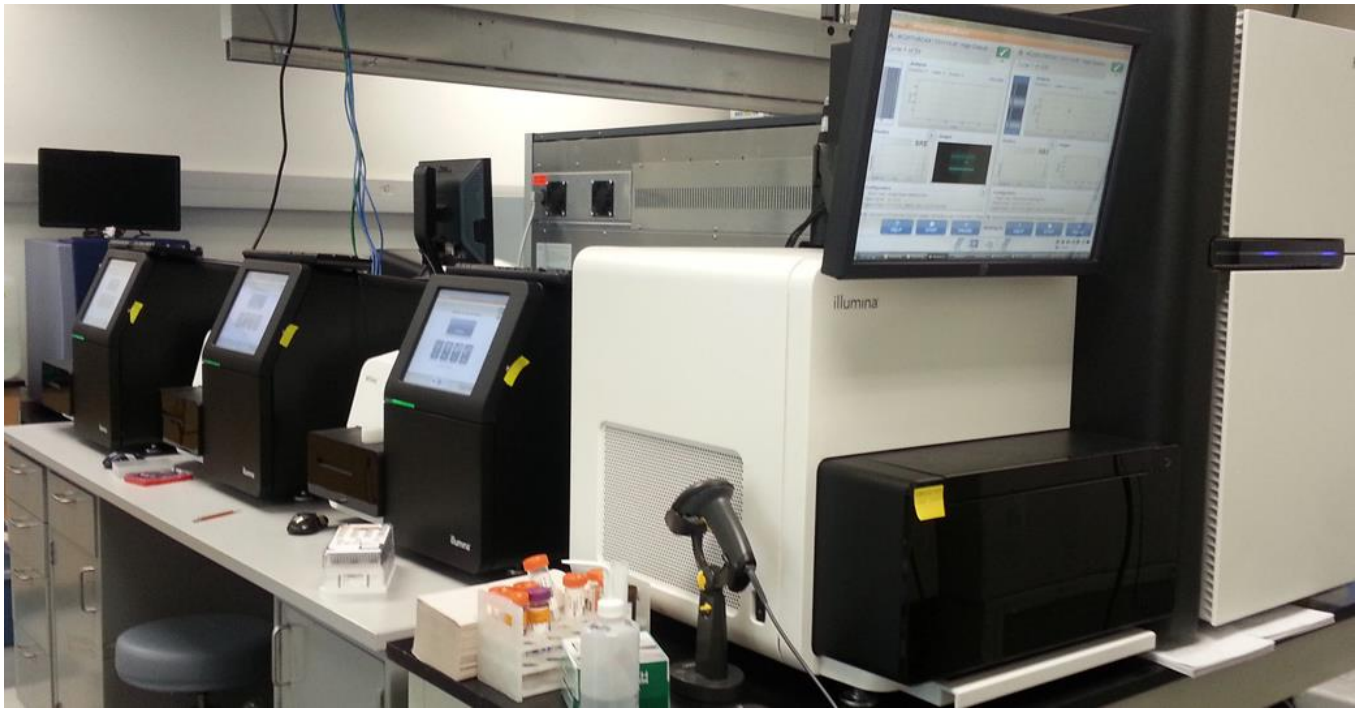
Human Chromosomes. Credit: Jane Ades, NHGRI



The Genome Assembly Workshop

Lutz Froenicke
DNA Technologies & Expression Analysis Cores
UC Davis Genome Center
2018

DNA Technologies & Expression Analysis Cores

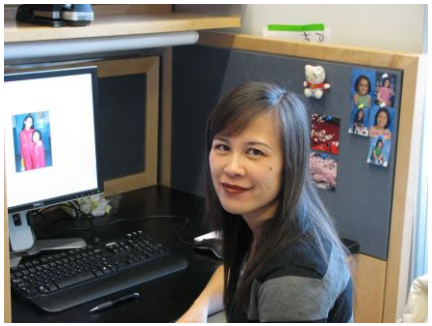
- HT Sequencing Illumina
- Long-Read & Linked Read Sequencing
  PacBio, Oxoford Nanopore, 10X Genomics
- HMW DNA isolation
- Illumina microarray (genotyping)

- Consultations → Experimental Design
  (Bioinformatics Core & DNA Tech Core)

- introducing new technologies to the campus
- shared equipment
- teaching (workshops)

The DNA Tech Core Team

Emily



Oanh

Diana

Siranoosh

Vanessa

Ruta

DNA Tech Genome Assemblv Tools

- 10X Genomics Chromium Genome "linked read"

- PacBio Sequel "super-long reads" >20kb
- PacBio Sequel "high-fidelity-long reads" (Q20,Q30, 10kb & 15kb)

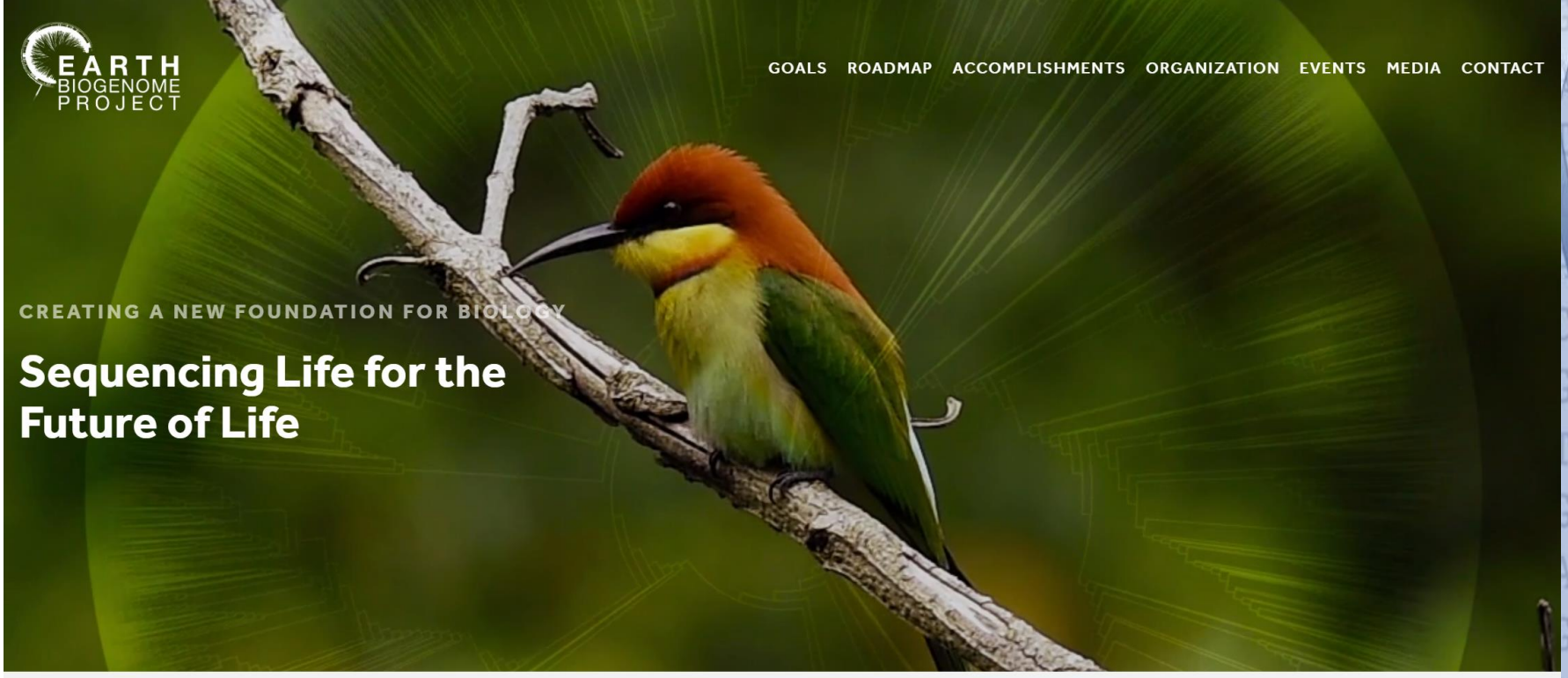
- PromethION Nanopore "super-long reads" >20kb
- MinION Nanopore "ultra-long reads" >100 kb

- Hi-C (chromosome scale genome scaffolding)

Bionano Saphyr at Luo lab, UCD

Optical Genome Mapping

mcluo@ucdavis.edu



CREATING A NEW FOUNDATION FOR BIOLOGY

Sequencing Life for the Future of Life

What is the Earth Biogenome Project?

Powerful advances in genome sequencing technology, informatics, automation, and artificial intelligence, have propelled humankind to the threshold of a new beginning in understanding, utilizing, and conserving biodiversity. For the first time in history, it is possible to efficiently sequence the genomes of all known species, and to use genomics to help discover the remaining 80 to 90 percent of species that are currently hidden from science.

A GRAND CHALLENGE

The Earth BioGenome Project (EBP), a *moonshot* for biology, aims to sequence, catalog and characterize the genomes of all of Earth's eukaryotic biodiversity over a period of ten years.

A GRAND VISION

Create a new foundation for biology to drive solutions for preserving biodiversity and sustaining human societies.

Olga Dudchenko
4 days ago • 1 min

First chromosome-length genome assembly of a snake

Today, we're sharing a chromosome-length genome assembly for the Burmese python, the first chromosome-length assembly (as far as we know) for a snake. The…

145 views Write a comment

Olga Dudchenko
Dec 5 • 2 min

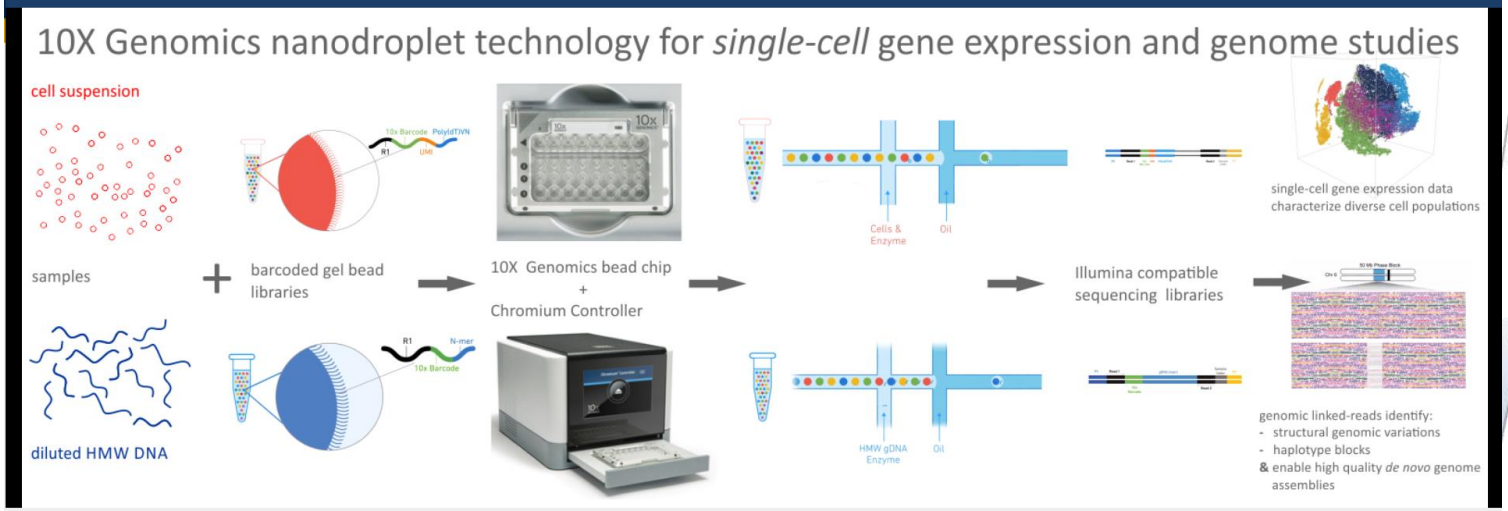
Welcome to the DNA Zoo!

Over a quarter of all assessed species are threatened with extinction [1]. The DNA Zoo is a consortium focused on facilitating conservation efforts through the rapid generation and release of high-quality genomics resources. We believe that these efforts can not only aid…

UCDAVIS Genome Center

DNA TECHNOLOGIES CORE

Expression Analysis Core



Welcome to the DNA Technologies & Expression Analysis Cores

HOLIDAY HOURS: We will be closed on the university holidays (December 24th, 25th and 31st as well as January 1st) but open weekdays otherwise.

We expect difficulties with the receipt of shipments for the week between the holidays. Please arrange shipments to arrive by December 20th.

The DNA Technologies and Expression Analysis Cores at the Genome Center offer high-throughput sequencing, genotyping, and microarray services, as well as training and consultation. Our goal is to enable access to high throughput genome-wide analyses at economical recharge rates, as a functional extension of your laboratory. We operate on the cost-recovery principle. We employ liquid handling robots to minimize sample handling variation and to provide fast turnaround times. We are a designated Campus Research Core Facility.

We offer the two complementary Next Generation Sequencing (NGS) technologies: Illumina sequencing, and PacBio (long read) sequencing, and provide the full spectrum of sequencing options and a wide range of library preparation services for both platforms. Genotyping is performed on the Fluidigm EP1 System for low to medium assay numbers, and Illumina Infinium arrays for high density array SNP genotyping. Gene expression analysis is carried out by RNA-seq on HiSeq sequencers. **Single-cell** transcriptome (high-throughput single-cell gene expression profiling) and genome analyses (linked-read whole genome sequencing) are enabled by our 10X Genomics Chromium System.

We offer annual Illumina and PacBio sequencing library preparation workshops, free consultations on project considerations and

Recent Posts

- PacBio News – High Fidelity Long Reads – Sequel V3 Chemistry
- Holiday Hours
- New online equipment reservation system ready-to go!
- Monthly Illumina Office Hours
- We now offer Nanopore Sequencing on the PromethION, HMW-DNA isolation & Hi-C

Latest Tweets

- Looks like good coffee! (great coffee genome data with a read lengths N50 of ~30 kb) https://t.co/dZcgCvfd8F, Dec 15
- "..... Single-cell studies revealed that transcription occurs in discontinuous bursts, suggesting that features of… https://t.co/GBs8vjCQuJ, Dec 14
- I am currently writing "The Art of Stupidity". Hope this is not taken yet. https://t.co/MCm7kxCmFu, Dec 14
- BTW, the TIN score is calculated by the RSeQC from the alignments and is designed to be analogous to the

long-read and linked-read sequencing for high quality genome assemblies

Experience

150+ species/varieties with PacBio

110+ species with 10x Genomics

as of December 2018

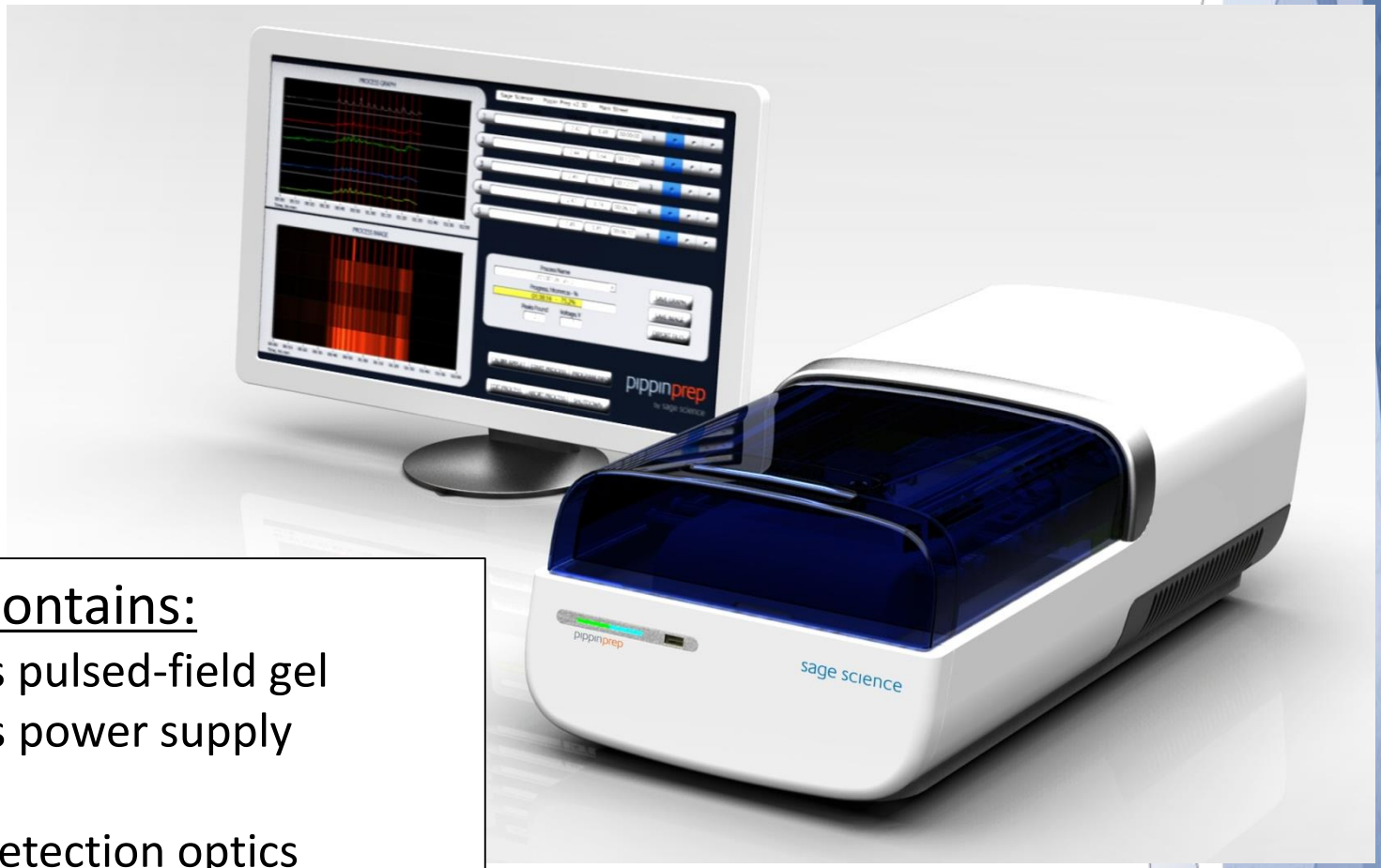
DNA Quality !!!

- HMW DNA isolation
  - Physical damage (PFGE image is not fully informative)
  - Chemical damage
  - Chemical contamination

  - Sample specific protocols?
  - Nuclei isolation, agarose plugs
  - Cell culture?
  - Rescue efforts (BluePippin; DNA damage reapir) tend to have minimal impact

The Blue Pippin Prep System

Automated Preparative Gel Electrophoresis for NGS

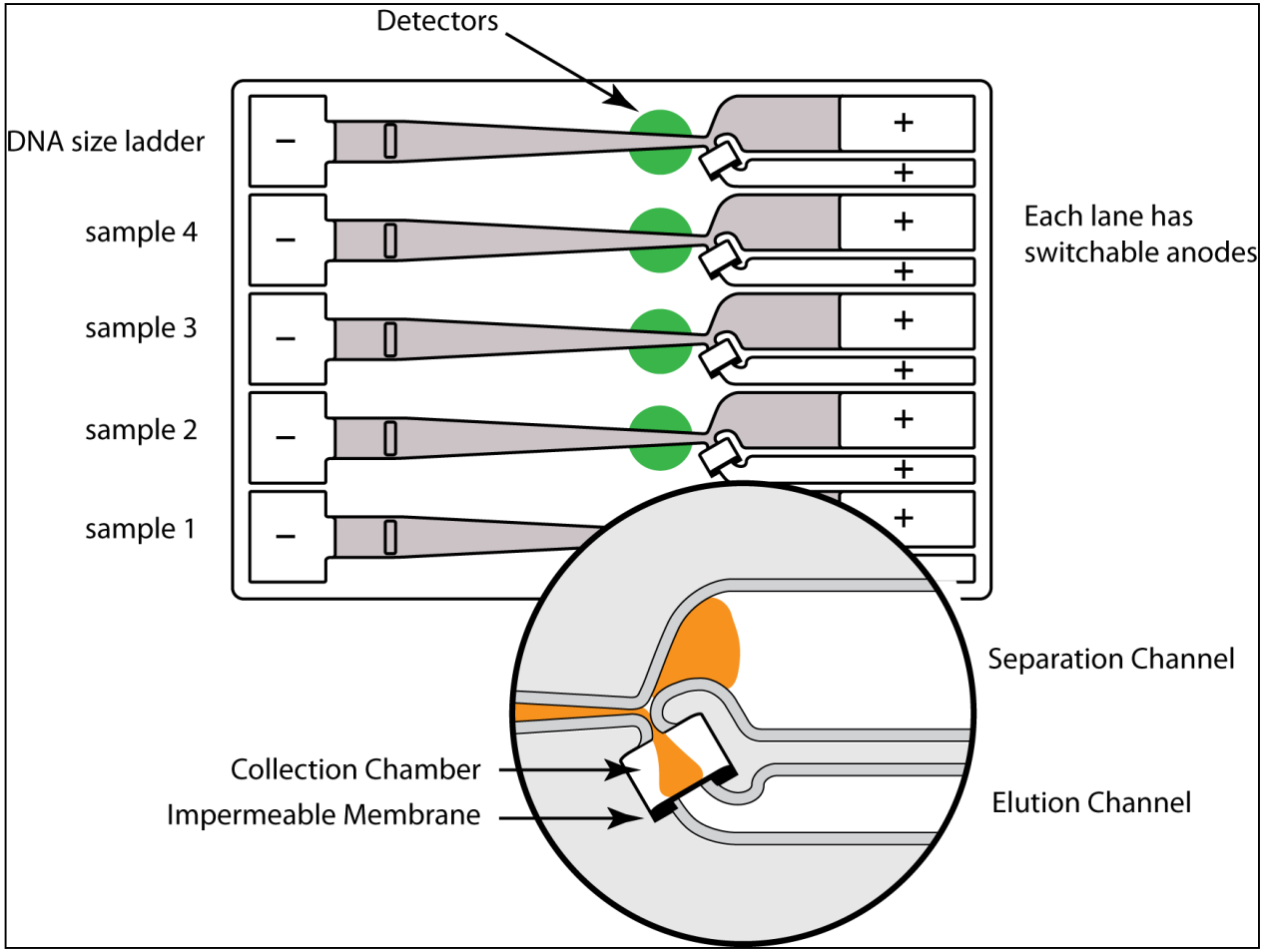


Instrument contains:

Electrophoresis pulsed-field gel electrophoresis power supply
Electrode array
Fluorescence detection optics
Single-board PC with control software

The BluePippin Prep System

Automated Preparative Gel Electrophoresis for NGS



Strengths and weaknesses

(diploid genomes)

	10X	Sequel SLR	Sequel HFLR	PromethION
Accuracy – single read	4	1	4	2
Accuracy - consensus	4	4	4	3
Contiguity	1-3 ?	3-4	3	4
Yield/$	4	4	2	4

HFLR: high-fidelity-long-reads

- In many case the resulting assembly quality will depend on the DNA sample quality as well genome organization (repeat content/length).
- Combining the strengths of two or three technologies seems most promising

Coffee

2017:
arabica 1.3 Gb
genome
Medrano et al.

PacBio &
ChICAGO scaffolding
Scaffold N50 = 2.24 Mb
Contig N50= 1.31 Mb

2015:
robusta 0.7 Gb

UCDAVIS

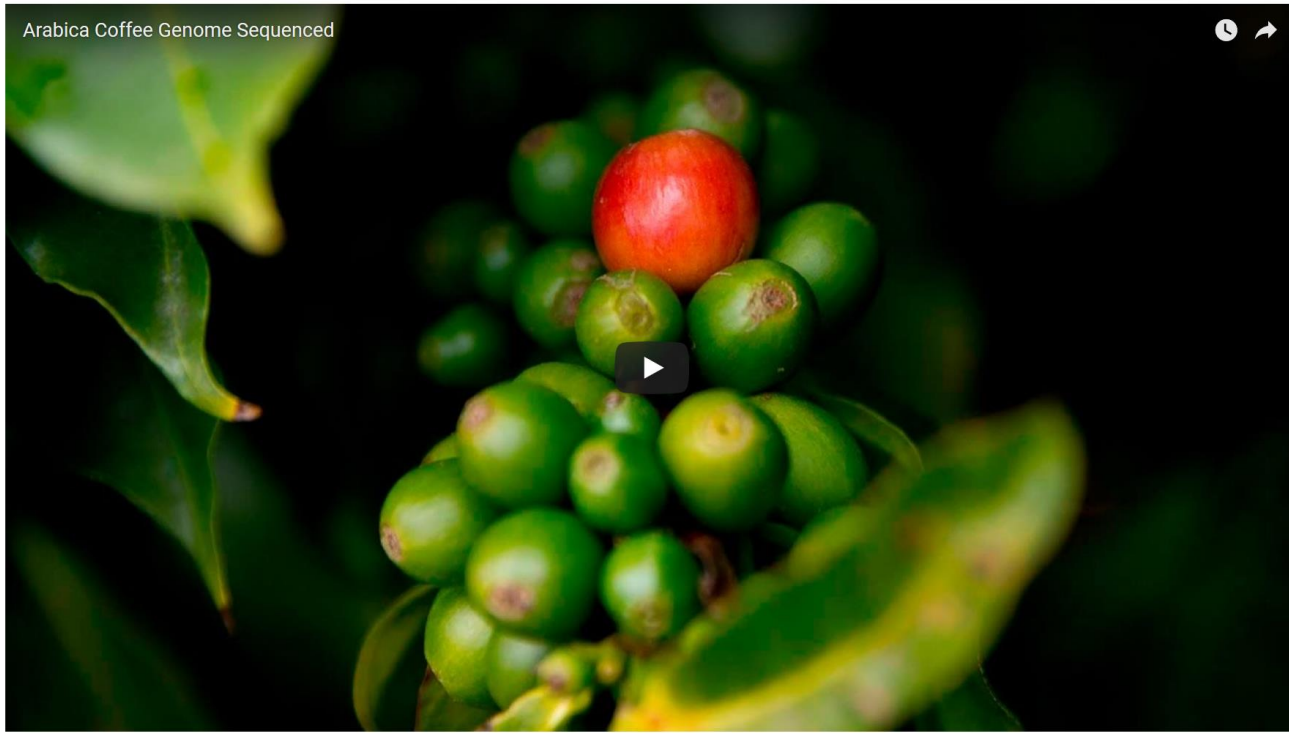
ABOUT US ADMISSIONS ACADEMICS RESEARCH

Home > News > Arabica Coffee Genome Sequenced

Arabica Coffee Genome Sequenced

Coincides With Birth of California-Grown Specialty Coffee Industry

By Pat Bailey on January 13, 2017 in Food & Agriculture



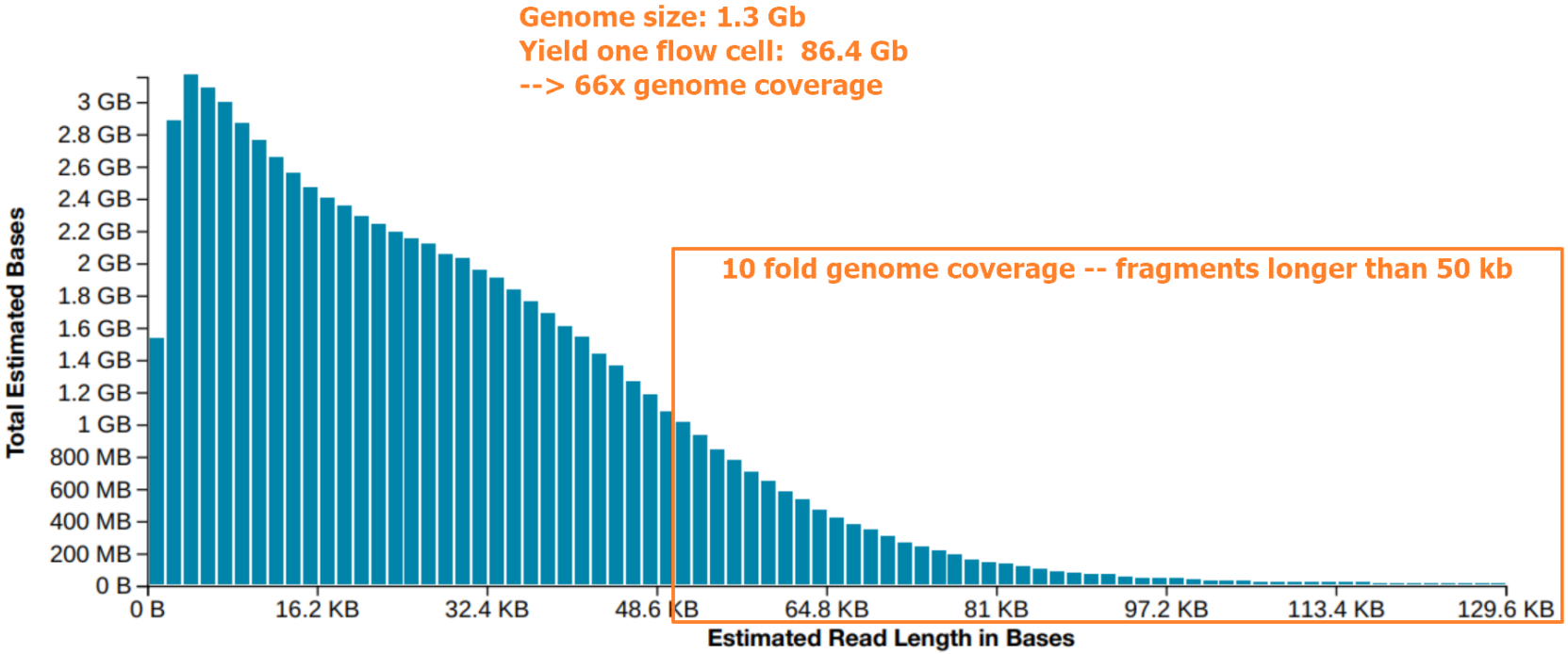
The first public genome sequence for *Coffea arabica*, the species responsible for more than 70 percent of global coffee production, was released today by researchers at the University of California, Davis.

Quick Summary

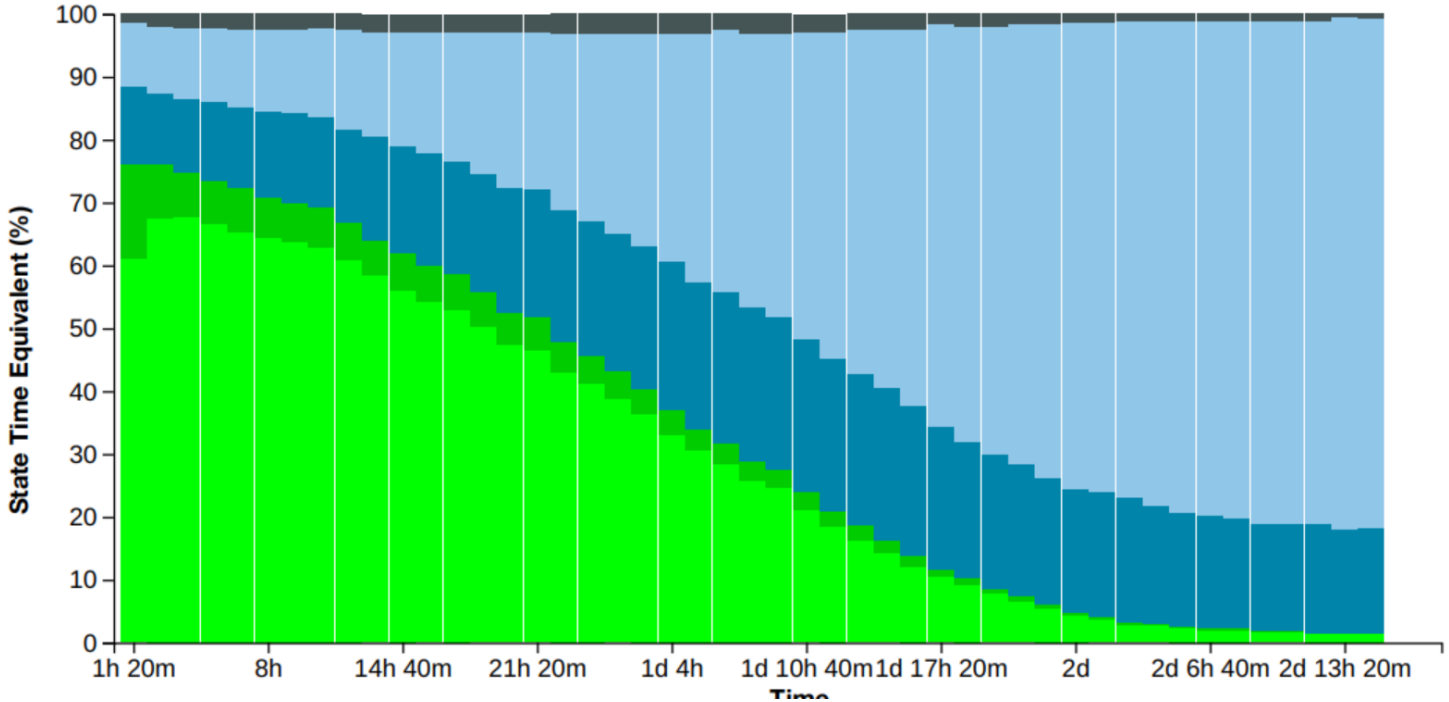
› **Will help develop disease-resistant varieties adaptable to climate**

PromethION read lengths

- RL histogram *arabica*

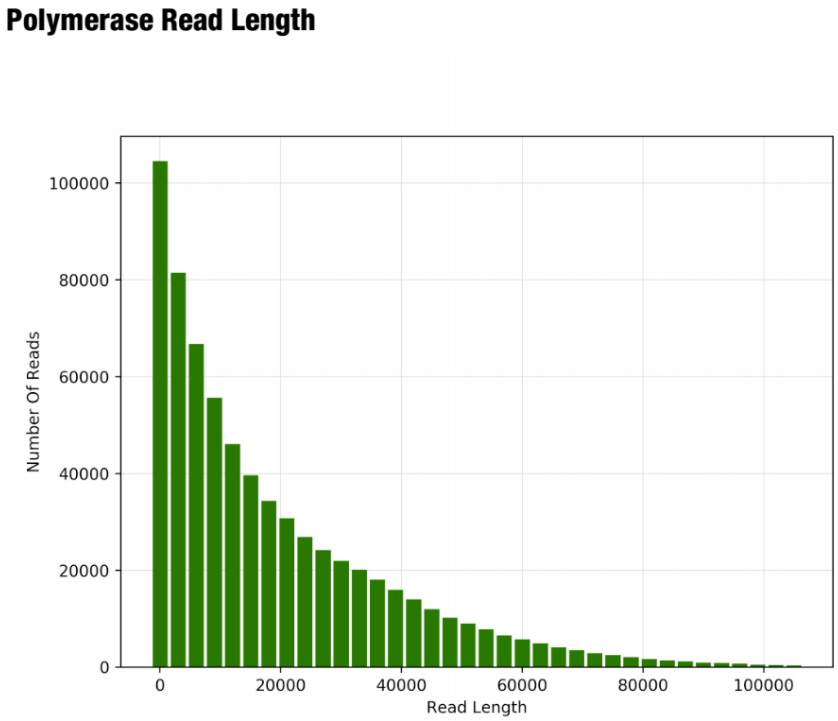


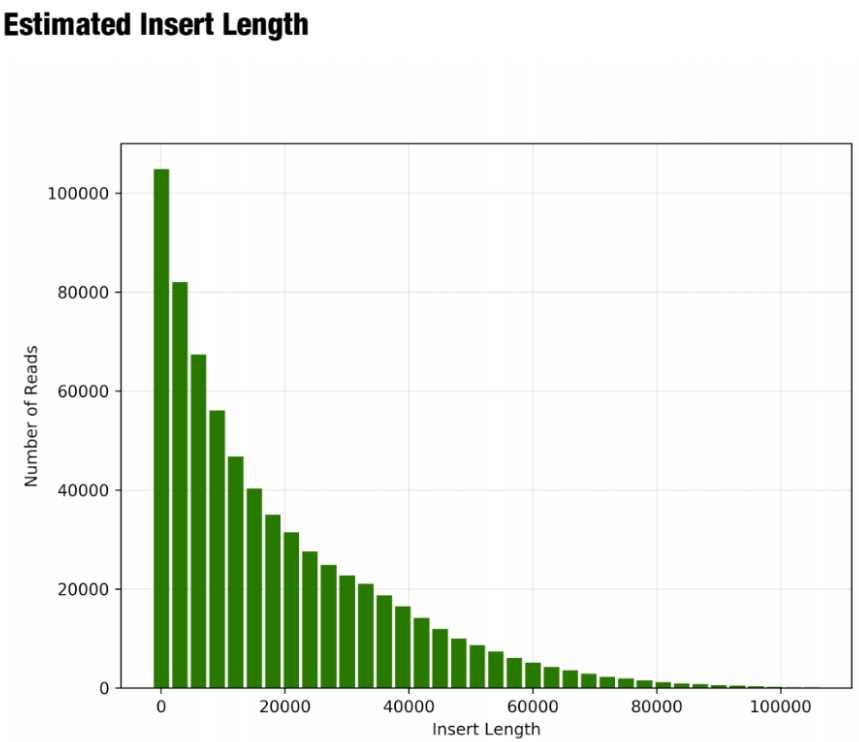
Yield over time



PacBio V3 chemistry 10 hour movies

- Monkey cell line



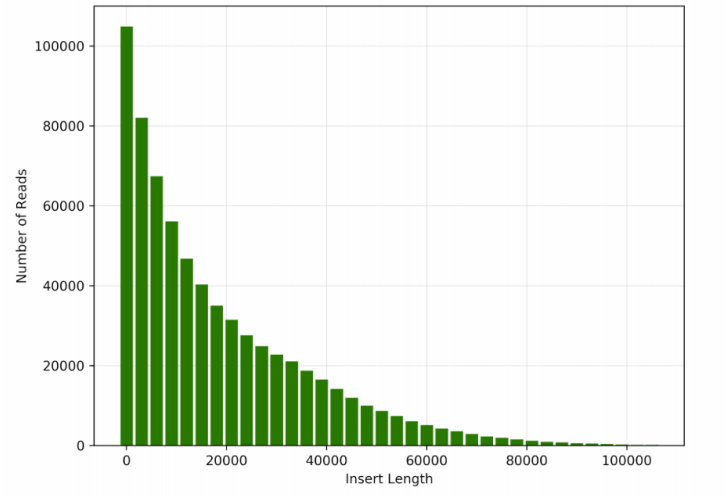


PacBio V3 chemistry 10 hour movies (not yet V2 library prep kit; December 2018)

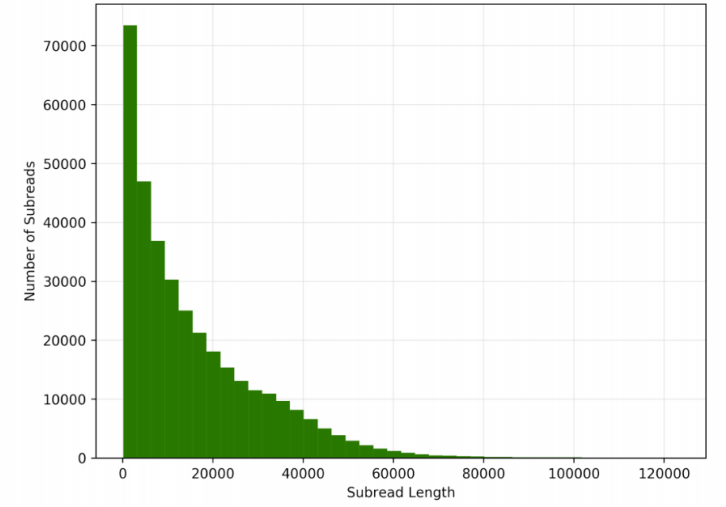
Monkey cell line
Polym.R. N50 35.3 kb
Subread N50 33 kb

Hydra
Polym.R. N50 29,5 kb
Subread N50 27.5 kb

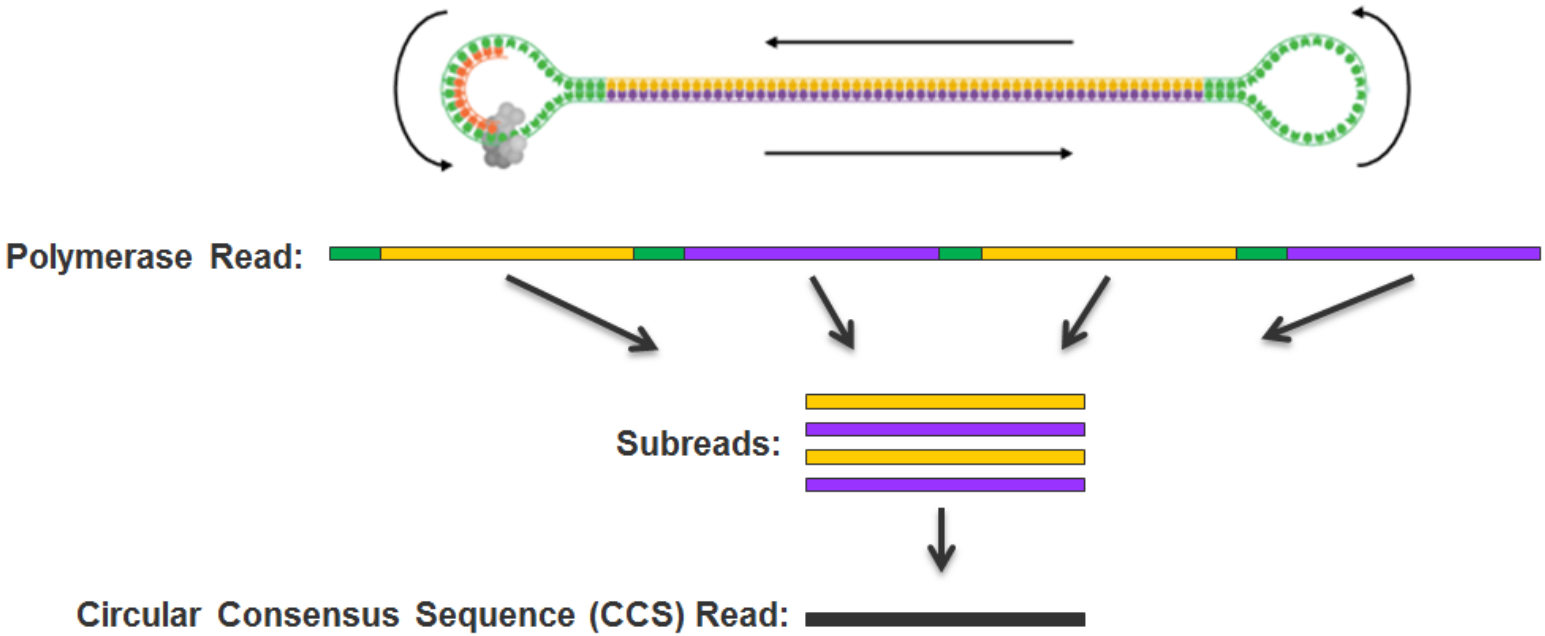
Estimated Insert Length



Subread Length



SMRT-bell adapters circular sequencing



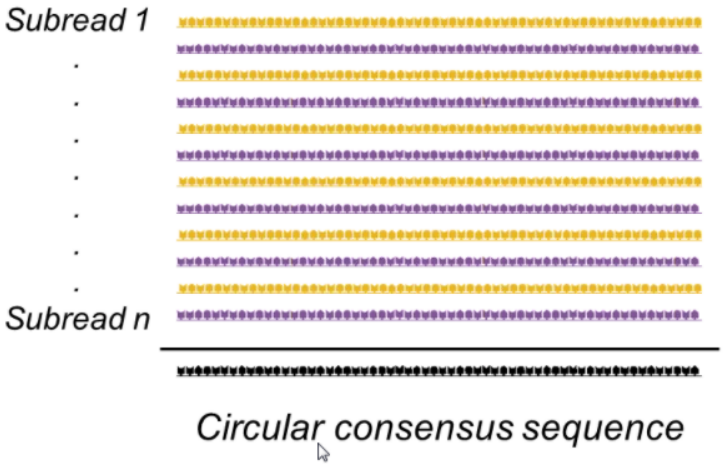
Source: PacBio

NEW CHEMISTRY PERFORMANCE

	v 6.0	v 5.1
Shorter-insert libraries		
- Throughput per SMRT Cell:	**up to 50 Gb**	up to 20 Gb
- Average read length:	**up to 100 kb**	up to 40 kb

Pre-extension:

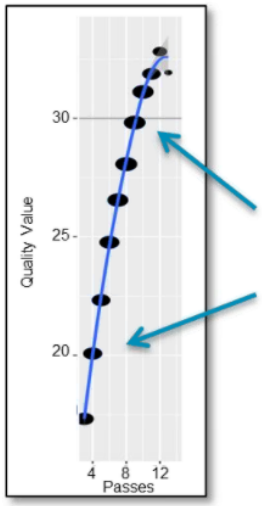
- Expedite polymerase into rolling circle synthesis
- Ensure undamaged template



Source: PacBio

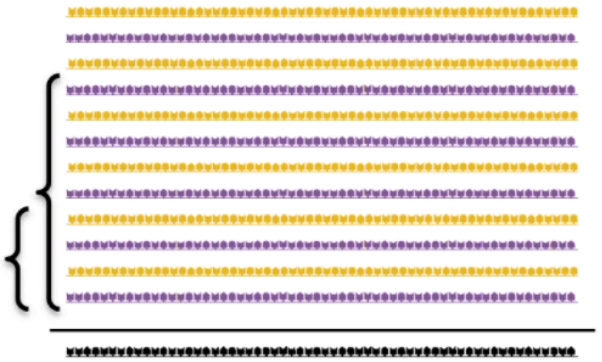
NEW CHEMISTRY PERFORMANCE

Shorter-insert libraries	**v 6.0**	v 5.1
- Throughput per SMRT Cell:	**up to 50 Gb**	up to 20 Gb
- Average read length:	**up to 100 kb**	up to 40 kb



9 passes for Q30 (99.9%)

4 passes for Q20 (99%)

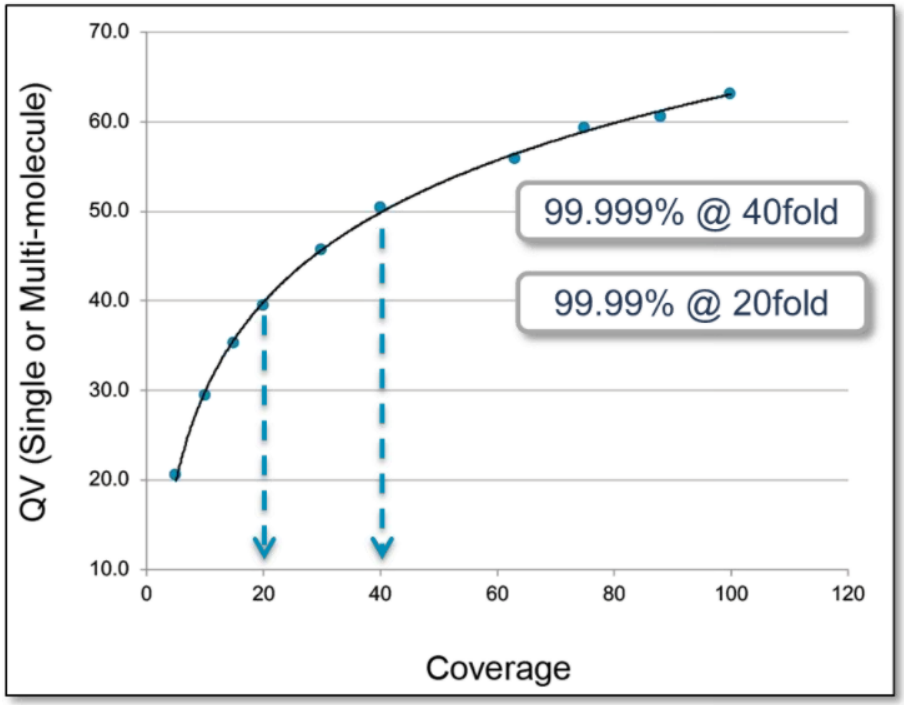


Circular consensus sequence

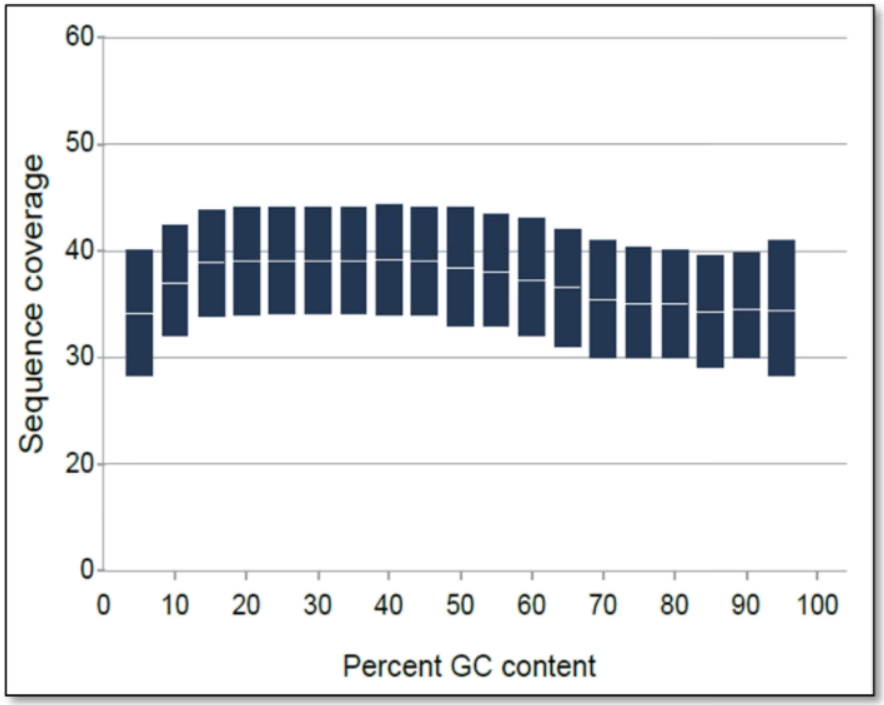
Source: PacBio

NEW CHEMISTRY PERFORMANCE

Consensus accuracy:



Minimal sequence GC% and complexity bias:



Source: PacBio

NEW CHEMISTRY PERFORMANCE

Shorter-insert libraries	v 6.0	v 5.1
- Throughput per SMRT Cell:	**up to 50 Gb**	up to 20 Gb
- Average read length:	**up to 100 kb**	up to 40 kb

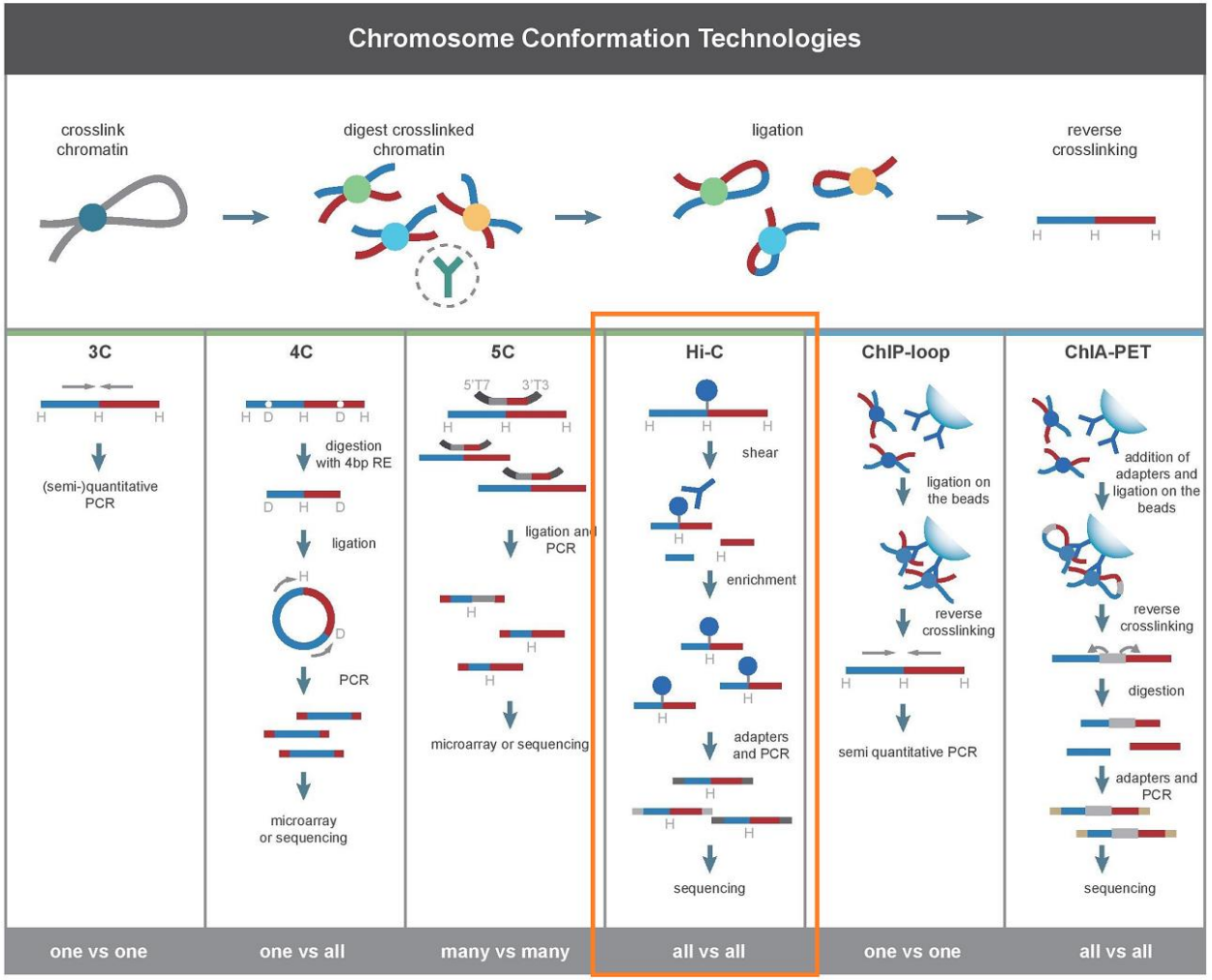
Insert Size	Movie time (hrs)	# Q20 CCS reads	# Q30 CCS reads
1,000	10	570,000	525,000
2,000	10	500,000	400,000
5,000	20	400,000	300,000
10,000	**20**	**250,000**	**125,000**

2.5 Gb Q20 data

1.2 Gb Q30 data

Source: PacBio

Hi-C chromosome-scale scaffolding



source: Wikipedia

Future's so bright

Thank you!

Let’s get started!