge Amplification + Cluster Generation + Clonal Amplification



1.

DNA

Flow cell

2.

DNA folds over into a bridge-like shape.

3.

Primer

4.

Complementary (Reverse) strand is made.

5.

Reverse Strand

Forward Strand

6.

Clonal copies of both forward and reverse strand in a cluster.

🟩 P5 (i5) oligo

🟥 P7 (i7) oligo

P5: AATGATACGGCGACCACCGAGA

P7: CAAGCAGAAGACGGCATACGAG

Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

Sequencing-by-synthesis (SBS)

Paired-End versus Single-End Reads Illumina sequencing

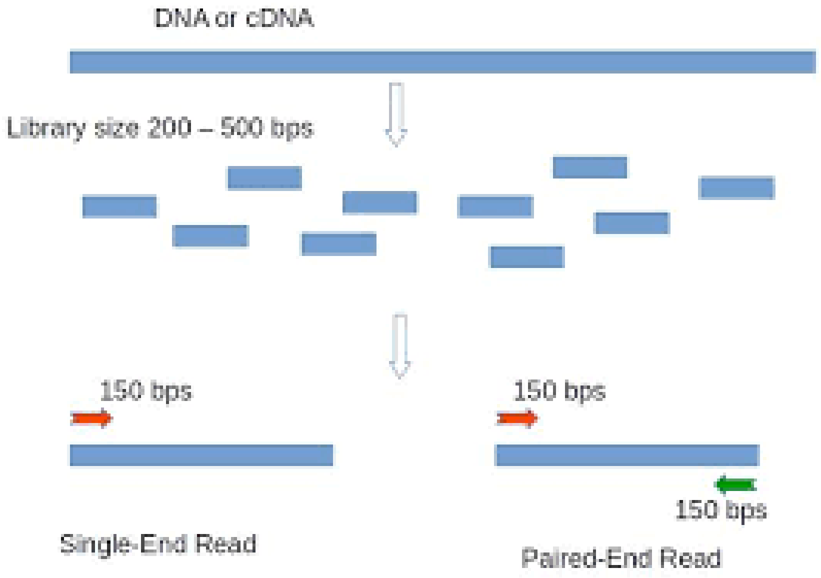


Image by 1010 Genome

££

££££

- One or both sides of the fragments can be sequenced using SBS
- While there are some advantages to Single-End sequencing, most of the applications we use will benefit most from Paired-End sequencing

Sequencing-by-synthesis (SBS)

Single Read **OR** Paired-End Read – Forward Strand

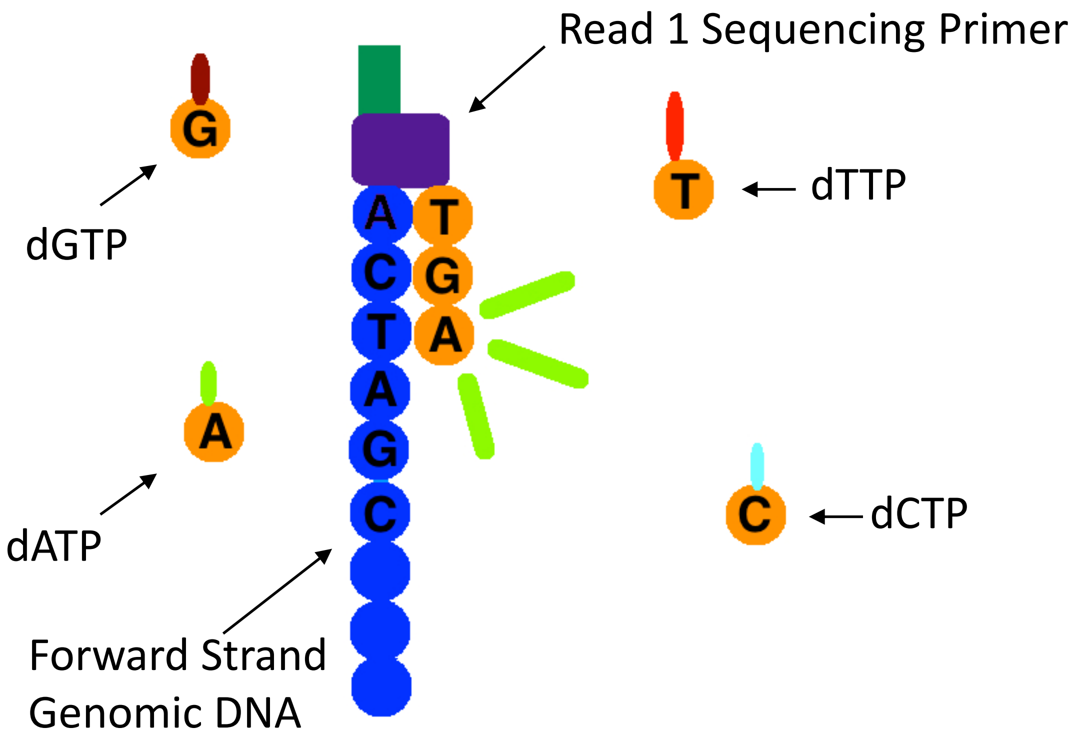


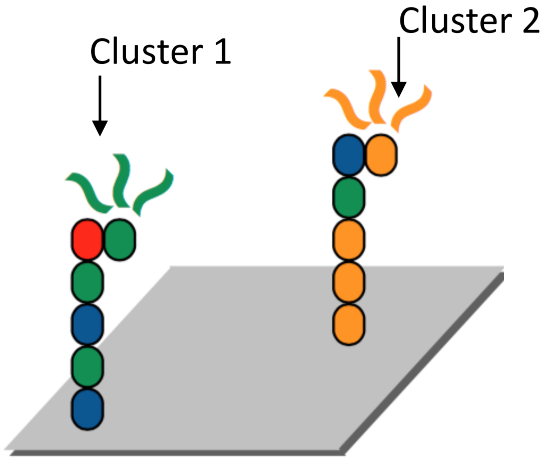
Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license

At the end of clonal amplification, all of the reverse strands are washed off the flow cell, leaving only forward strands. A primer attaches to the forward strands adapter primer binding site, and a polymerase adds a fluorescently tagged dNTP to the DNA strand. Only one base can be added per round due to the fluorophore acting as a blocking group; however, the blocking group is reversible.

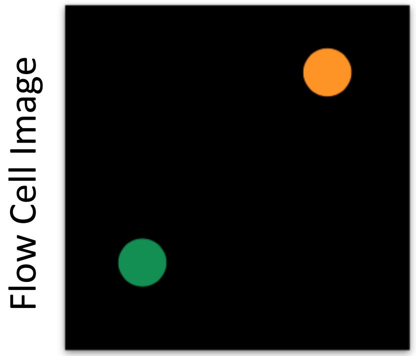
Using the four-color chemistry*, each of the four bases has a unique emission, and after each round, the machine records which base was added. Once the colour is recorded the fluorophore is washed away and another dNTP is washed over the flow cell and the process is repeated.

Text *mostly* from Wikipedia ☺

Sequencing-by-synthesis (SBS)



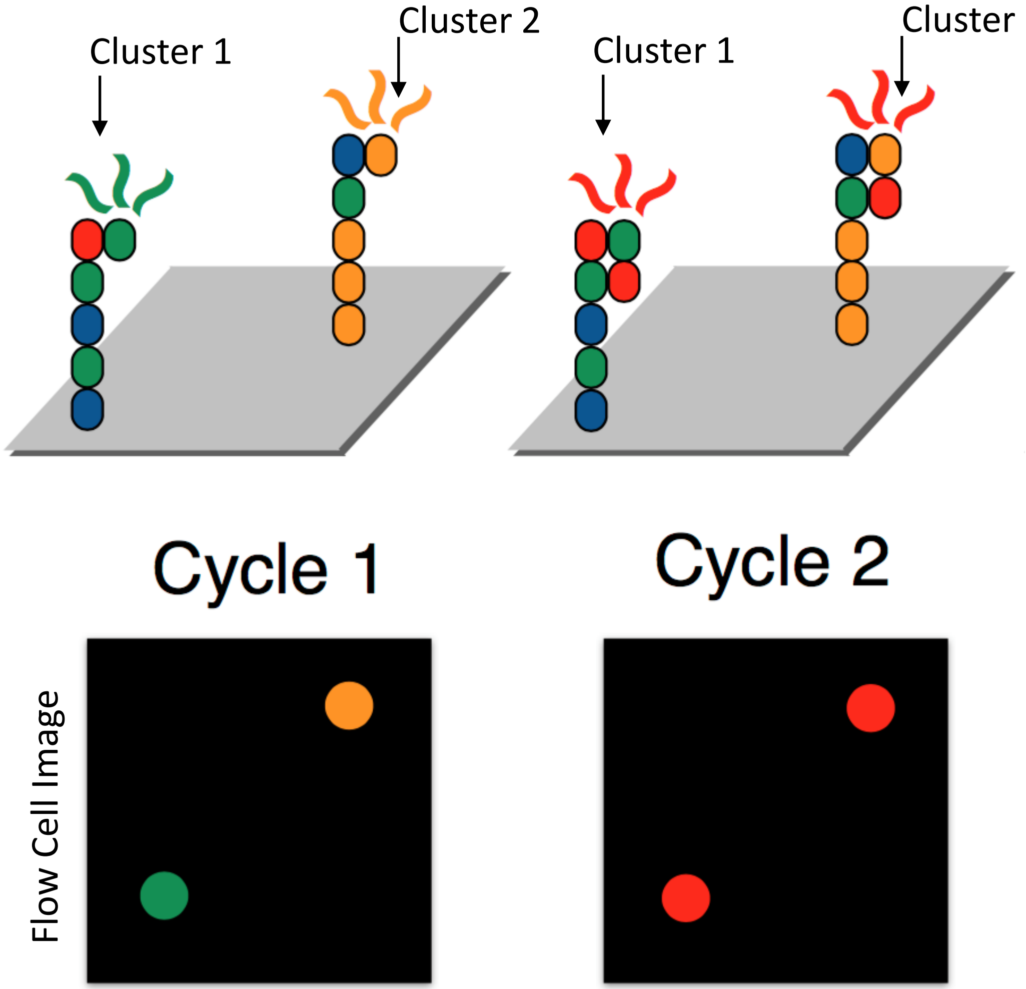
Cycle 1



Cluster 1 Sequence: T
Cluster 2 Sequence: C

Image by data-science-sequencing-github.io

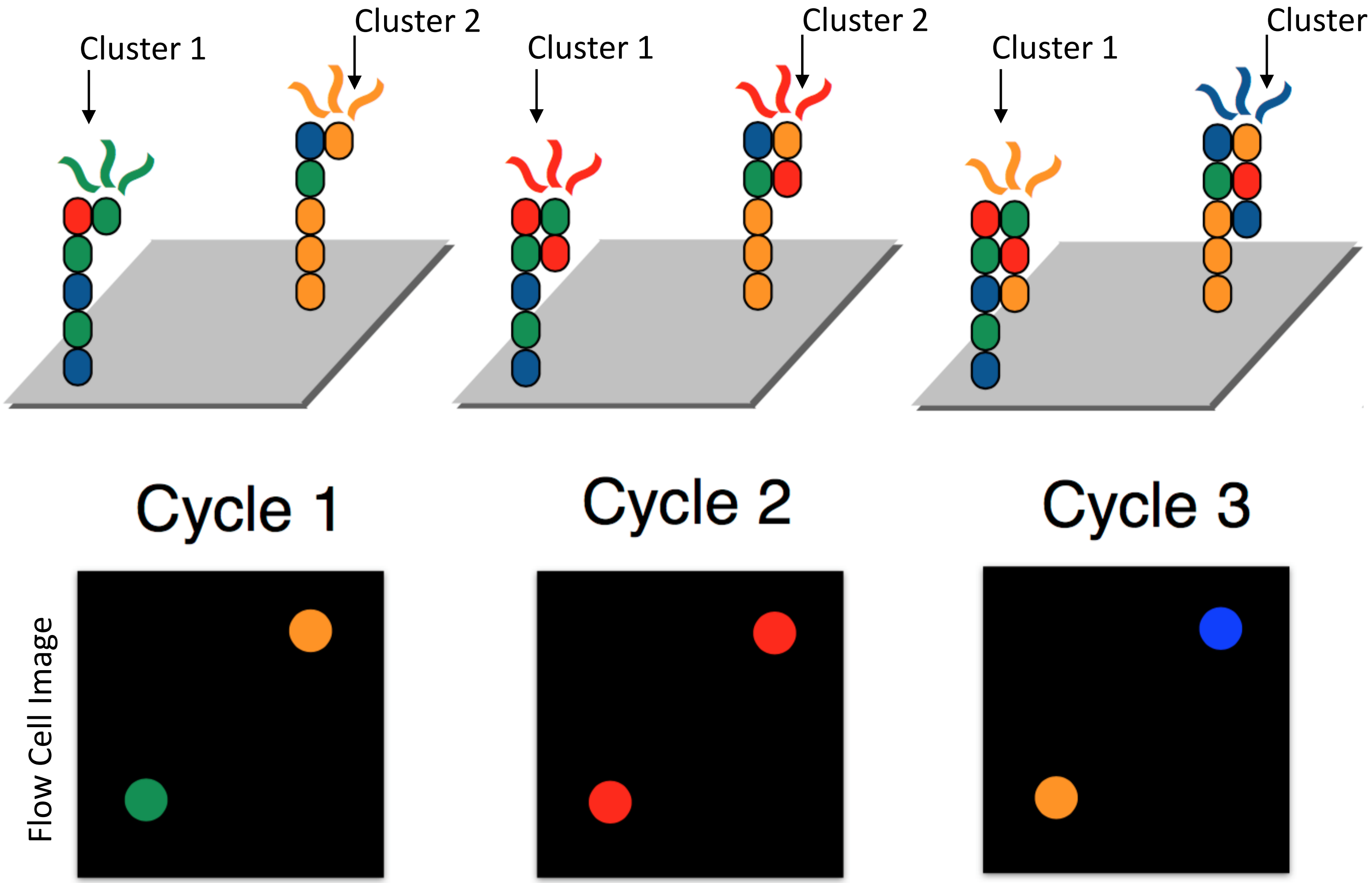
Sequencing-by-synthesis (SBS)



Cluster 1 Sequence: T A
Cluster 2 Sequence: C A

Image by data-science-sequencing-github.io

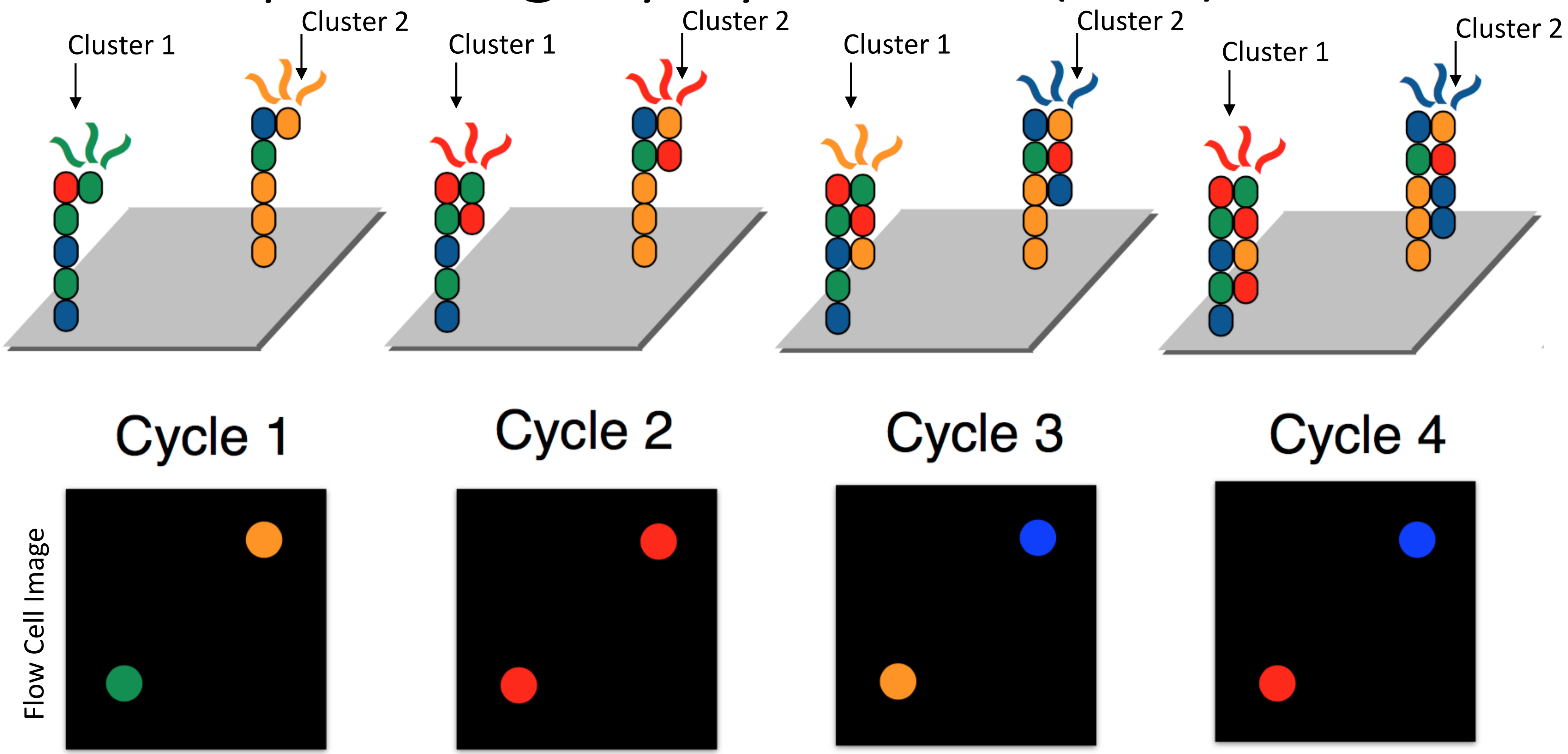
Sequencing-by-synthesis (SBS)



Cluster 1 Sequence: T A C
Cluster 2 Sequence: C A G

Image by data-science-sequencing-github.io

Sequencing-by-synthesis (SBS)

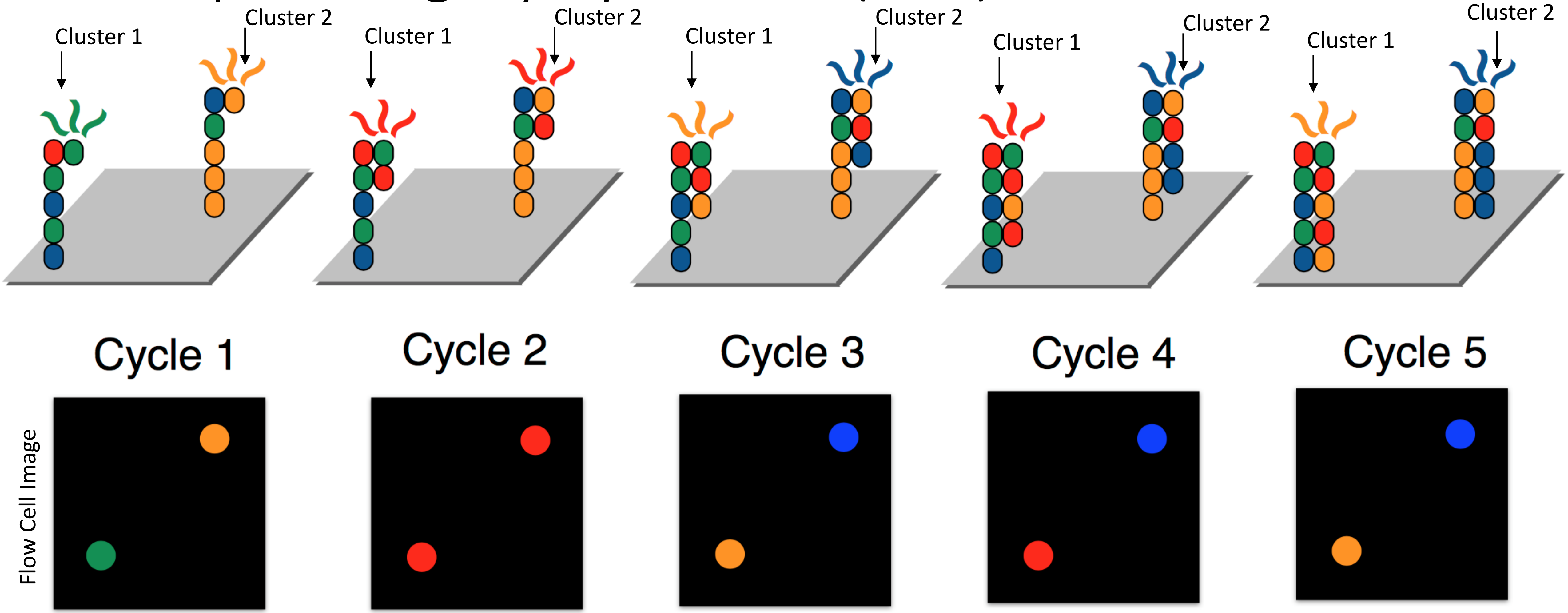


Cluster 1 Sequence: T A C A

Cluster 2 Sequence: C A G G

Image by data-science-sequencing-github.io

Sequencing-by-synthesis (SBS)



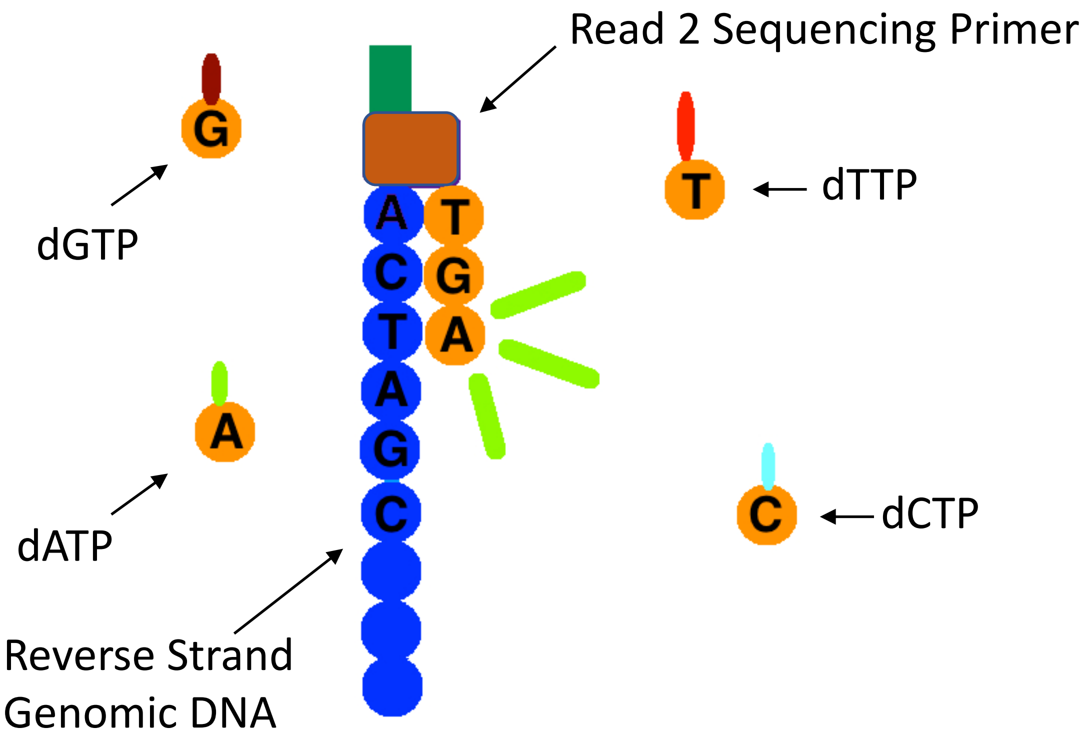
Cluster 1 Sequence: T A C A C

Cluster 2 Sequence: C A G G G

Image by data-science-sequencing-github.io

Sequencing-by-synthesis (SBS)

Paired-End Read – Reverse Strand

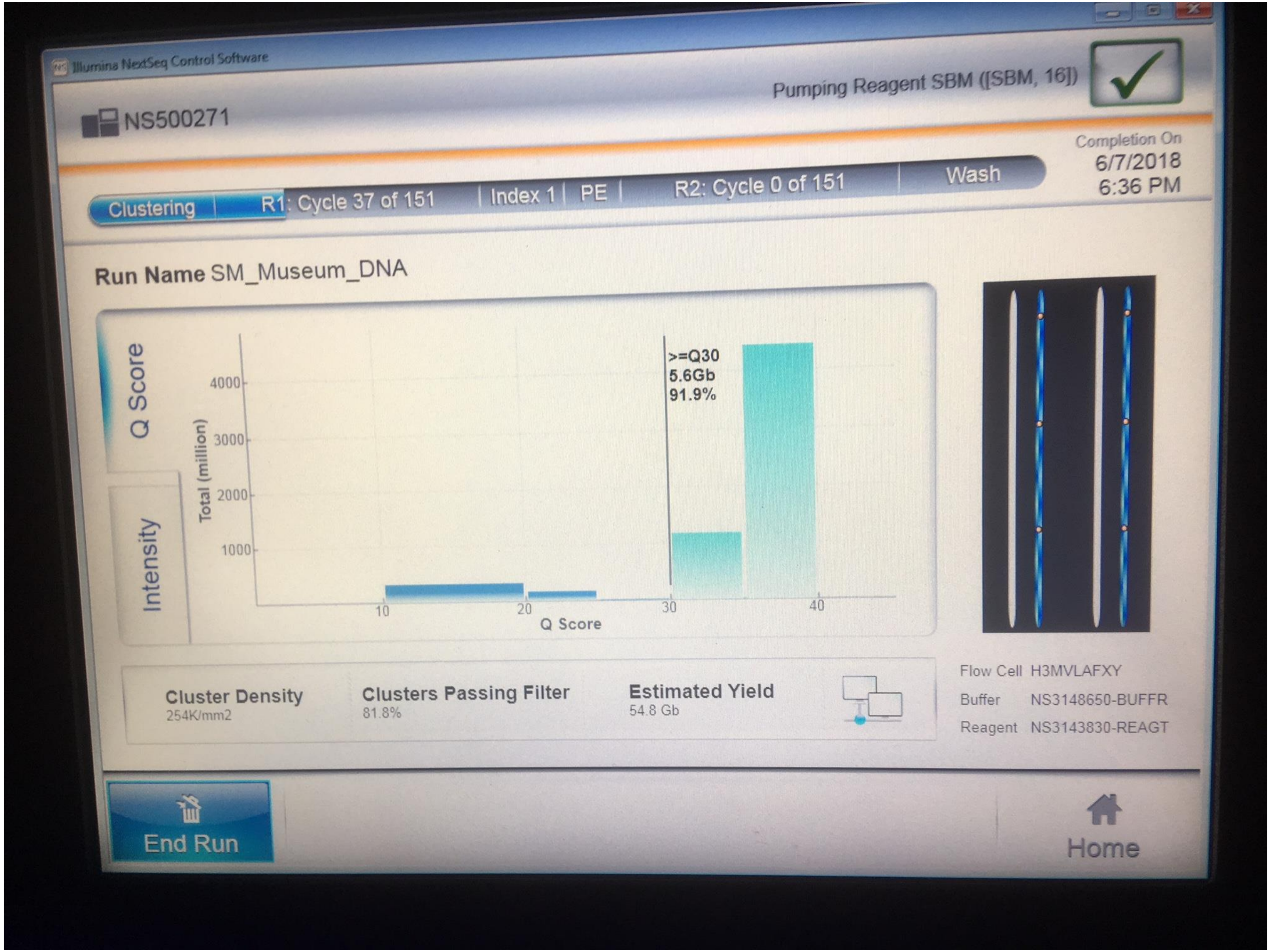


Once the Read 1 DNA strand has been read, the strand that was just added is washed away. Then, the index 1 primer attaches, polymerizes the index 1 sequence, and is washed away. The strand forms a bridge again, and the 3' end of the DNA strand attaches to an oligo on the flow cell. The index 2 primer attaches, polymerizes the sequence, and is washed away.

A polymerase sequences the complementary strand on top of the arched strand. They separate, and the 3' end of each strand is blocked. The forward strand is washed away, and the process of sequence by synthesis repeats for the reverse strand.

Text *mostly* from Wikipedia ☺

Image by D.M. Lapato used under a Creative Commons Attribution-Share Alike 4.0 International license



Illumina NextSeq Control Software

NS500271

Pumping Reagent SBM ([SBM, 16])

Clustering | R1: Cycle 37 of 151 | Index 1 | PE | R2: Cycle 0 of 151 | Wash

Completion On
6/7/2018
6:36 PM

Run Name SM_Museum_DNA

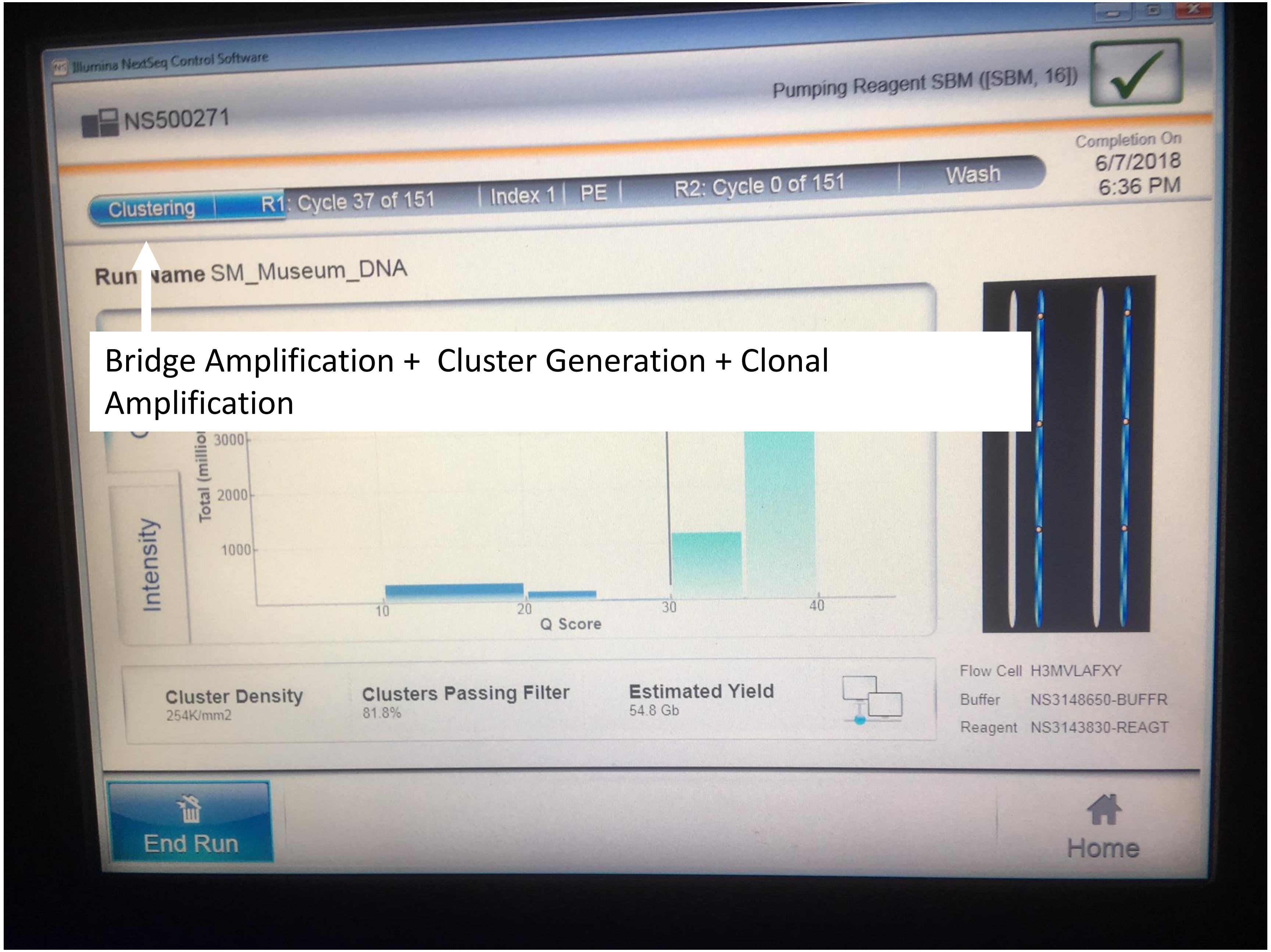
Cluster Density
254K/mm2

Clusters Passing Filter
81.8%

Estimated Yield
54.8 Gb

Flow Cell H3MVLAFXY
Buffer NS3148650-BUFFR
Reagent NS3143830-REAGT

NextSeq 500 @ NHM



Illumina NextSeq Control Software

NS500271

Pumping Reagent SBM ([SBM, 16])

Completion On 6/7/2018 6:36 PM

Run Name SM_Museum_DNA

Cluster Density 254K/mm2

Clusters Passing Filter 81.8%

Estimated Yield 54.8 Gb

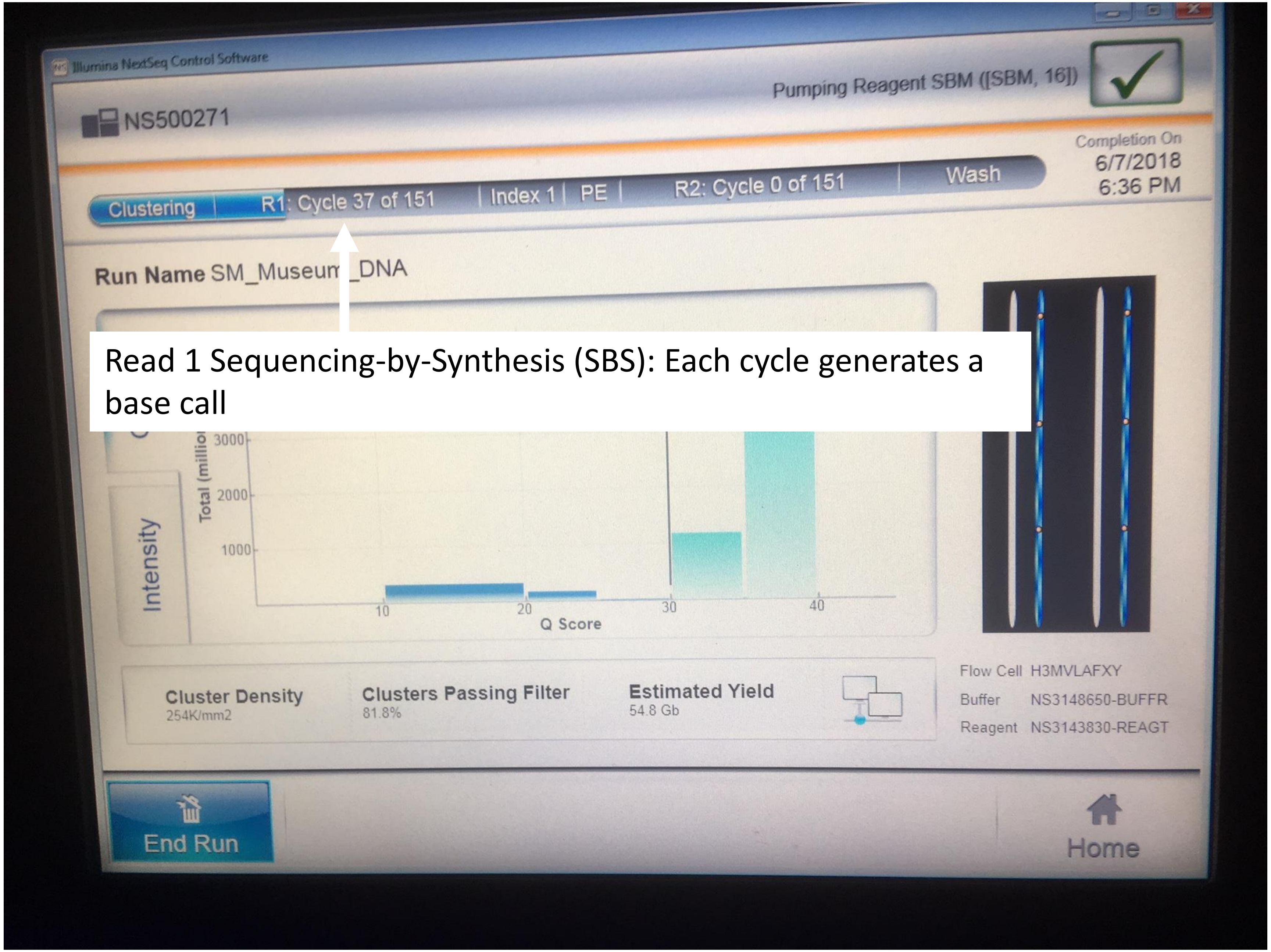
Flow Cell H3MVLAFXY

Buffer NS3148650-BUFFR

Reagent NS3143830-REAGT

Bridge Amplification + Cluster Generation + Clonal Amplification

NextSeq 500 @ NHM



Illumina NextSeq Control Software

NS500271

Pumping Reagent SBM ([SBM, 16])

Completion On
6/7/2018
6:36 PM

Run Name SM_Museum_DNA

Read 1 Sequencing-by-Synthesis (SBS): Each cycle generates a base call

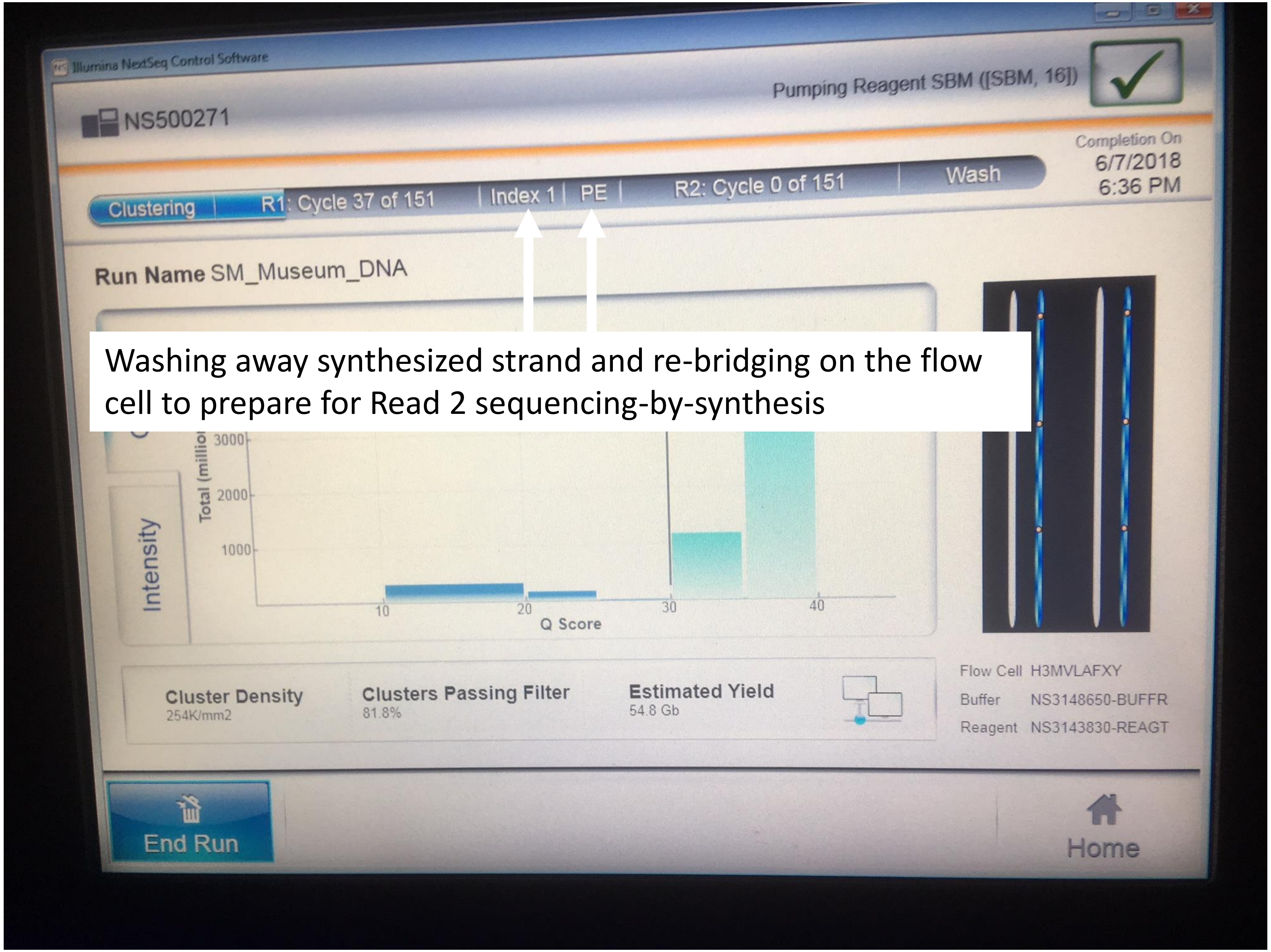
Cluster Density
254K/mm2

Clusters Passing Filter
81.8%

Estimated Yield
54.8 Gb

Flow Cell H3MVLAFXY
Buffer NS3148650-BUFFR
Reagent NS3143830-REAGT

NextSeq 500 @ NHM



Illumina NextSeq Control Software

NS500271

Pumping Reagent SBM ([SBM, 16])

Completion On
6/7/2018
6:36 PM

Run Name SM_Museum_DNA

Washing away synthesized strand and re-bridging on the flow cell to prepare for Read 2 sequencing-by-synthesis

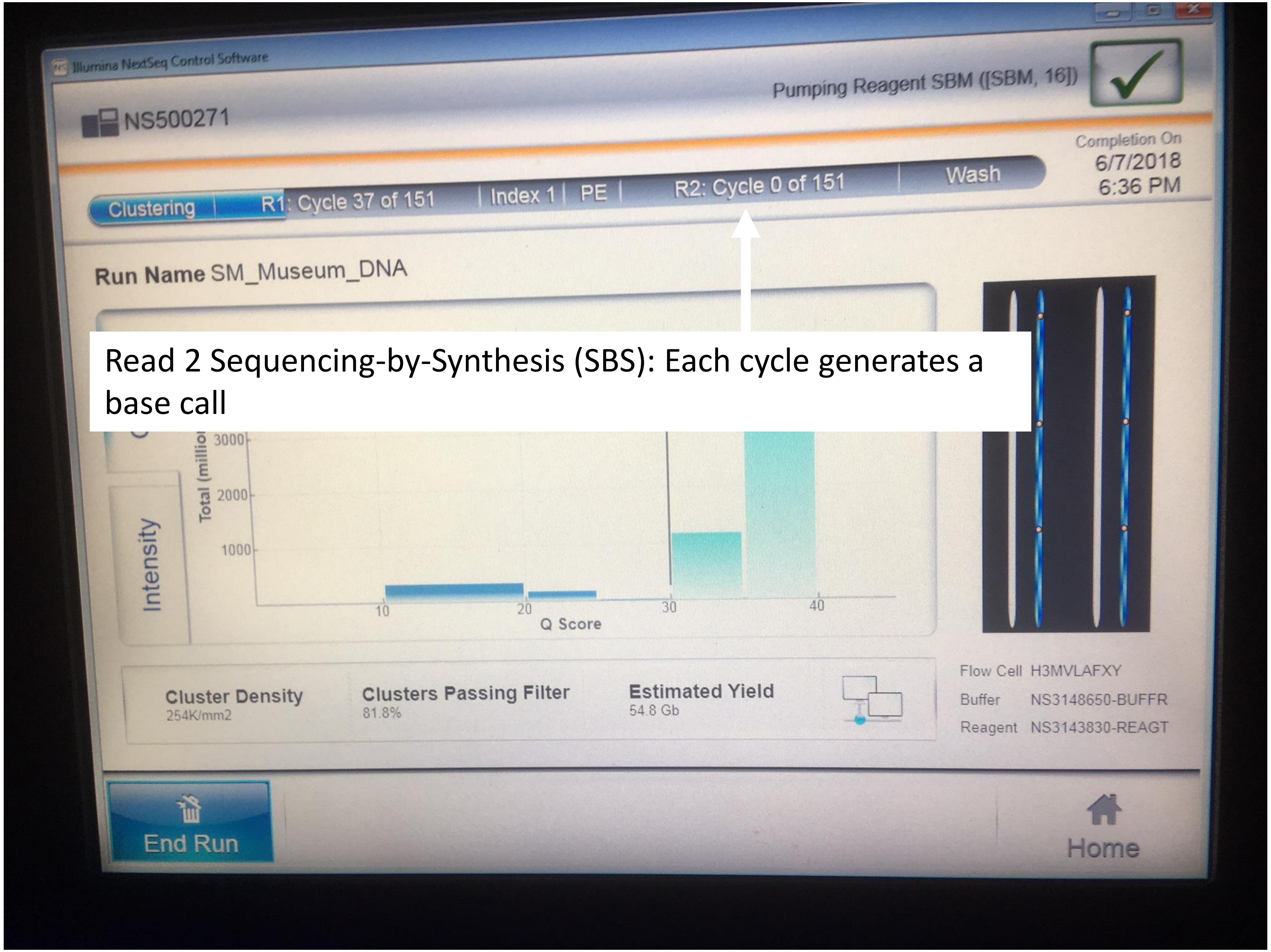
Cluster Density
254K/mm2

Clusters Passing Filter
81.8%

Estimated Yield
54.8 Gb

Flow Cell H3MVLAFXY
Buffer NS3148650-BUFFR
Reagent NS3143830-REAGT

NextSeq 500 @ NHM



Illumina NextSeq Control Software

NS500271

Pumping Reagent SBM ([SBM, 16])

Completion On
6/7/2018
6:36 PM

Run Name SM_Museum_DNA

Cluster Density 254K/mm2

Clusters Passing Filter 81.8%

Estimated Yield 54.8 Gb

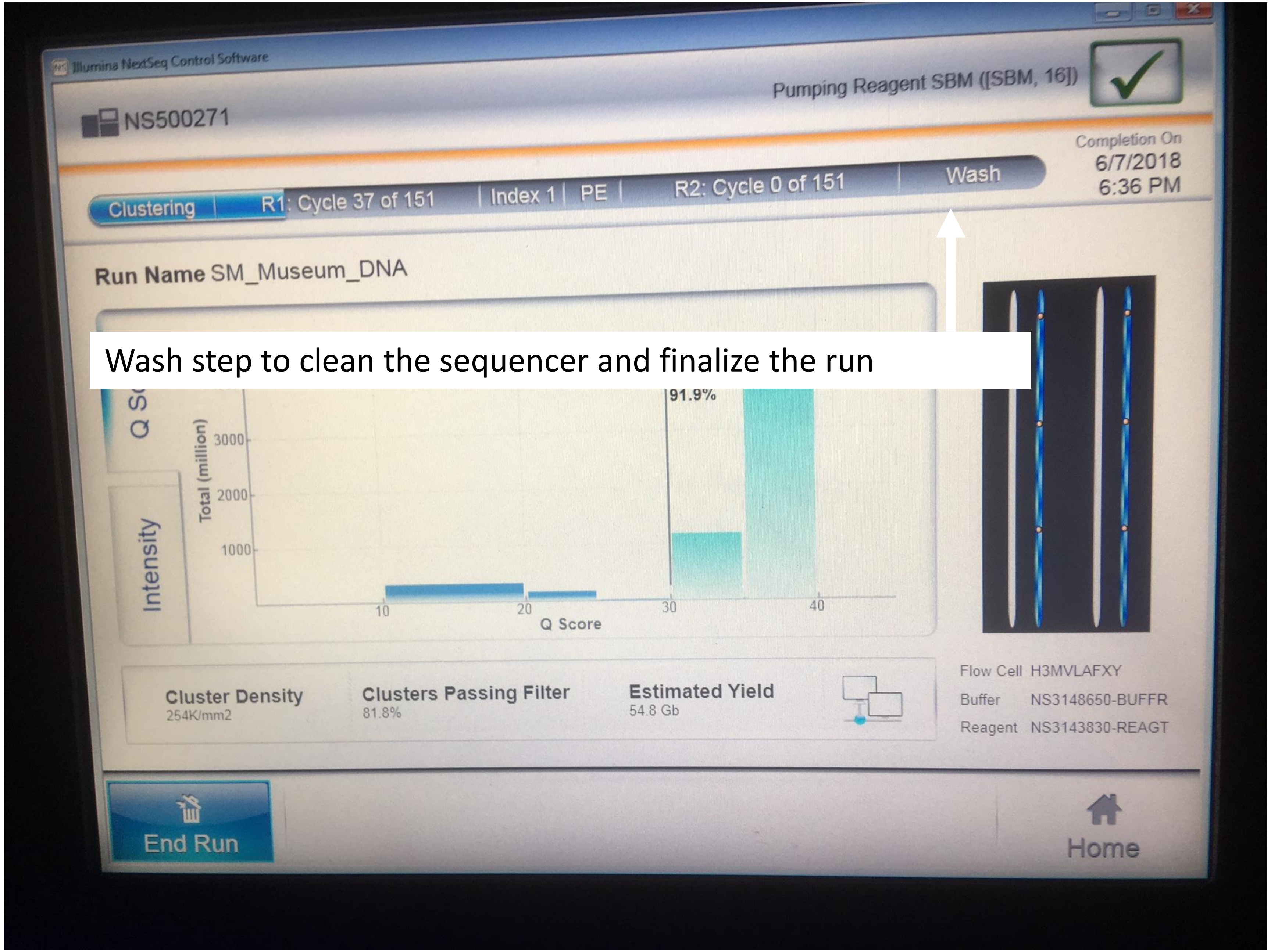
Flow Cell H3MVLAFXY

Buffer NS3148650-BUFFR

Reagent NS3143830-REAGT

Read 2 Sequencing-by-Synthesis (SBS): Each cycle generates a base call

NextSeq 500 @ NHM



Illumina NextSeq Control Software

NS500271

Pumping Reagent SBM ([SBM, 16])

Clustering | R1: Cycle 37 of 151 | Index 1 | PE | R2: Cycle 0 of 151 | Wash

Completion On
6/7/2018
6:36 PM

Run Name SM_Museum_DNA

Cluster Density 254K/mm2

Clusters Passing Filter 81.8%

Estimated Yield 54.8 Gb

Flow Cell H3MVLAFXY

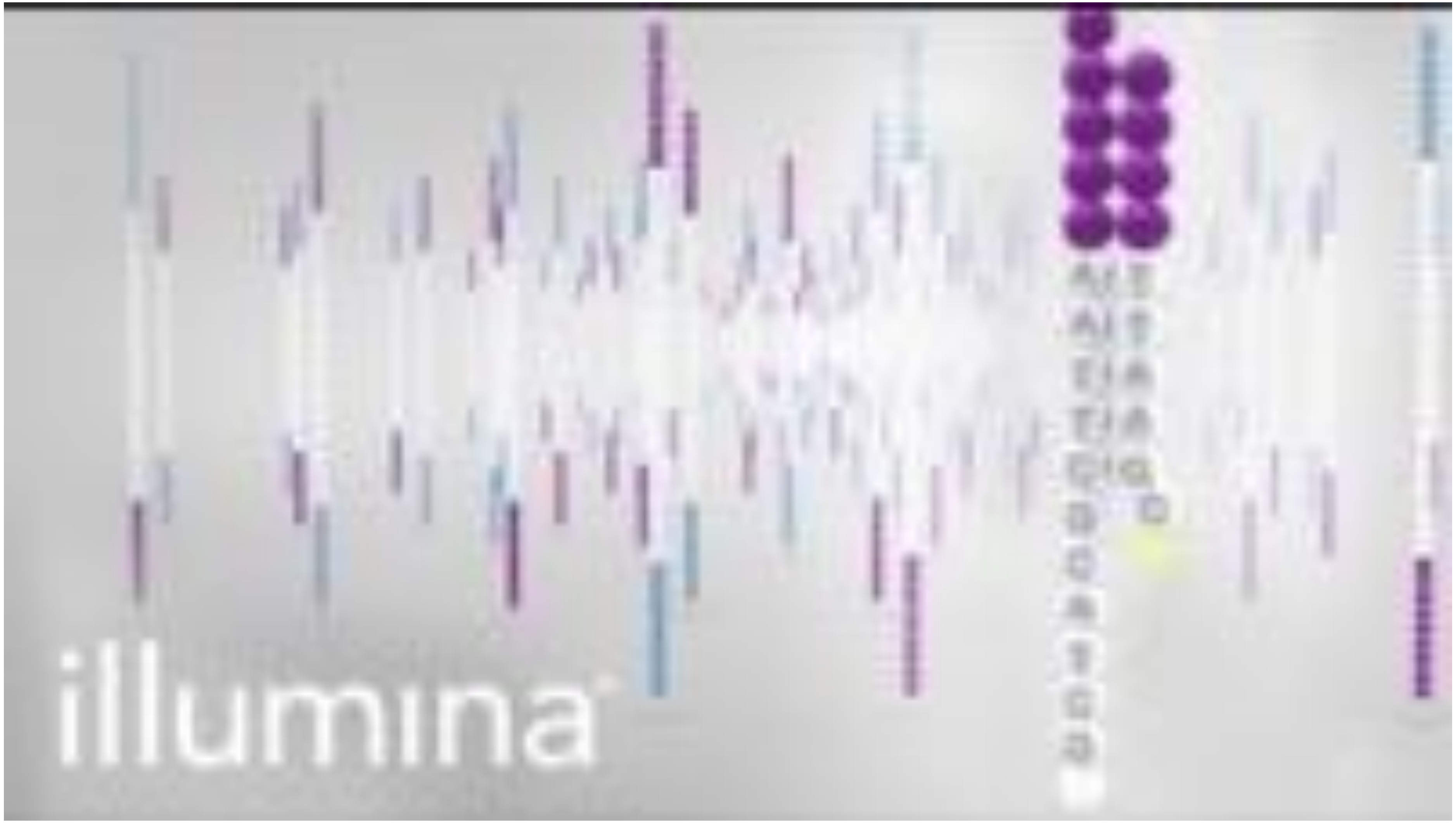
Buffer NS3148650-BUFFR

Reagent NS3143830-REAGT

Wash step to clean the sequencer and finalize the run

NextSeq 500 @ NHM

Animated explanation of SBS



Sequencing Power for Every Scale

The broadest portfolio offering available

Sequencing System	iSeq™	MiniSeq™	MiSeq®	NextSeq®	HiSeq®	HiSeq® X	NovaSeq®
					4000	Five/Ten	6000
Output per run	1.2 Gb	7.5 Gb	15 Gb	120 Gb	1.5 Tb	1.8 Tb	1 Tb - 6 Tb[1]
Instrument price	$19.9K	$49.5K	$99K	$275K	$900K	$6M[2]/$10M[2]	$985K
Installed base[3]	NA	~600	~6,000	~2,400	~2,300[4]		~285

1. Output per run for the S1, S2 and S4 flow cells equal 1 Tb, 2 Tb and 6 Tb, respectively assuming two flow cells per run
2. Based on purchase of 5 and 10 units for HiSeq X Five and HiSeq X Ten, respectively
3. Based on end of fiscal year 2017
4. Combined HiSeq family

Sequencing Power for Every Scale

The broadest portfolio offering available

@NHM ↓

@NHM ↓

Sequencing System	iSeq™	MiniSeq™	MiSeq®	NextSeq®	HiSeq®	HiSeq® X	NovaSeq®
					4000	Five/Ten	6000
Output per run	1.2 Gb	7.5 Gb	15 Gb	120 Gb	1.5 Tb	1.8 Tb	1 Tb - 6 Tb[1]
Instrument price	$19.9K	$49.5K	$99K	$275K	$900K	$6M[2]/$10M[2]	$985K
Installed base[3]	NA	~600	~6,000	~2,400	~2,300[4]		~285

1. Output per run for the S1, S2 and S4 flow cells equal 1 Tb, 2 Tb and 6 Tb, respectively assuming two flow cells per run
2. Based on purchase of 5 and 10 units for HiSeq X Five and HiSeq X Ten, respectively
3. Based on end of fiscal year 2017
4. Combined HiSeq family

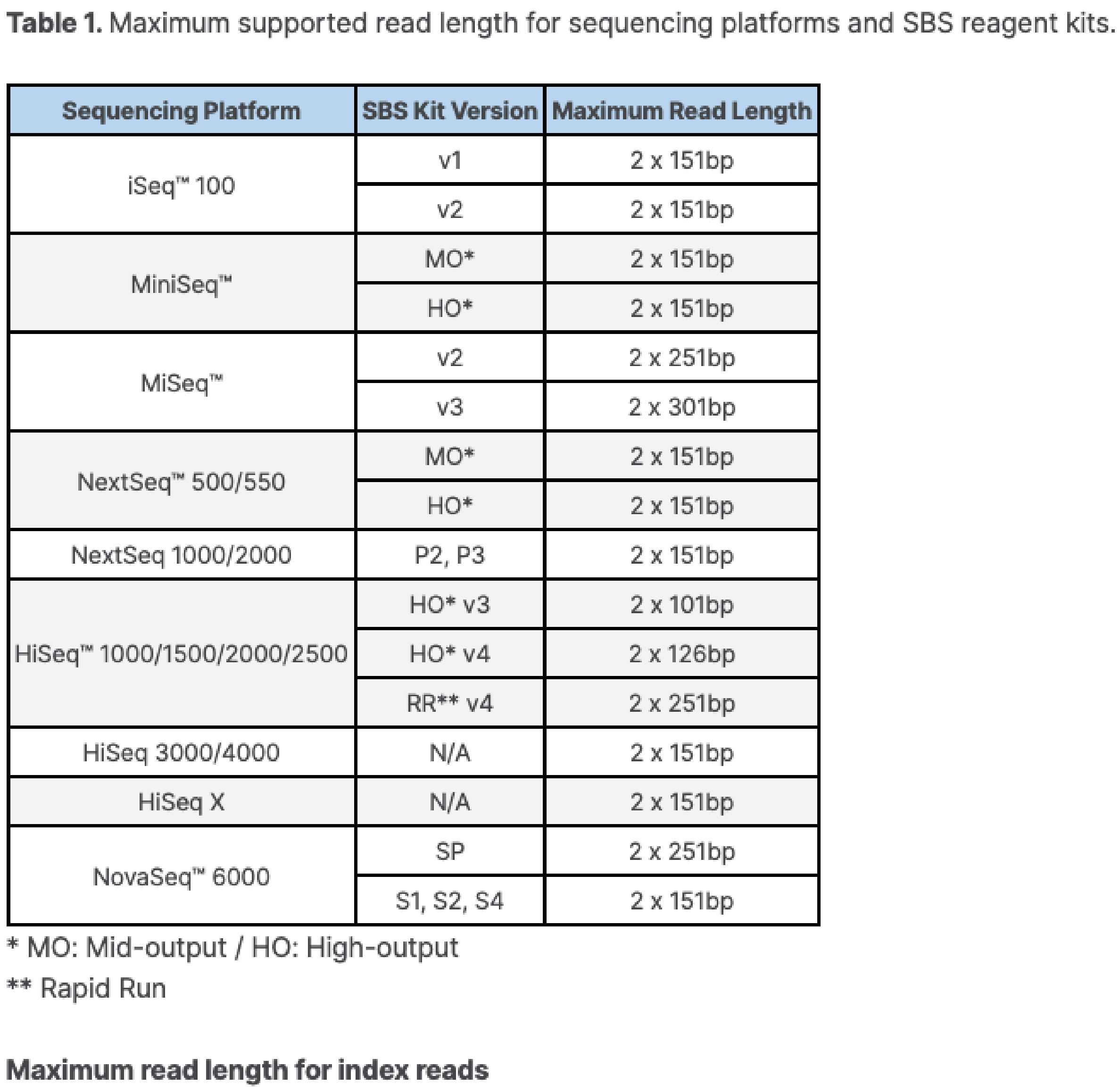


Table 1. Maximum supported read length for sequencing platforms and SBS reagent kits.

Sequencing Platform	SBS Kit Version	Maximum Read Length
iSeq™ 100	v1	2 x 151bp
	v2	2 x 151bp
MiniSeq™	MO*	2 x 151bp
	HO*	2 x 151bp
MiSeq™	v2	2 x 251bp
	v3	2 x 301bp
NextSeq™ 500/550	MO*	2 x 151bp
	HO*	2 x 151bp
NextSeq 1000/2000	P2, P3	2 x 151bp
HiSeq™ 1000/1500/2000/2500	HO* v3	2 x 101bp
	HO* v4	2 x 126bp
	RR** v4	2 x 251bp
HiSeq 3000/4000	N/A	2 x 151bp
HiSeq X	N/A	2 x 151bp
NovaSeq™ 6000	SP	2 x 251bp
	S1, S2, S4	2 x 151bp

* MO: Mid-output / HO: High-output

** Rapid Run

Maximum read length for index reads

@NHM →

@NHM →

Unit 1: Introduction to short read sequencing and library preparation

Bioinformatics Lab

GitHub

https://github.com/nhm-herpetology/museum-NGS-training

Overview…

- We will be doing most things via command line
- Many different ways to perform the same task
- Good to know your options so you can optimise the use of your time and troubleshoot
- WinSCP versus PuTTY examples of making a directory

Let’s make some common directories

- mkdir NGS_course
- cd NGS_course
- mkdir Unit_1
- cd Unit_1
- mkdir Data
- cd Data
- mkdir raw-fastq