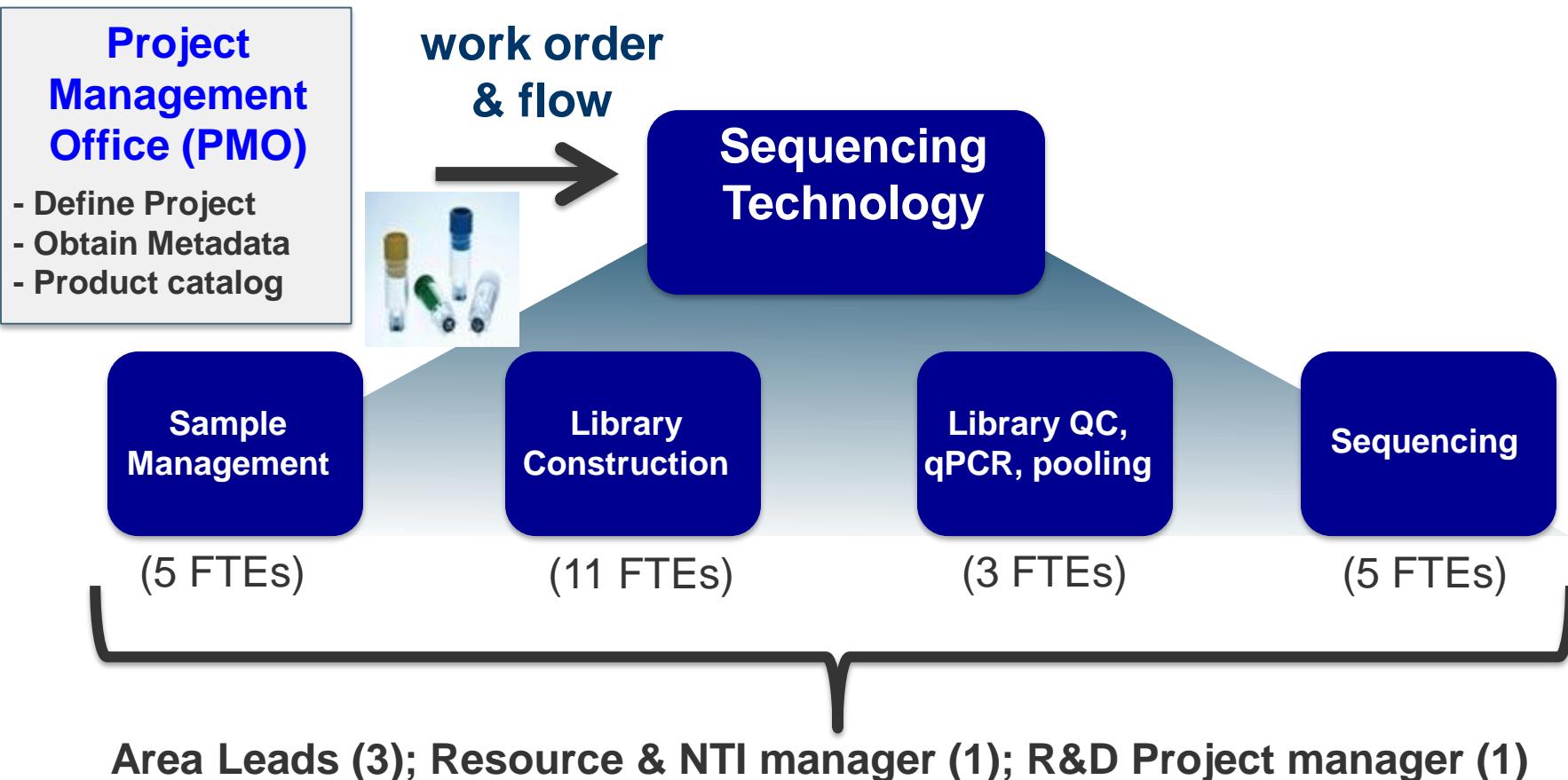




JGI Sequencing Technologies: Overview & Capabilities

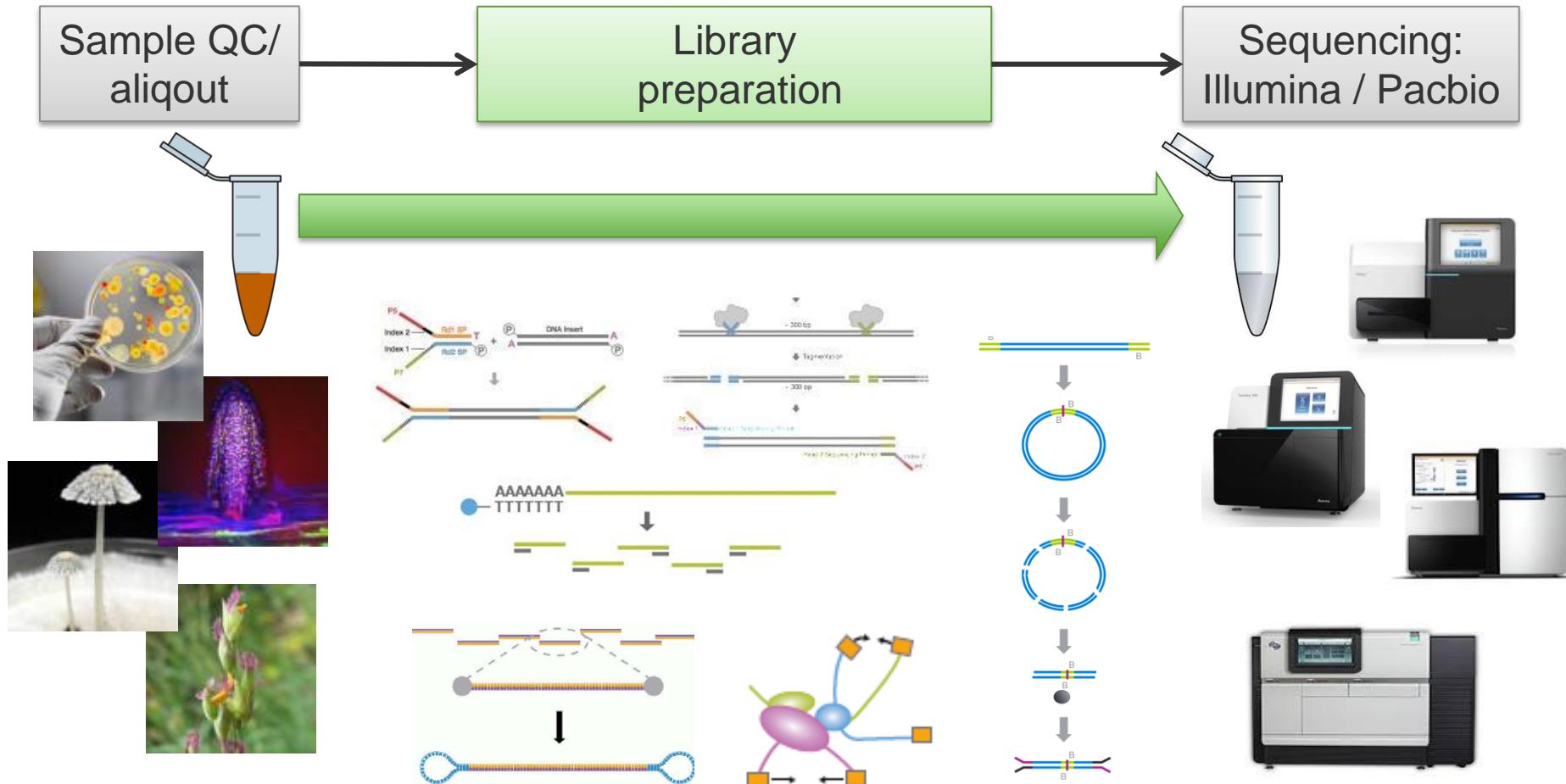
Chris Daum
cdaum@lbl.gov
September 26, 2016

Sequencing Technologies Group



- Streamlined Process from Sample In to Sequence Data Out
- Perform Process Optimization & Development:
 - New Preps, Applications, Sequencing Technologies
 - Continuous Improvement & Lean Manufacturing Six-Sigma

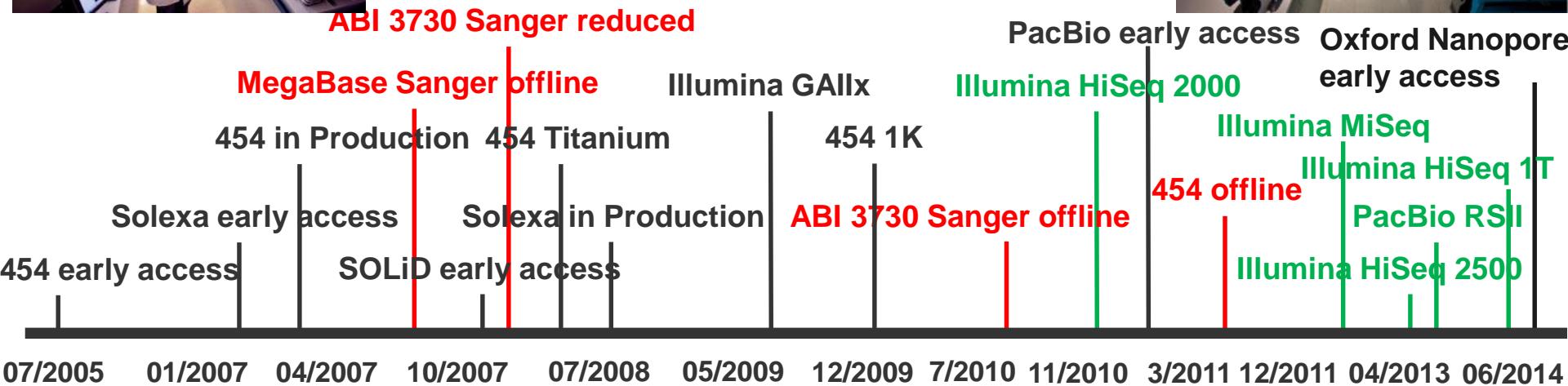
Sequencing Technologies: Sample to Data



Supported by ITS-Clarity LIMS tracking system

Staying State of the Art

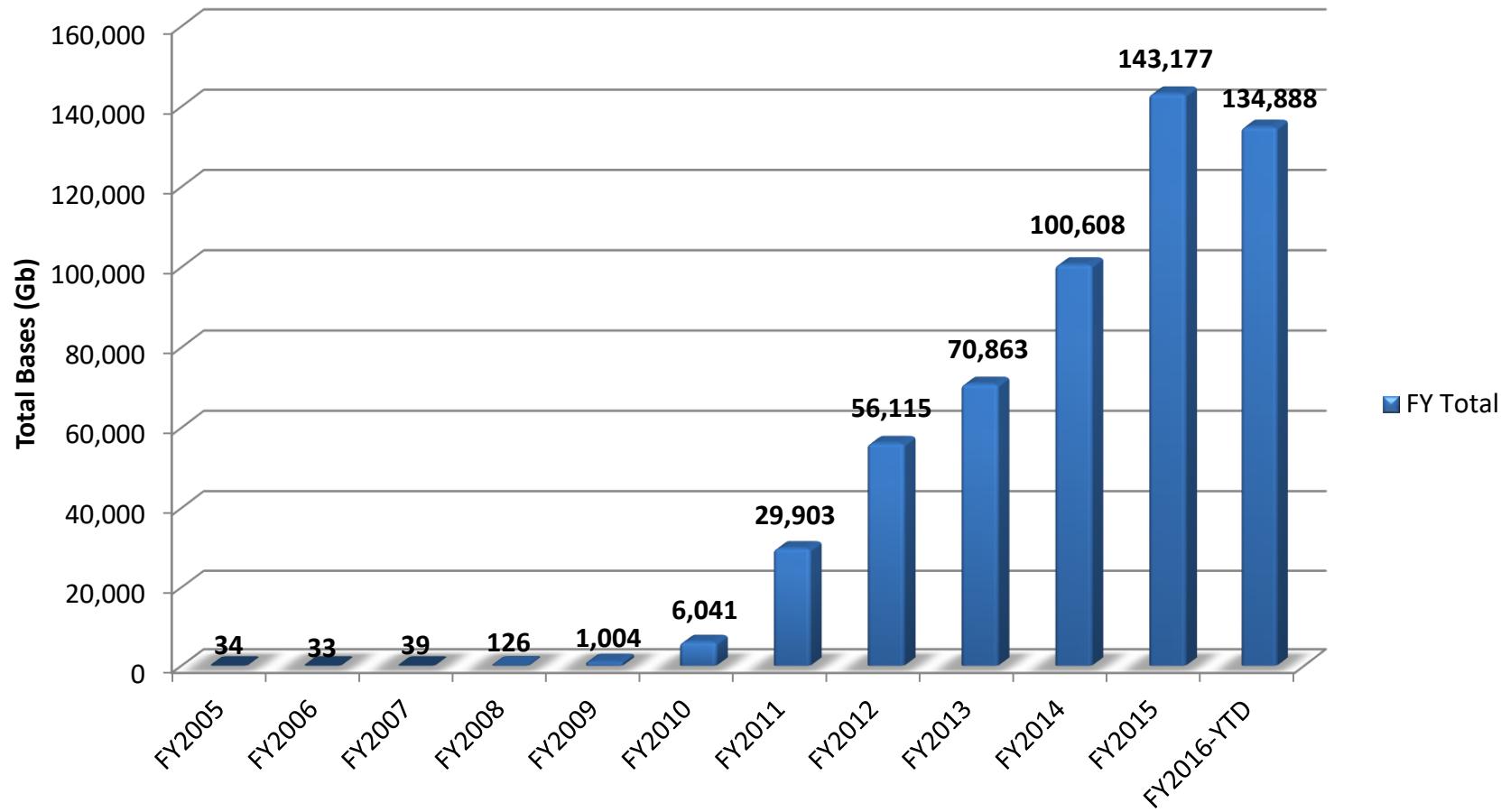
Sanger Sequencing to Next-Gen Sequencing by Synthesis



JGI Yearly Base Output



FY Total Bases (Gb) Sequenced - All Platforms



JGI Sequencing Platforms Portfolio



							
	Illumina HiSeq 1T	Illumina HiSeq 2500	Illumina HiSeq 2000	Illumina NextSeq 500	Illumina MiSeq	PacBio RSII	PacBio Sequel
Units	3	3	2	1	5	3	2
Reads (Single-Read)	>1,500 Million per Flowcell	200 Million per Flowcell	>1,000 Million per Flowcell	400 Million per Flowcell	>10 Million per Flowcell	0.06 Million per SMRT Cell	0.4 Million per SMRT Cell
Readlength	2 X 150bp Max*	2 X 250bp Max	2 X 150bp Max*	2 X 150 Max	2 X 300bp Max	>12,500bp Avg; >40,000bp Max	12,000bp Avg; >40,000bp Max
Total Bases	500 Gb per Flowcell	130 Gb per Flowcell	350 Gb per Flowcell	>100 Gb per Flowcell	5-20 Gb per Flowcell	>0.4 Gb per SMRT Cell for 2hr runs; >0.8Gb for 4hr runs	>2.7 Gb per SMRT Cell for 2hr runs; >5.0 Gb for 4hr runs
Run Time	7 Days for 2 X 150	4.5 Days for 2 X 250	16 Days for 2 X 150	1 Days for 2 X 150	2 Days for 2 X 300	0.08-0.12 Days (2-4 hours)	0.08-0.12 Days (2-4 hours)
Applications	Primary Sequence Generator at JGI	Rapid output HiSeq	Supplement / Backup Platform	Rapid mid-range output; Single Cell	16S/18S iTags, Library QC, R&D	Assembly improvement, de novo, SynBio validation, methylation/ epigenetics	Early Access Testing & Validation of Technology

Major Investment in Automation: Process

Sample
Archive &
Retrieval



SAM

Sample QC &
Aliquot



Starlet x 2

Library
Construction



Sciclone G3 x3

Library QC,
qPCR, pooling



Star x 2



Automated
Decapper



BioMicroLab
Volume Check



BioTek Synergy H1
Plate Reader

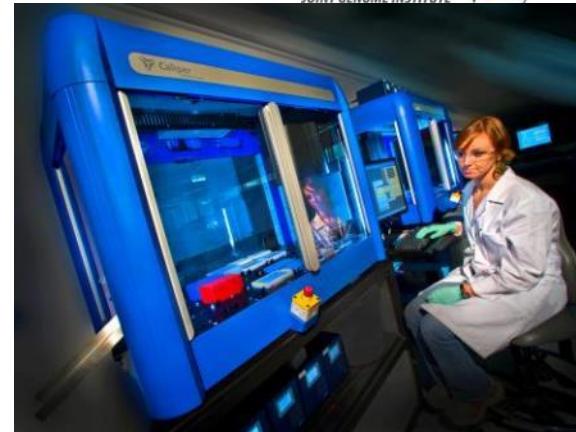


5ul PCR
Labcyte Echo

Maximize Consistency, Throughput and Reliability

Automated Library Creation

- **Plate based Automated library preps:**
 - Implemented into Production in late 2011
 - 3 PerkinElmer Sciclone NGS robots
 - 12 production sample prep methods are supported
 - >24,000 sample sequencing libraries prepared in last year



- **Supporting Equipment:**



ScicClone Automated Preps

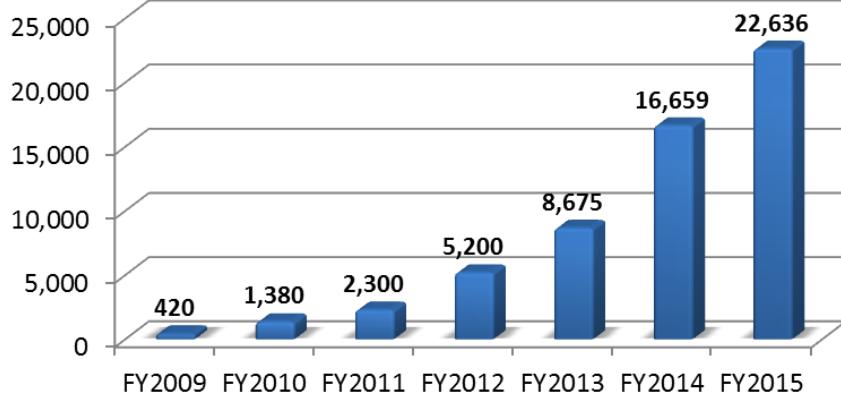


Supported Preps

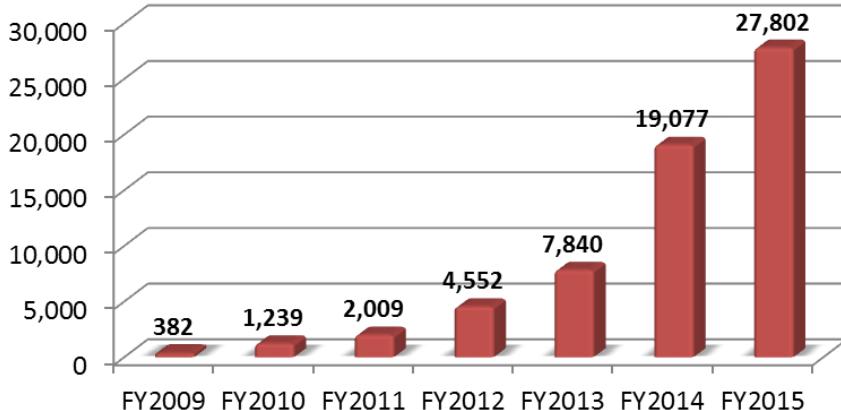
- Illumina gDNA PCR-free WGS Fragment Library prep
- Illumina TruSeq RNA-seq stranded library preps:
 - PolyA selection of mRNA for eukaryotes
 - rRNA depletion for microbes & metatranscriptomes
- Illumina small RNA & miRNA for eukaryotes
- Illumina iTags (16S proks; 18S euks; Fungal ITS)
- Illumina Exome Capture (NimbleGen SeqCap) prep for targeted resequencing
- Illumina NexteraXT prep
- Illumina Methyl-Seq (bisulfite conversion) prep
- PacBio DNA 2kb libs, and >10kb libs with enzymatic shearing

Continued Growth in Number of Samples Handled

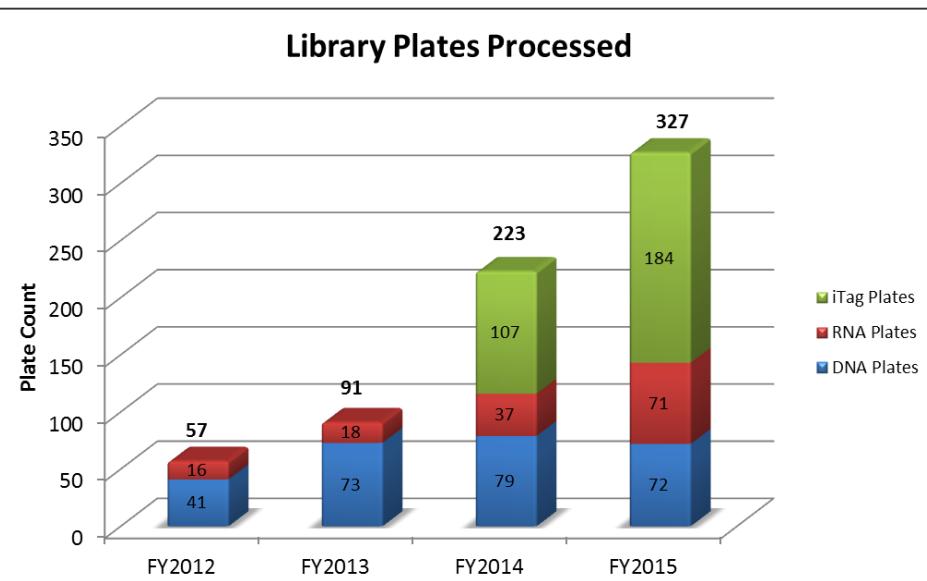
Samples Received



Libraries Created



Library Plates Processed



A Growing Portfolio of Library Capabilities

Genome: DNA

WGS
WGS: Tight insert
LMP
Pacbio long reads



Genome assembly
Structural variation
Comparative genomics
Genotyping

Transcriptome: RNA

Stranded RNA-seq
Poly A
rRNA depletion
FFPE/LCM
smRNA
Iso-seq



Gene annotation
Gene expression

Epigenome: DNA-protein

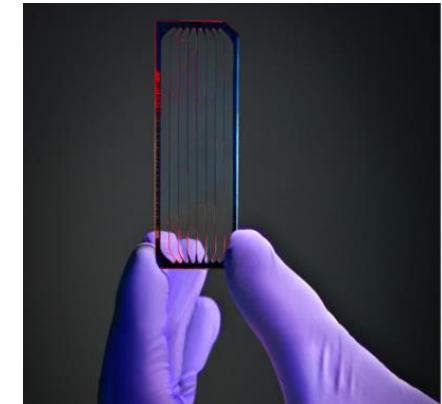
Bisulfite-seq
FAIRE-seq
ChIP-seq
ChIA-PET



Methylation
Gene regulation
Chromatin conformation
Protein binding

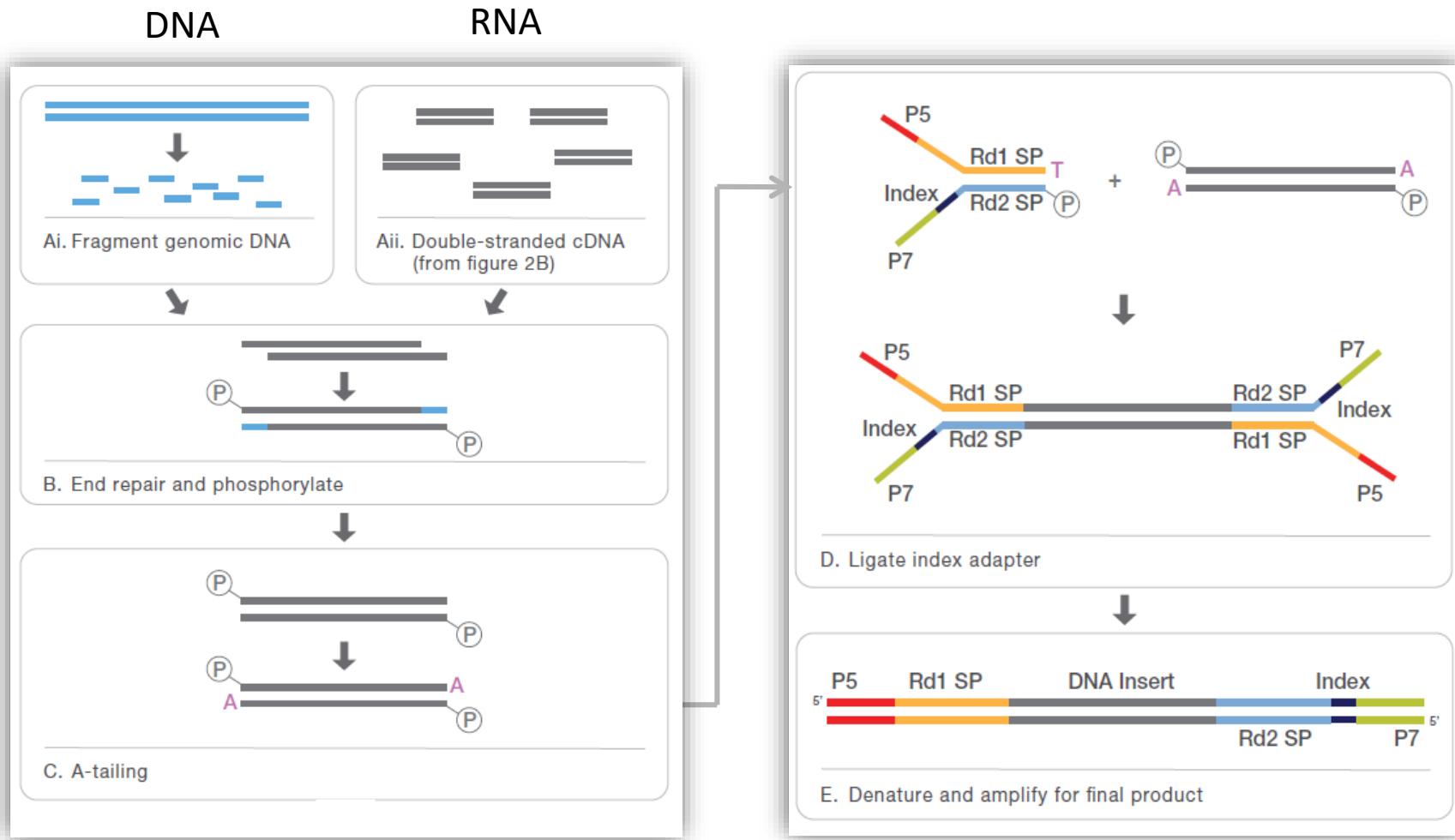
Illumina Sequencing Platform

- Technology Overview
- Technology Updates
- Application: 16S iTag Sequencing



Illumina Sample Prep

Fragment Library Creation:



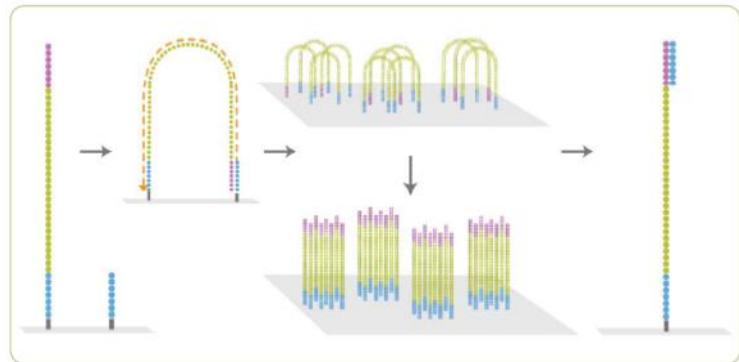
Illumina Sequencing – SBS Technology



Apply Templates to Flow Cell

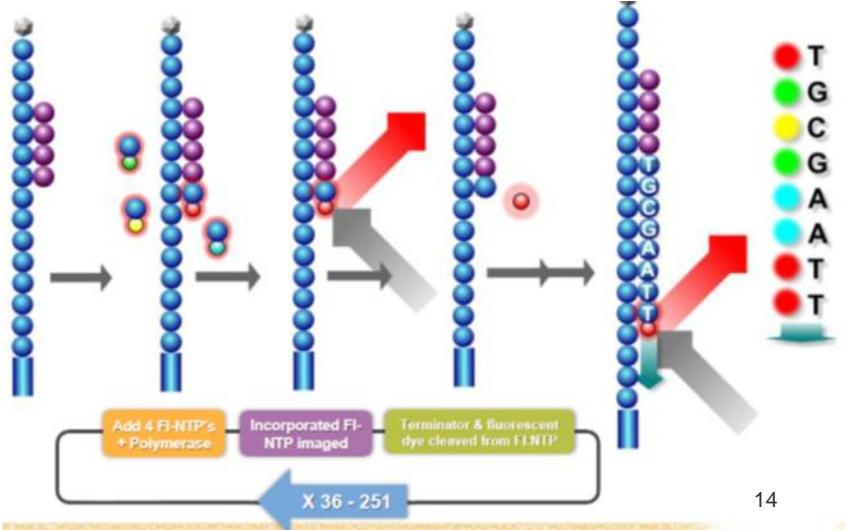


Cluster Generation by PCR (Bridge Amplification)



Load Flow Cell on Sequencer

Sequencing by Synthesis with Reversible Terminators



Illumina Technologies

NextSeq 500 –
midrange sequencer



NeoPrep – automated
microfluidic library
preparation



HiSeq X Ten – set of 5 or 10 ultra-high throughput sequencers for
human whole genome sequencing; Illumina's \$1,000 human genome



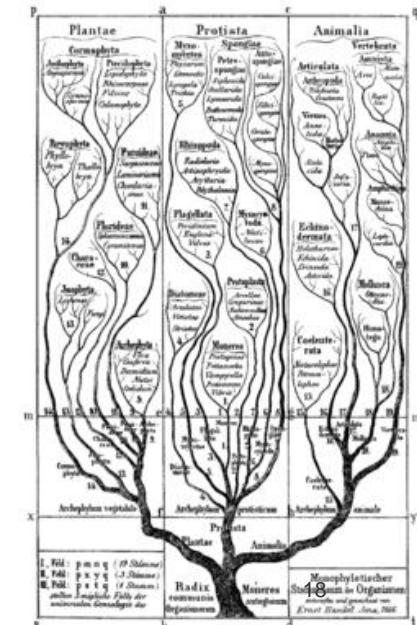
Illumina - Project Firefly

- **Semiconductor sequencing on a CMOS sensor**
 - Nanowells are fabricated over photodiodes and clusters are built directly on top of the CMOS sensors
 - Based on technology from Avantome (Mostafa Ronaghi) that Illumina acquired in 2008
 - Utilizes 1-channel SBS chemistry:
 - A (removable fluor), T (permanent fluor), G (no fluor) and C (no fluor, but tagged).
 - After initial incorporation & imaging, chemistry step is run to remove fluor from A and add fluor to C, a secondary image is taken.
 - Using these two images per cycles and changes in fluorescence, all 4 bases may be determined.
 - Projected launch in late 2017. Initial launch specs:
 - \$30K for sequencer & sample prep instruments; \$100 per run
 - 3.5-13hr run times, up to 2x150 readlengths
 - 4M reads, 1Gb output



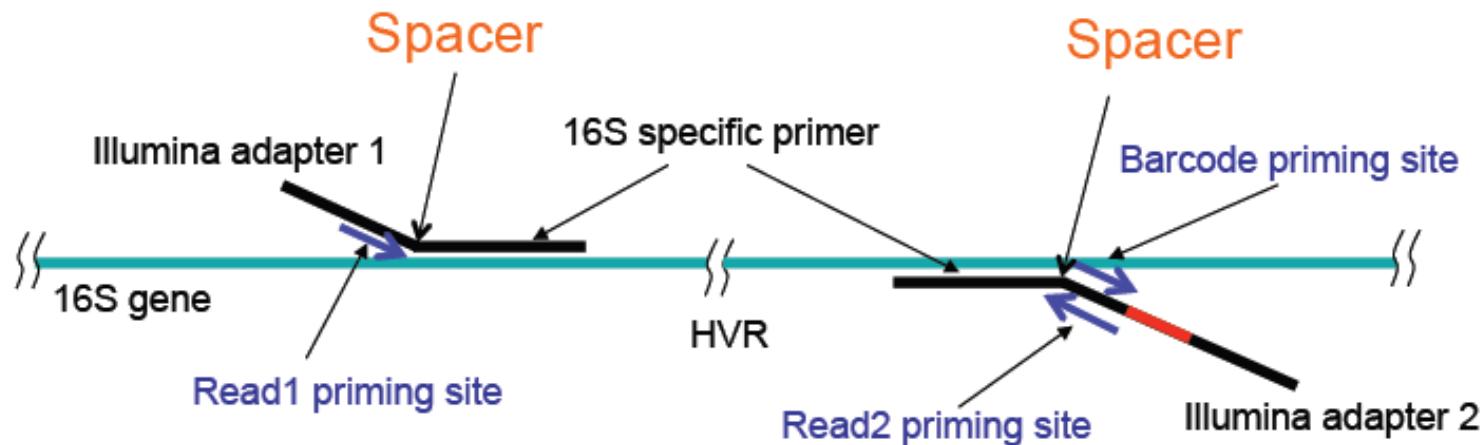
16S iTag sequencing

- **iTags – microbial community profiling with 16S sequencing on MiSeq**
 - Target 16S rRNA v4-v5 region with specific primers to generate amplicon libraries
 - Also have primers specific for Eukaryotic 18Sv4 and Fungal ITS
 - Utilize staggered adapters to introduce sequencing diversity and improve data quality
 - Up to 184 samples pooled per 2x300 MiSeq run



16S tagging on MiSeq

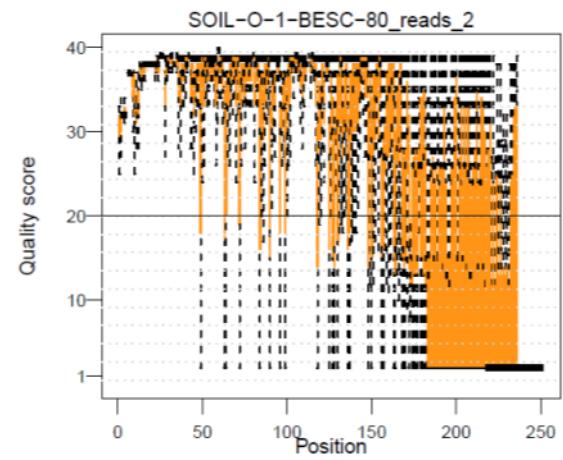
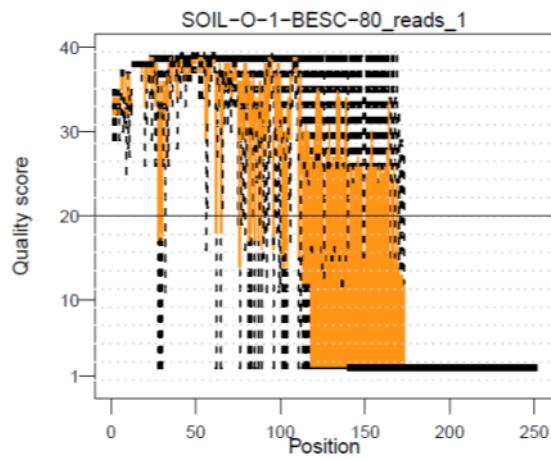
- Conserved nature of 16SrRNA means low diversity sequence between samples, which causes quality issues with Illumina sequencing technology
- Utilize staggered adapters with varying length spacer sequence:
 - Spacer is 0, 1, 2, 3, 4 nt. in length
- Once the amplicons are pooled the spacers introduce diversity and desynchronize sequence data, ensuring high quality



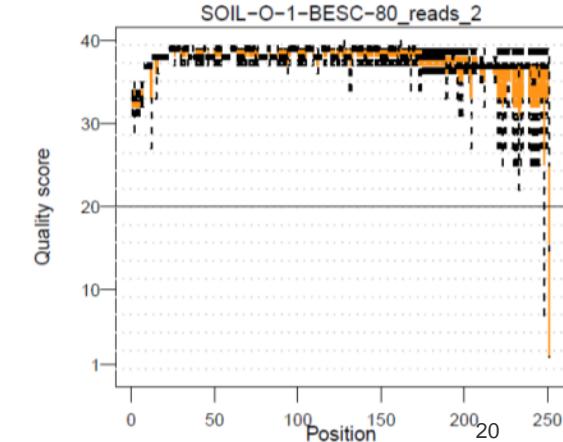
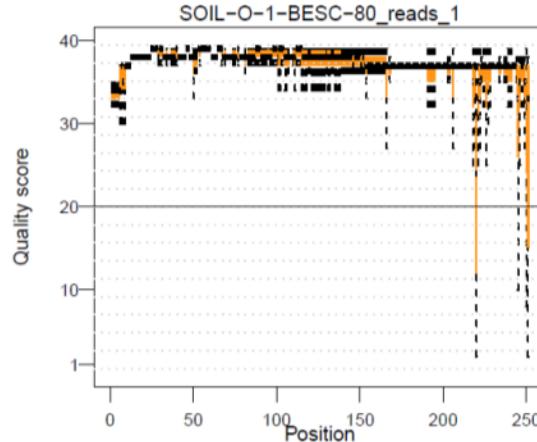
iTags Staggered Adapters

Staggered adapters in action:

Non-Staggered
-poor read quality



Staggered
-improved read
quality



PNA Oligos for depletion of chloroplast and mitochondrial amplicons



Peptide Nucleic Acid (PNA) Oligos

- PNA oligos designed to bind to plant organelle sequences are added to the 16S prep
- During the PCR amplification PNAs prevent the extension of the targeted molecules
- 16S preps for endophyte samples with PNA increases total usable reads

Chemistry & Biology
Article

Targeted Disruption of the CCR5 Gene in Human Hematopoietic Stem Cells Stimulated by Peptide Nucleic Acids

Erica B. Schleifman,¹ Ranjit Bindra,¹ Jean Leif,² Jacob del Campo,¹ Faye A. Rogers,¹ Pradeep Uchil,³ Olaf Kutsch,⁴

Leonard D. Shultz,⁵ Priti Kumar,⁶ Dale L. Greiner,² and Peter M. Glazer^{1,*}

¹Departments of Therapeutic Radiology and Genetics, Yale University School of Medicine, New Haven, CT 06510, USA

PacBio Sequencing Platform

- Technology Overview
- Technology Updates
- Application: PacBio only Microbes



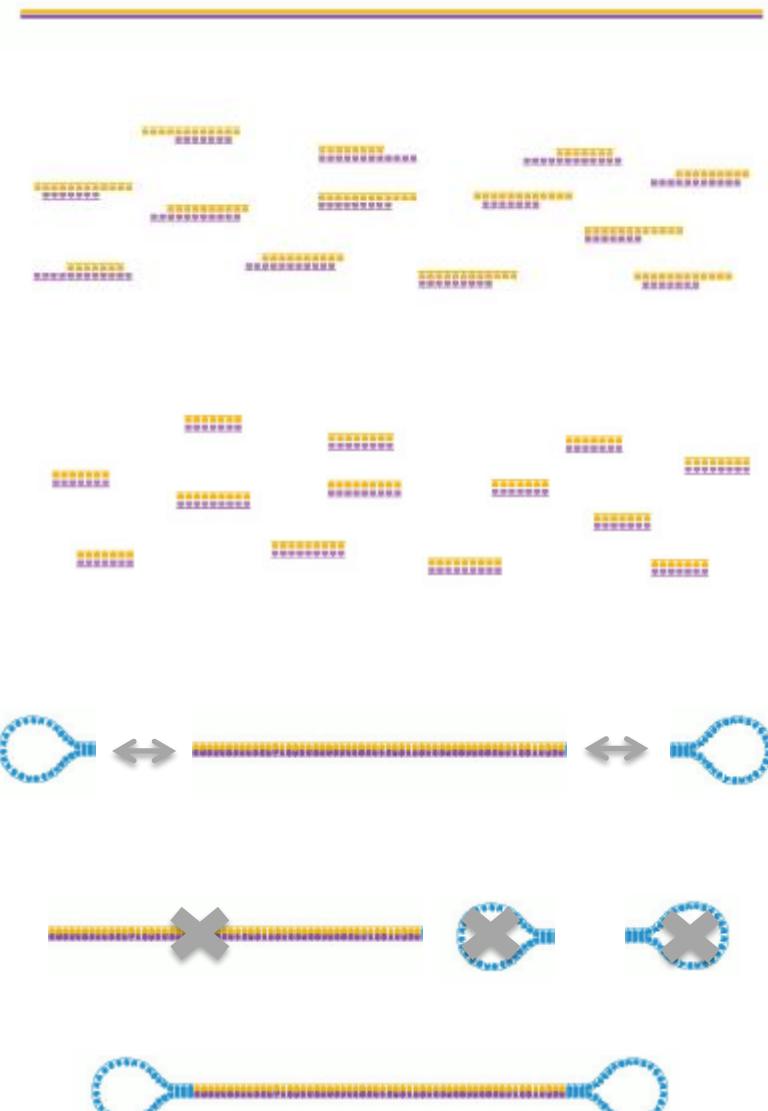
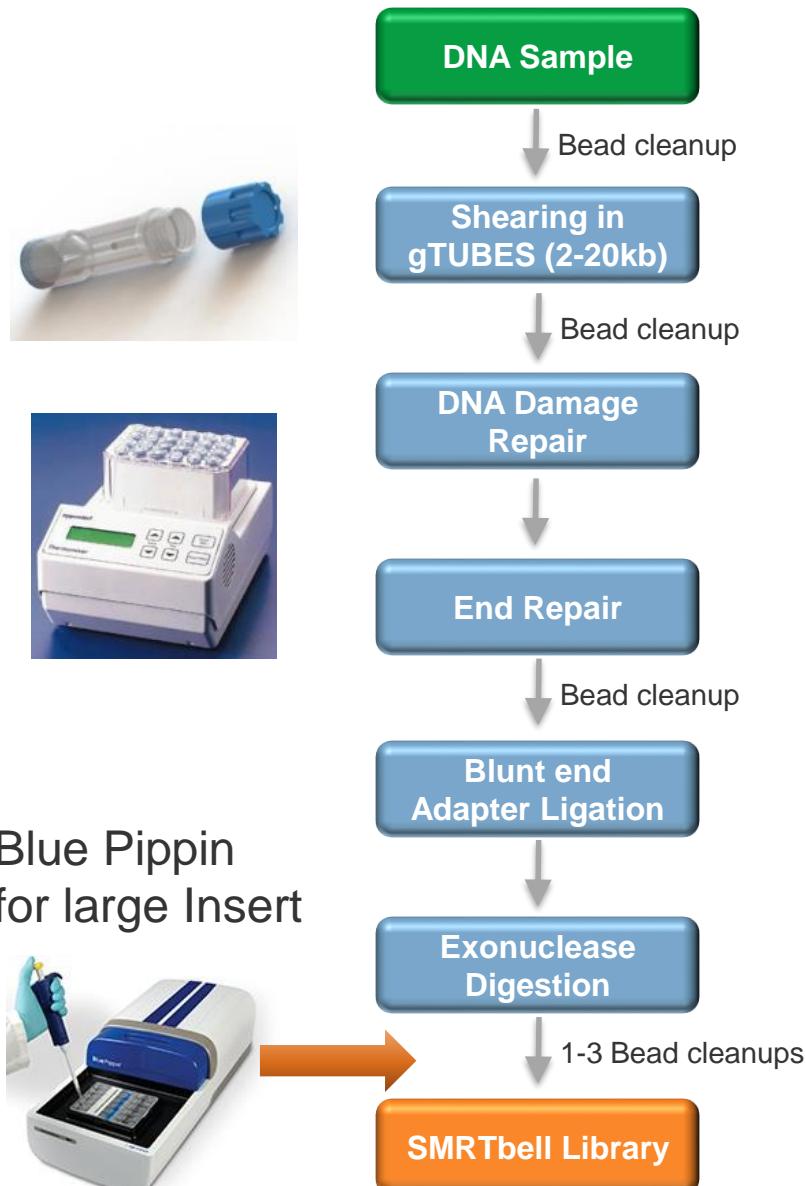
PacBio Single-Molecule, Real-Time (SMRT) Sequencing



- **Single Molecule:** no amplification needed
 - Reduced bias (high/low GC, “hard stops”) with no amplification
 - Wide range of insert sizes and does not require multiple/difficult preps
- **Real Time:**
 - Long read length (avg. ~13kb, max >60kb)
 - Fast prep & sequencing (1 day prep, 2-6 hr/cell movie time, up to 16 cells can be queued for a run)
 - Base modification data with every run of unamplified sample template

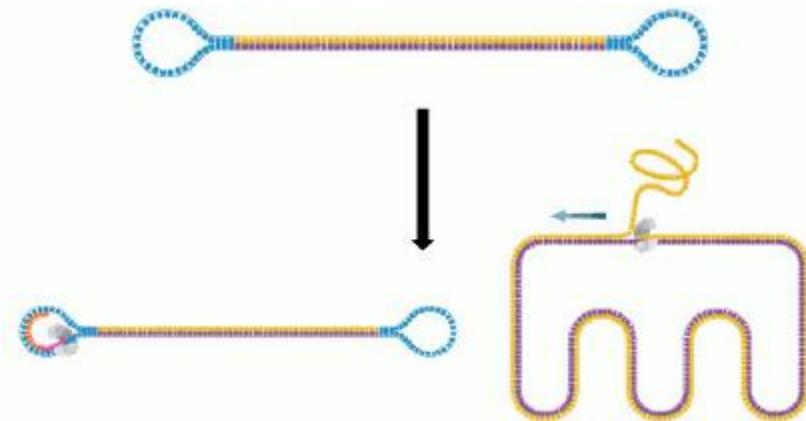
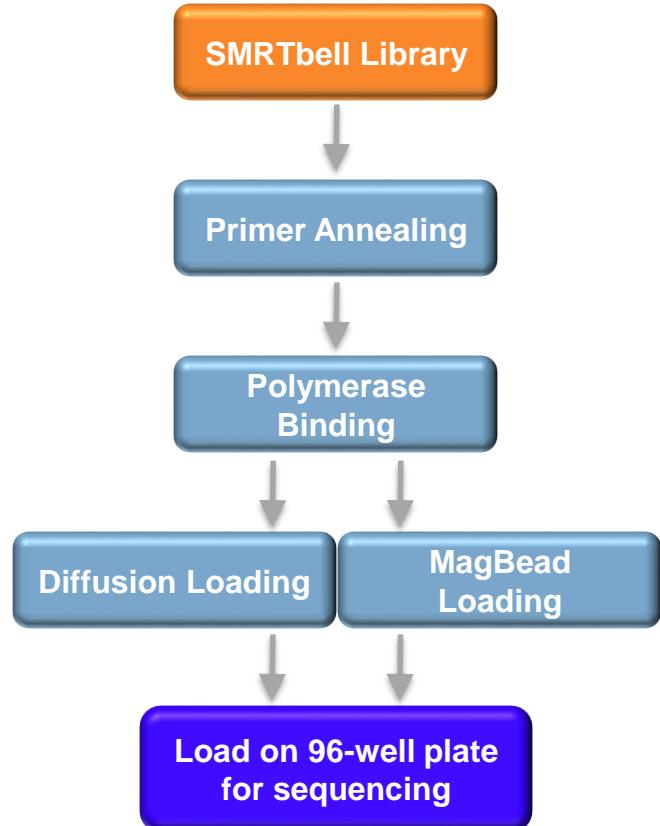


Bench Workflow: SMRTbell Library Construction (1-2 days)



Completed SMRTbell template

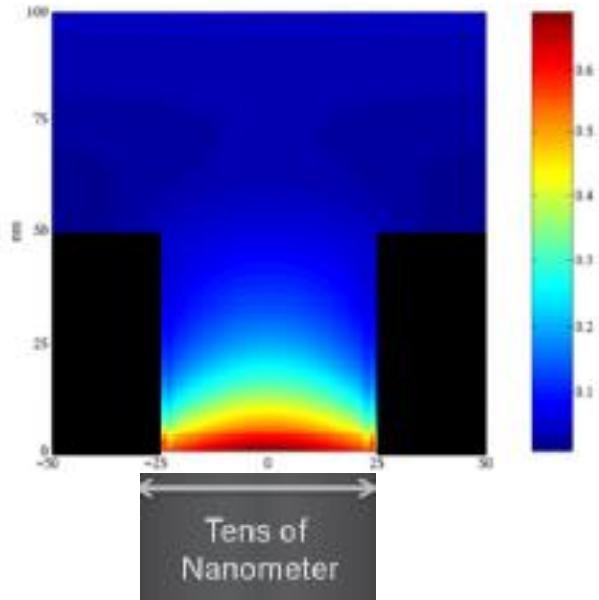
Bench Workflow: Sequencing Preparation & run (1 day)





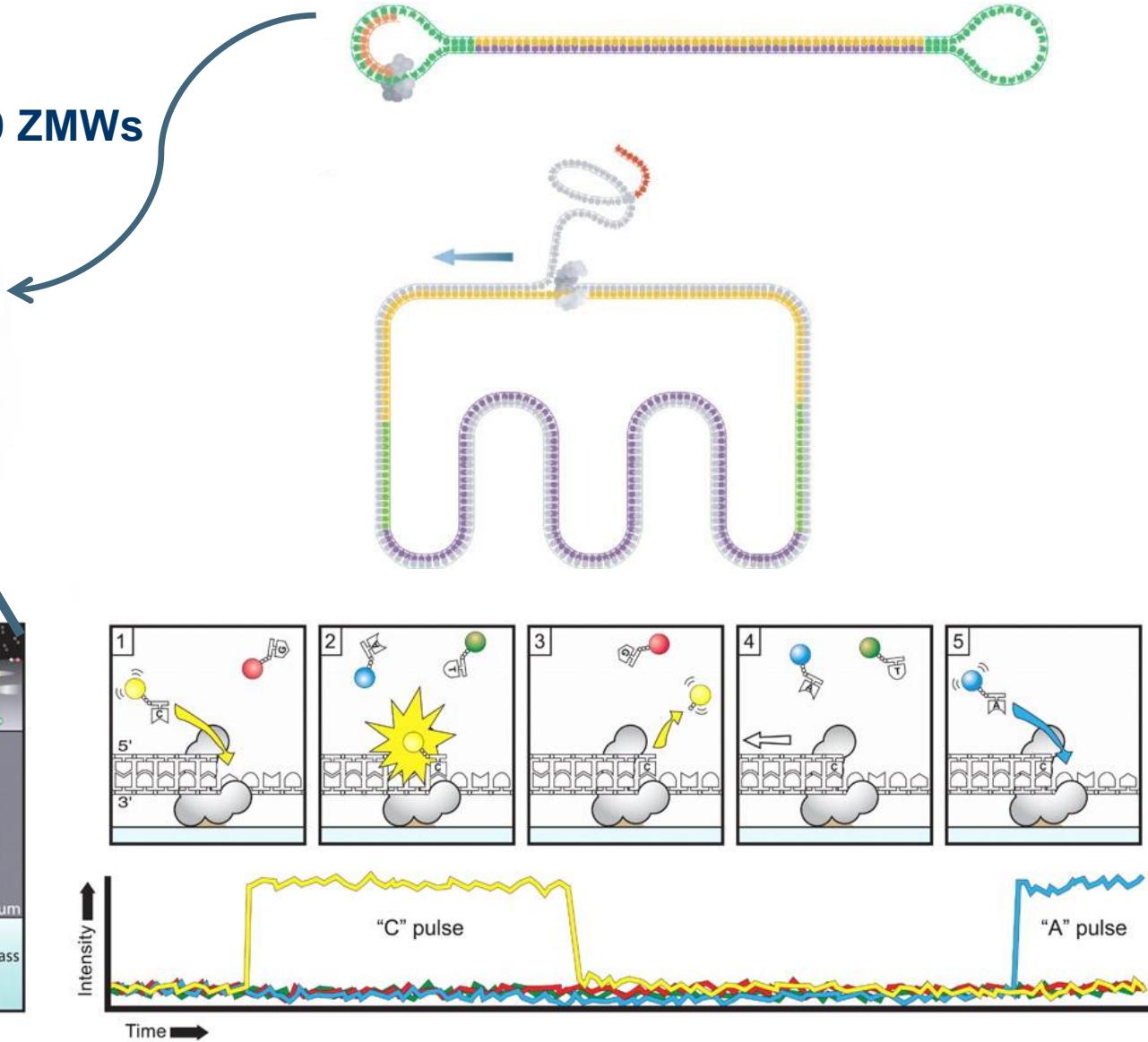
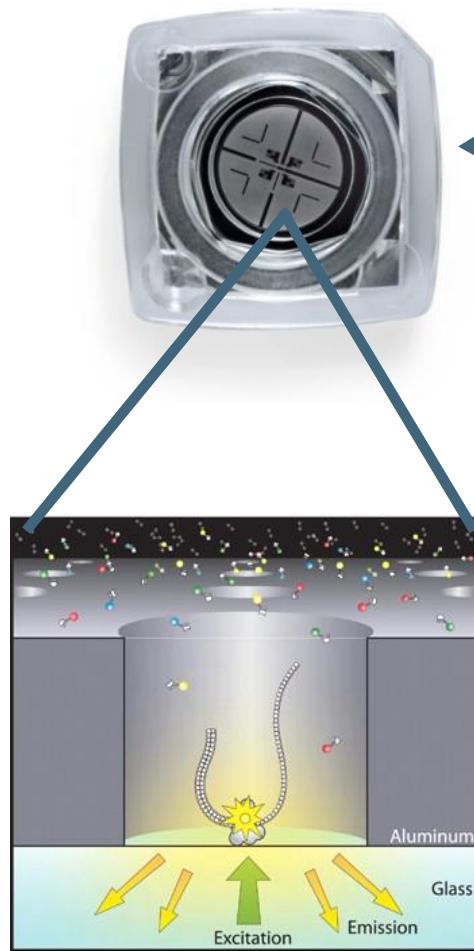
Multiplexed
ZMWs

- 1 SMRT cell = 150,292 ZMWs
 - Chambers for sequencing
 - Zero-mode waveguide (ZMW):
 - Cylindrical well (~50 nanometer diameter) too small for wavelength of light to pass, however allows fluorescence detection at very bottom.



PacBio Sequencing

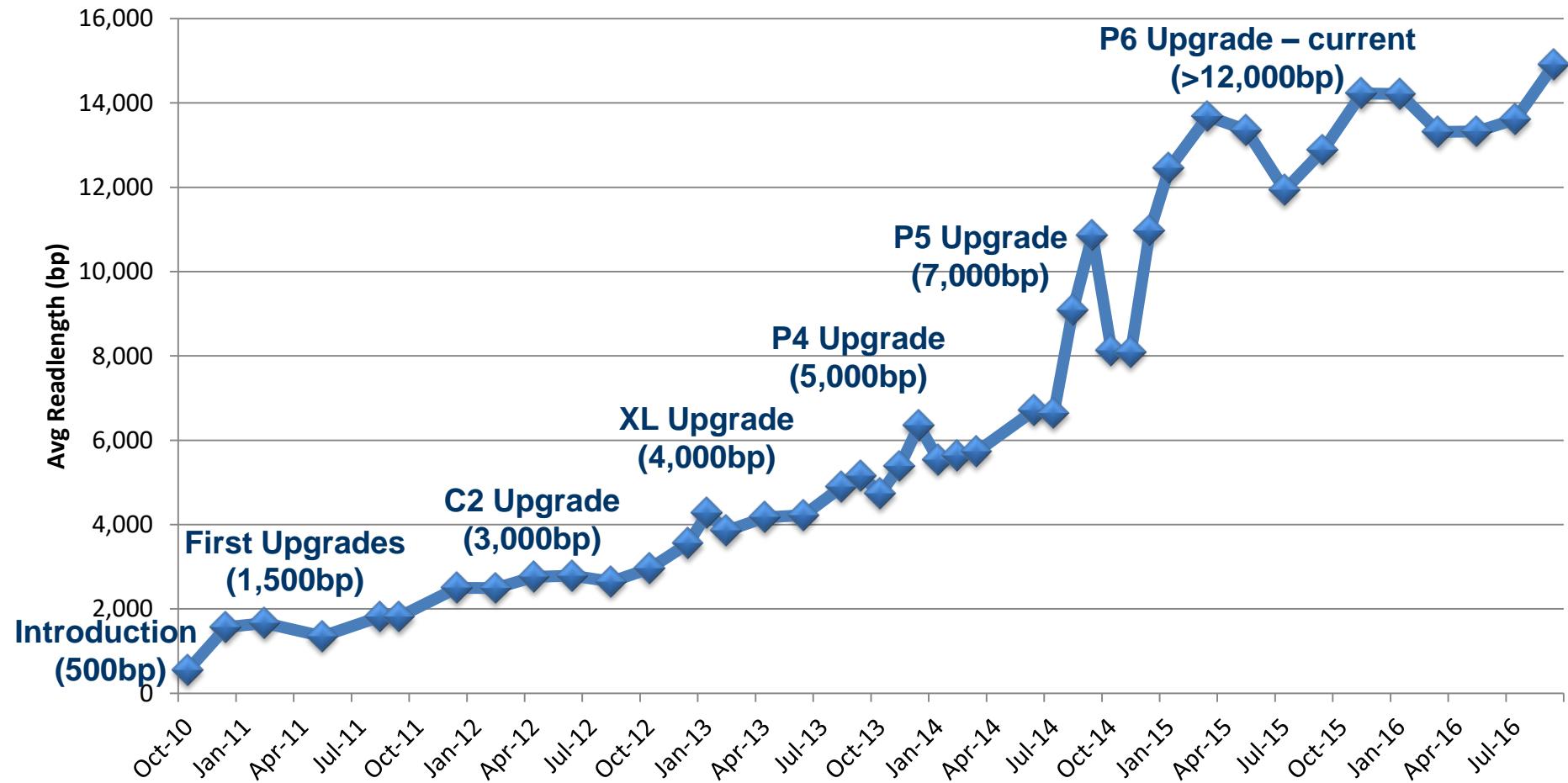
1 SMRT cell = 150,000 ZMWs



JGI PacBio RSII Readlength



PacBio RS Avg Polymerase Readlength since October 2010



PacBio Sequel System

- **Announced October 2015:**
 - Limited availability with 3 installation sites: JGI, Mount Sinai, Roche
 - Two Sequels installed in December 2015 at JGI

- **Primary Platform Benefits:**

Same SMRT Technology
0.5-6.0 hour movies

7X Increased capacity with 1 Million ZMWs/Cell
5Gb+/Run

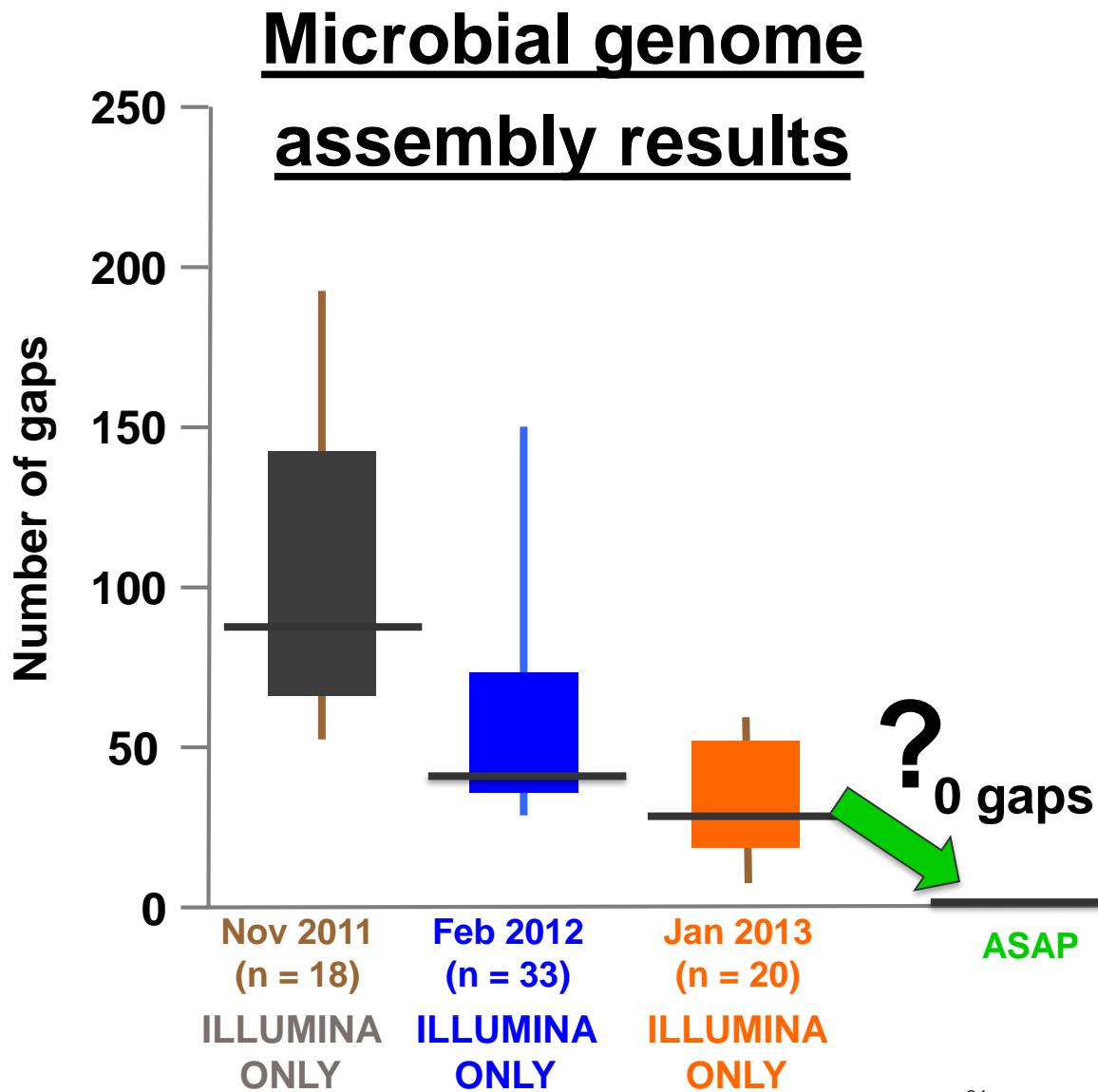
50% Reduced per base costs compared to RSII



PacBio only Microbes

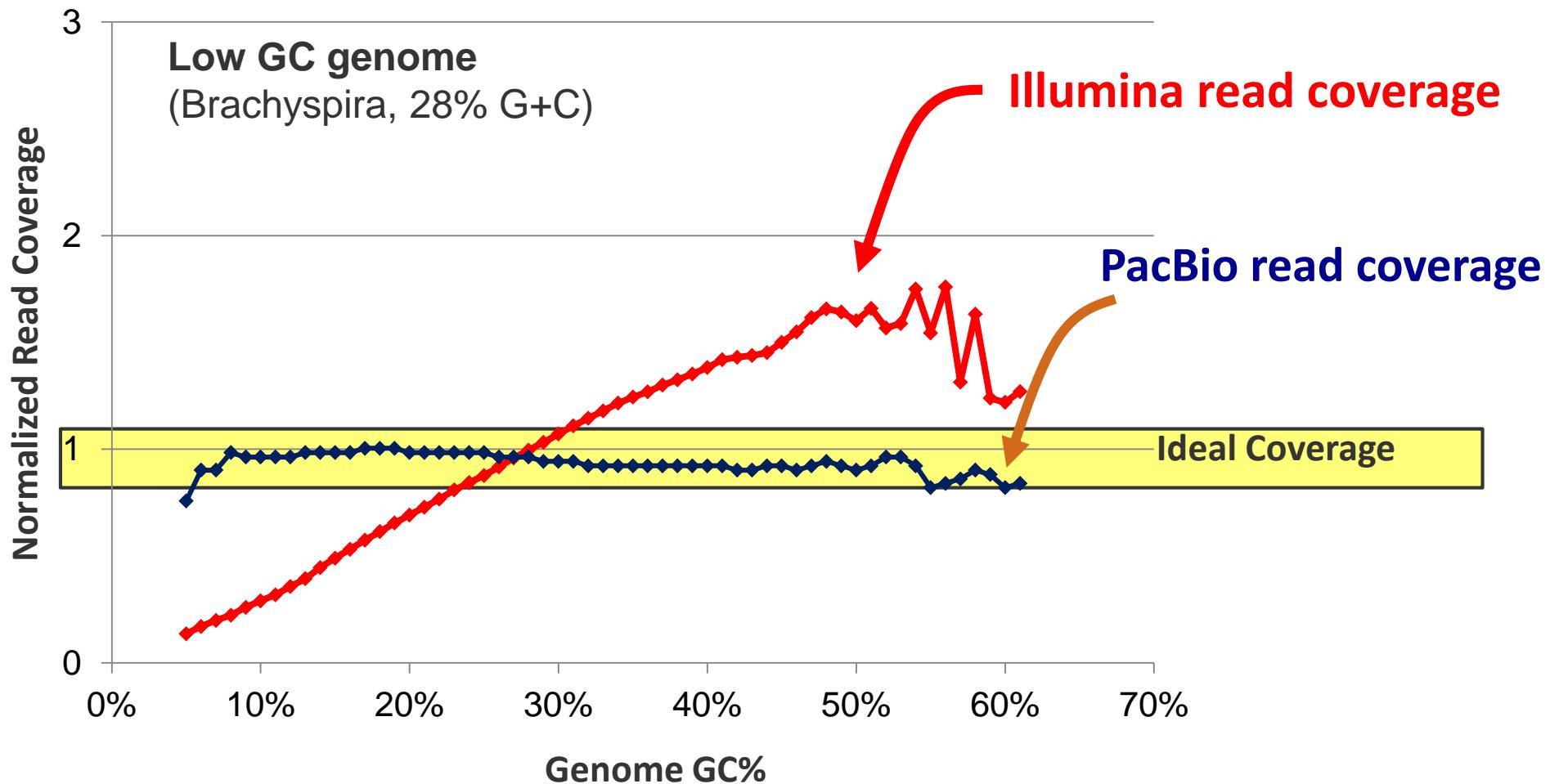
Short read only genome assembly is improving over time

- Better libraries**
- Longer reads**
- Better chemistry**
- Better algorithms**



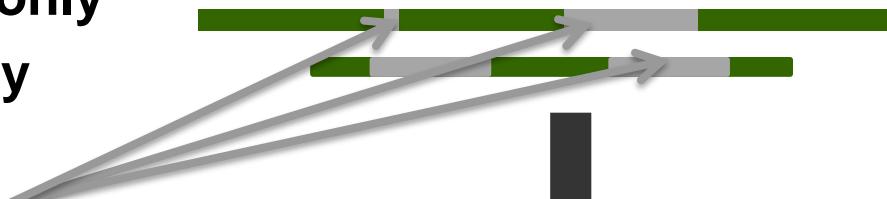
Why are there Gaps?

- Single Molecule Sequencing Read Coverage versus GC%



Using PacBio to improve Assemblies

**Short read only
assembly**



"Captured" gaps caused by
i. repeats
ii. regions of low coverage



PacBio data
can help

Longer reads
(to span repeats)

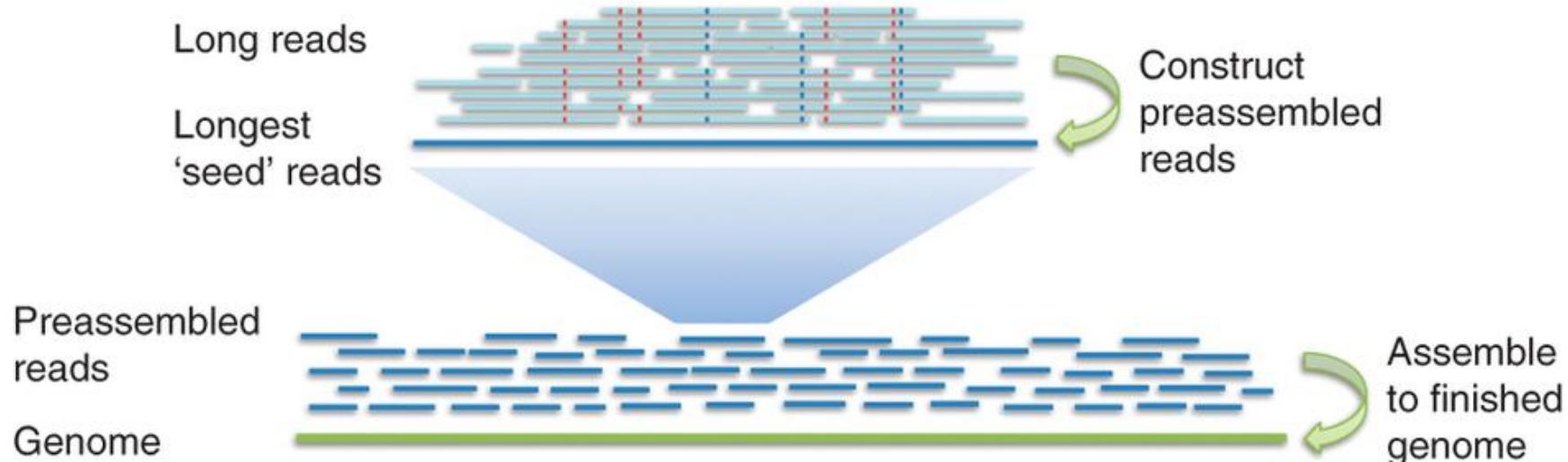
Less biased coverage
(to complement illumina coverage 'gaps')

A better assembly



Use Accurate PacBio Long Reads for de novo Assembly

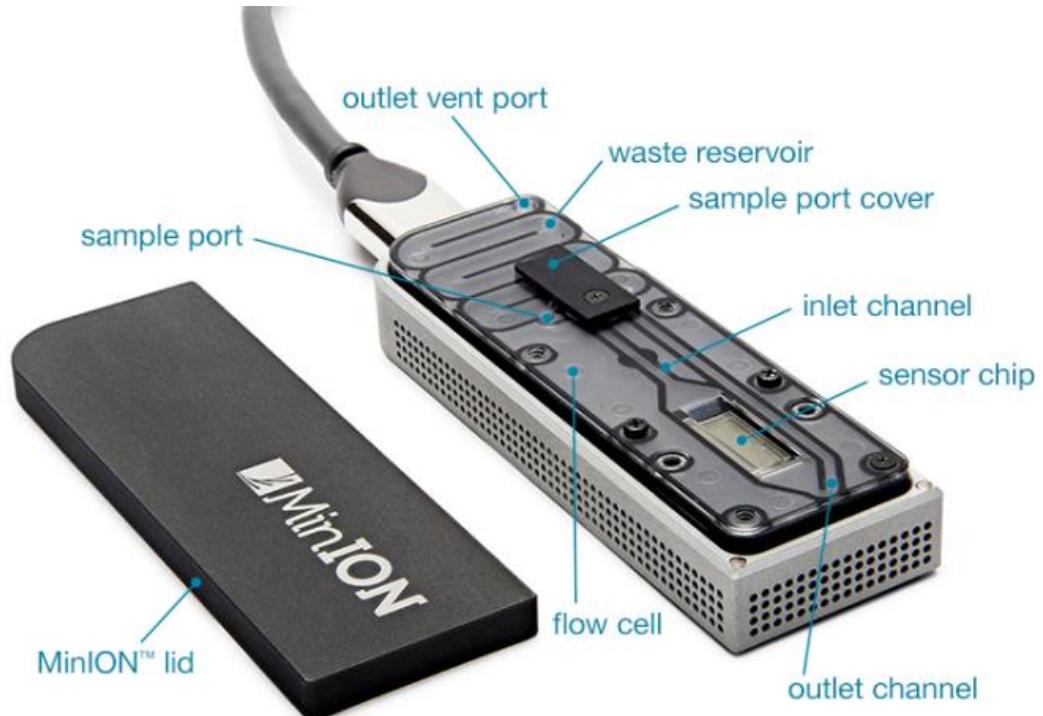
Hierarchical Genome Assembly Process (HGAP)



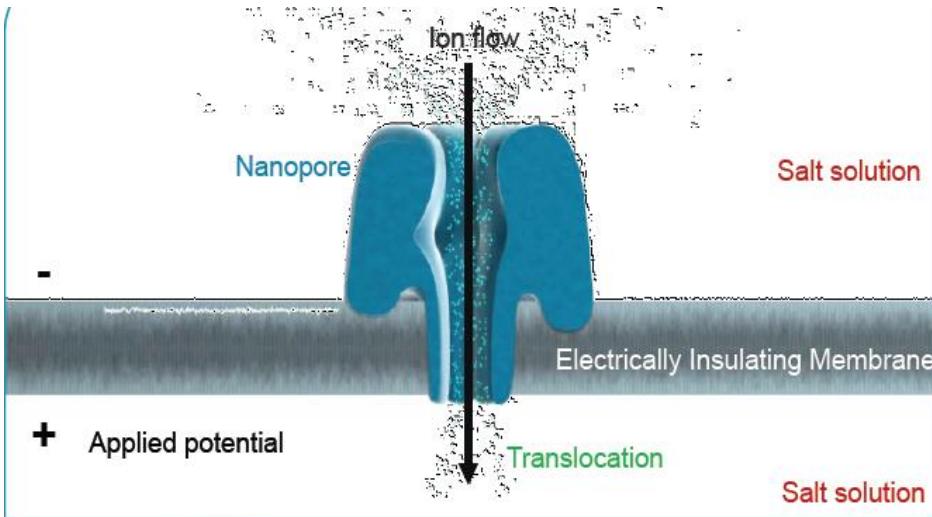
JGI now does de novo sequencing of microbes & fungal genomes for high-quality assemblies.

Emerging Sequencing Technologies

Oxford Nanopore Technology

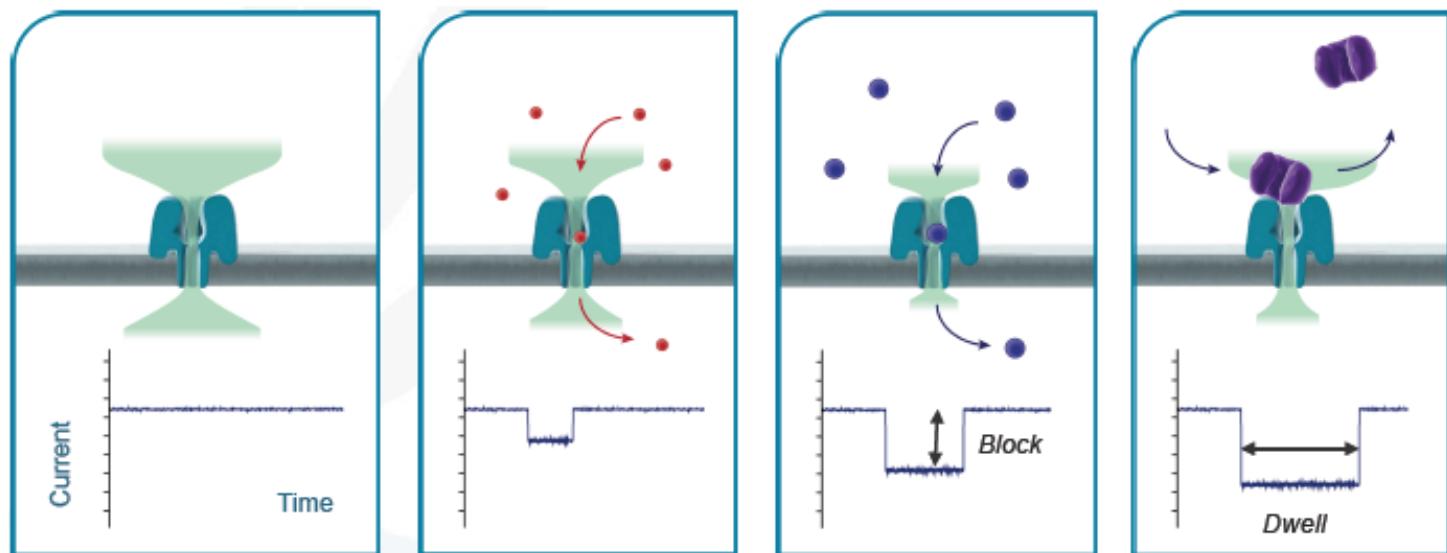


ONT– Overview

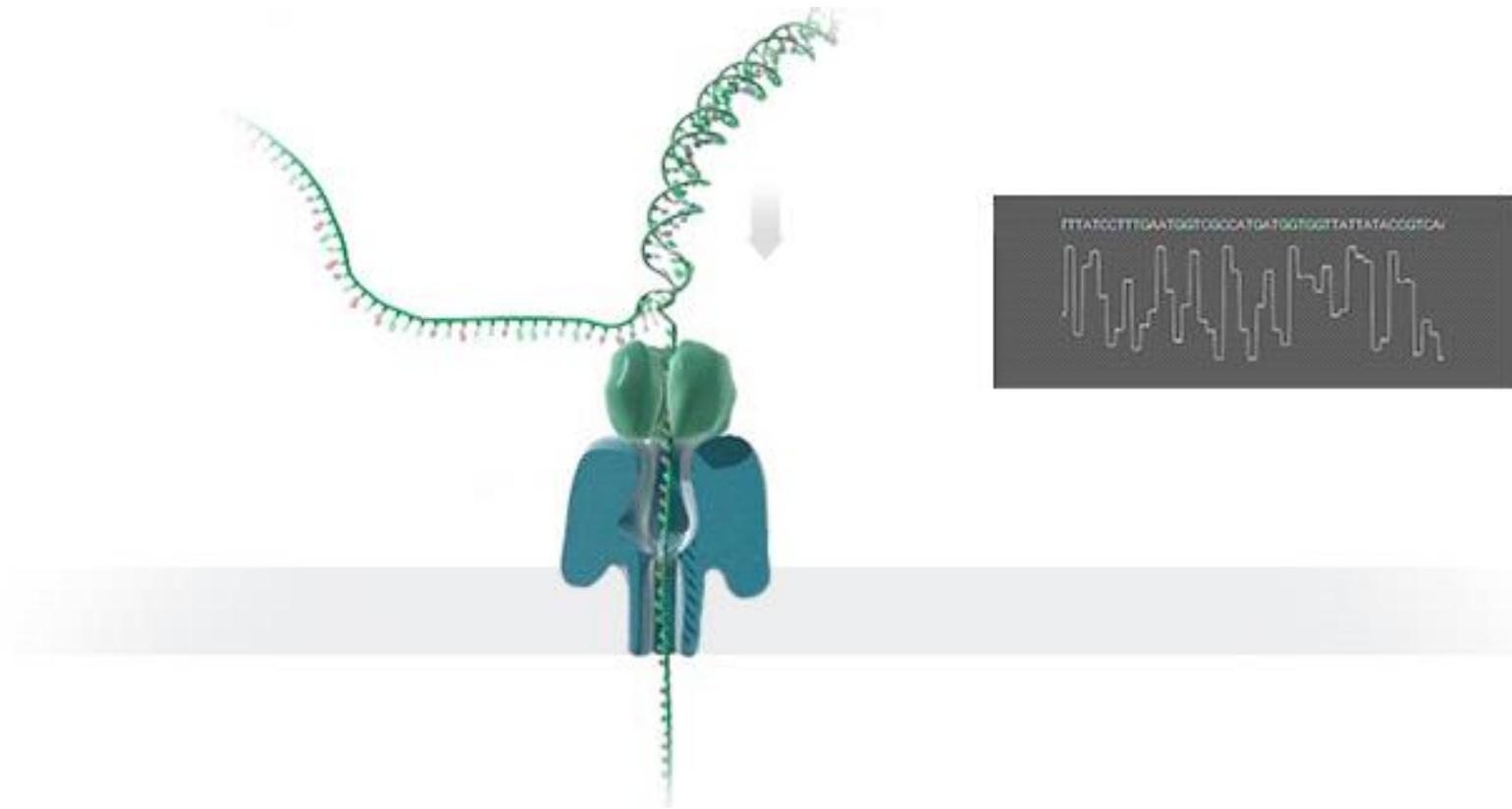


Detecting current Changes as molecules move through pores

Kmers are detected from single DNA strands passing through the pores

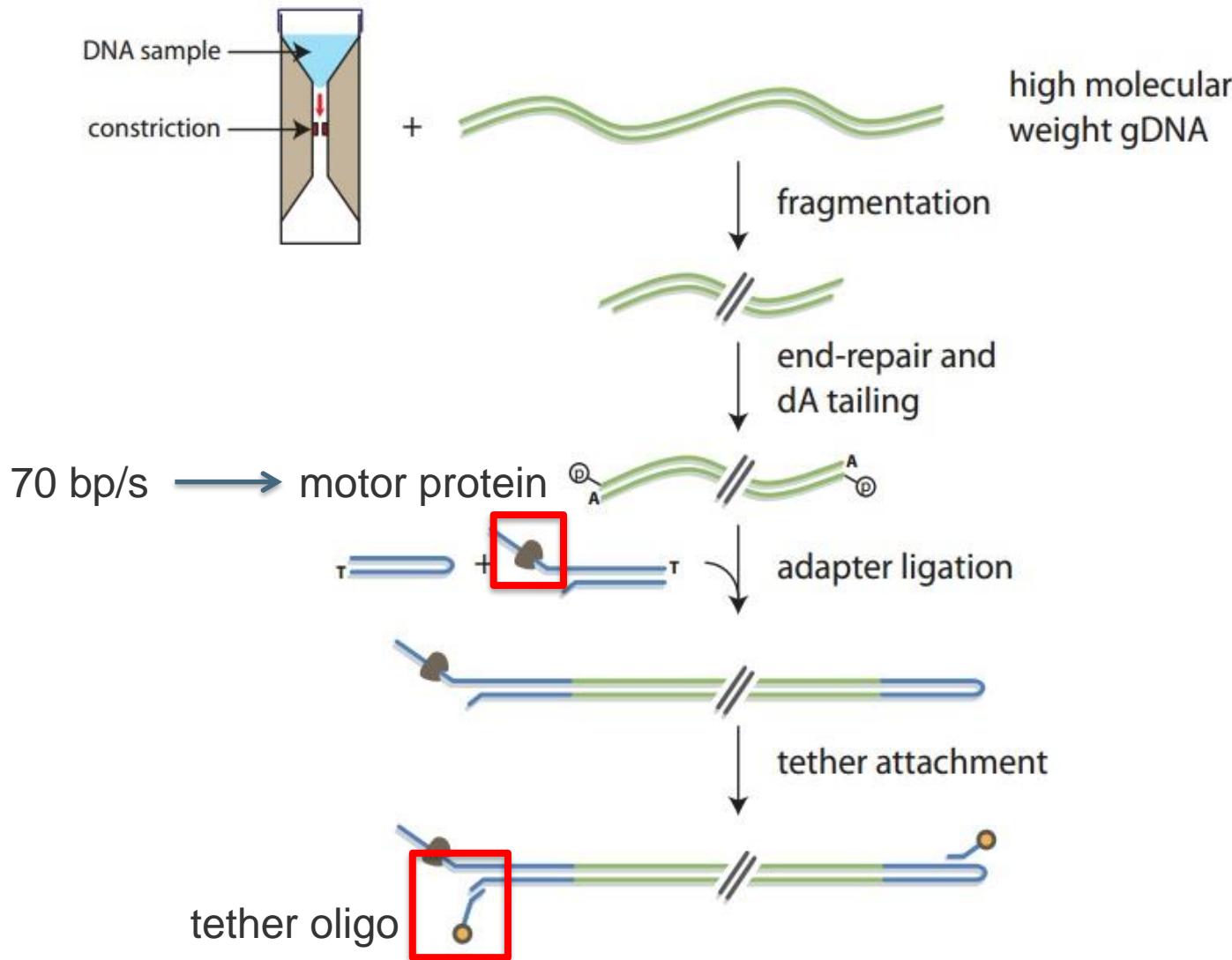


ONT Sequencing

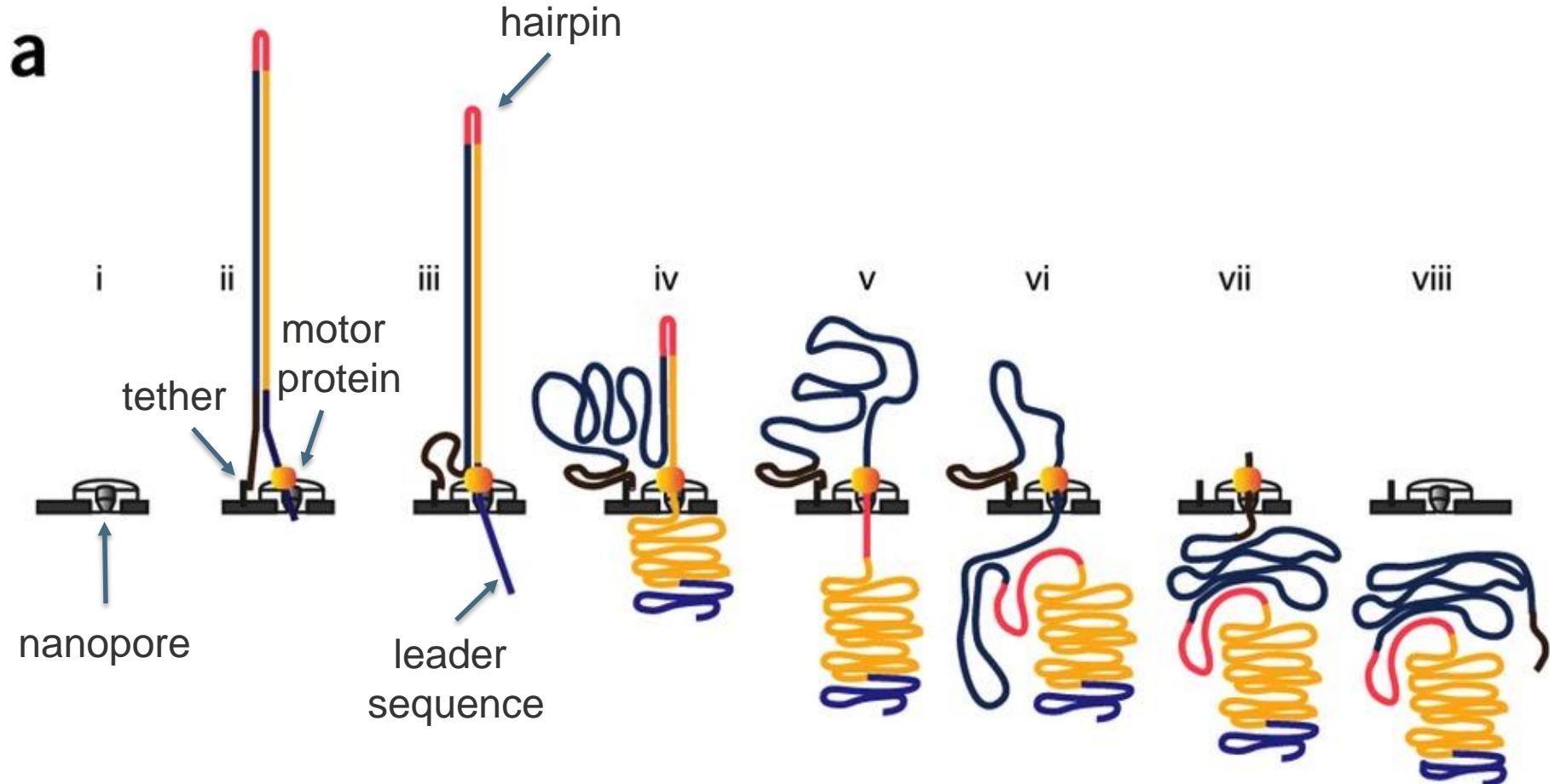


**No polymerase or amplification needed →
as long as DNA intact, long reads possible**

Simple Library Preparation



Nanopore Sequencing



ONT MinION Hardware

MinION (Early Access Version)

- 512 pores
- <1Gb per run



MKI (2016)

- 512 pores; 1 Gb per run
- Faster, better S:N, longer runs



Source: Oxford Nanopore Technologies

ONT Hardware Updates



VOLTRAX

- Automated sample preparation
- Late 2016



PROMETHION

- >300X throughput than MinION
- JGI early access site in 2016
- 1Tb+ runs?

Upcoming Releases – SmidgION

- Mobile sequencing for iPhone
- 256 Channels, low power mgmt tricks
- Flow cell clips in
- 4-5 hours on phone battery
- 250 MB/hour
- Late 2017 early access



Questions? cgdaum@lbl.gov



The JGI is more than just machines!