Solanaceae Coordinated Agricultural Project

Next generation sequencing

Allen Van Deynze

UC Davis

November 16th, 2010

USDA United States Department of Agriculture National Institute of Food and Agriculture

Sol CAP

Marker development considerations

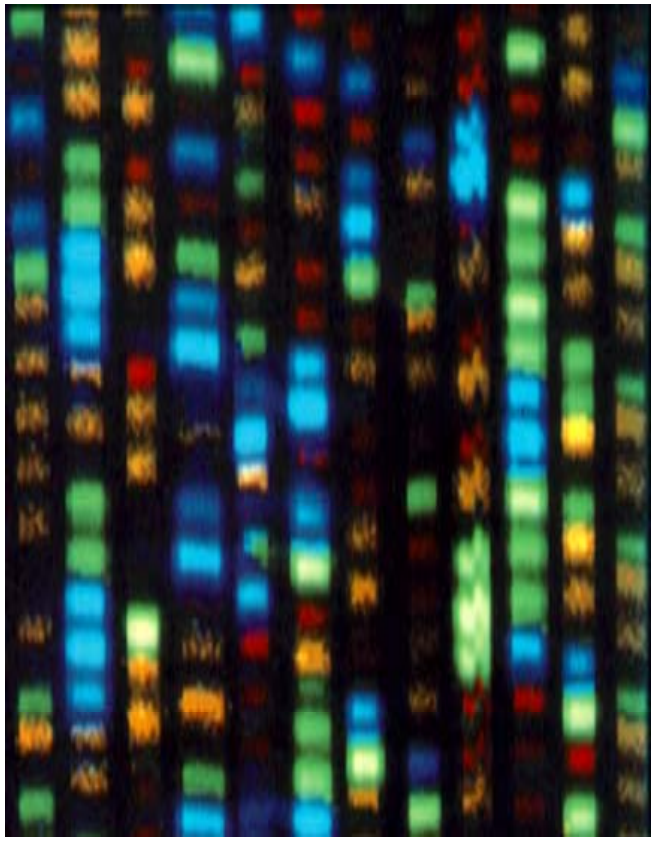
- How to sequence?
- What part of the DNA to sequence?

Talk 2

- What lines to sequence?
- How many lines to sequence?

Sequencing DNA

The goal of sequencing DNA is to tell the order of the bases, or nucleotides, that form the inside of the double-helix molecule.

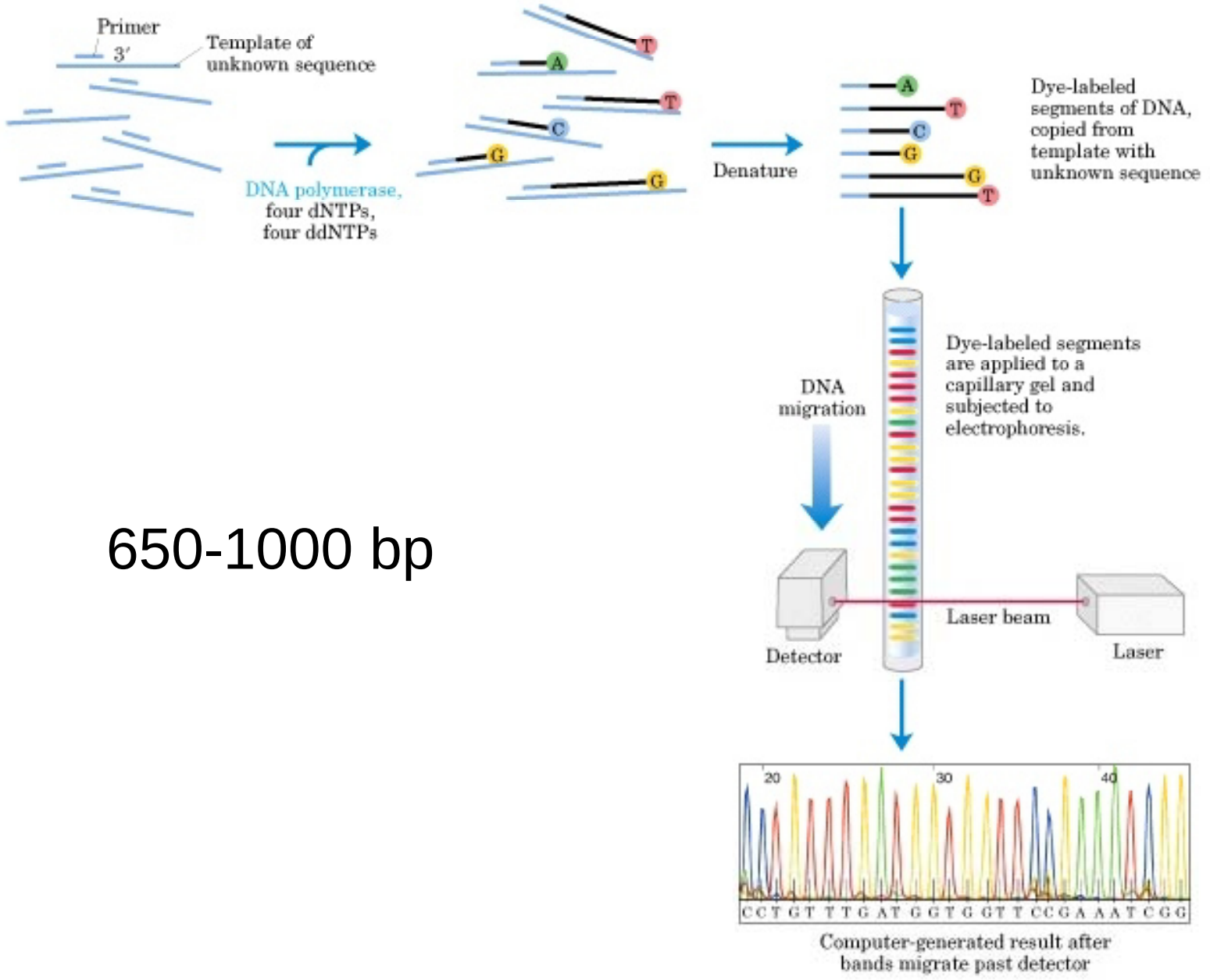


High throughput sequencing methods

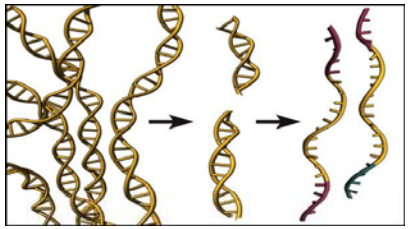
- Sanger/Dideoxy
- 2nd generation (NextGen)
- 3rd generation

Sanger Dideoxy DNA sequencing

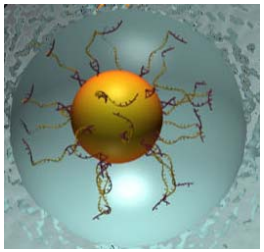
650-1000 bp



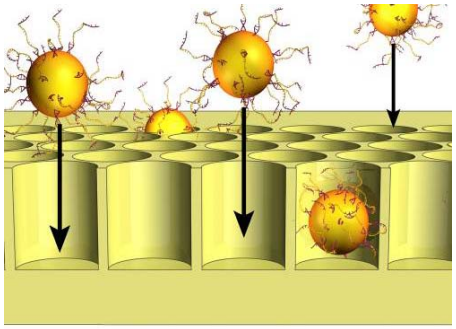
454-Pyrosequencing



Construct Single stranded adaptor ligated DNA

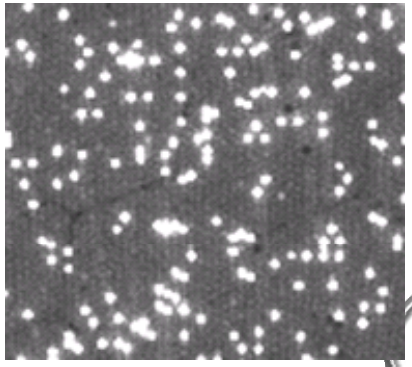


Perform emulsion PCR



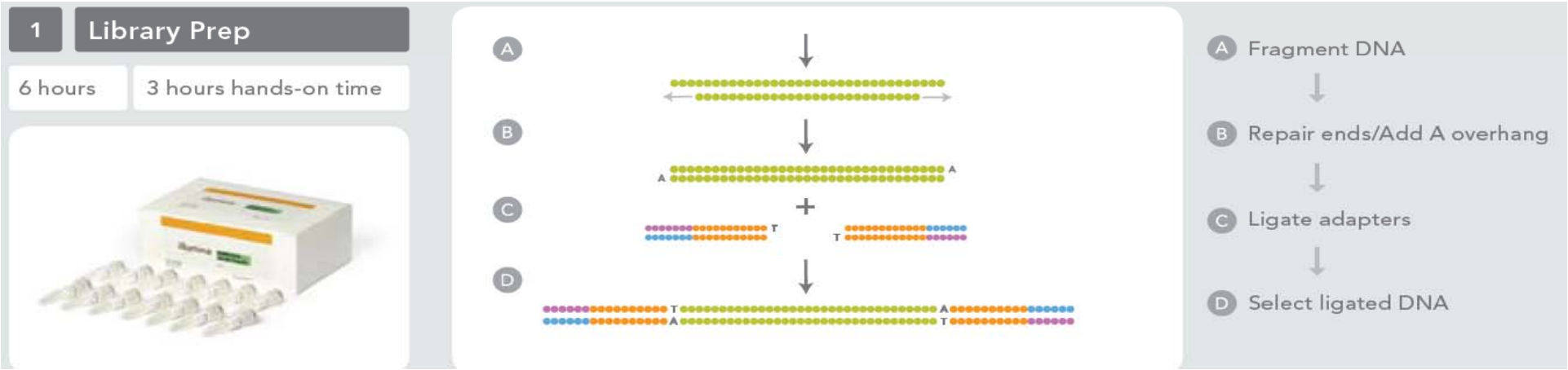
Depositing DNA Beads into the PicoTiter™Plate

- Sequencing by Synthesis:
  - Simultaneous sequencing of the entire genome in hundreds of thousands of picoliter-size wells
  - Pyrophosphate signal generation



Solexa/Illumina Sequencing

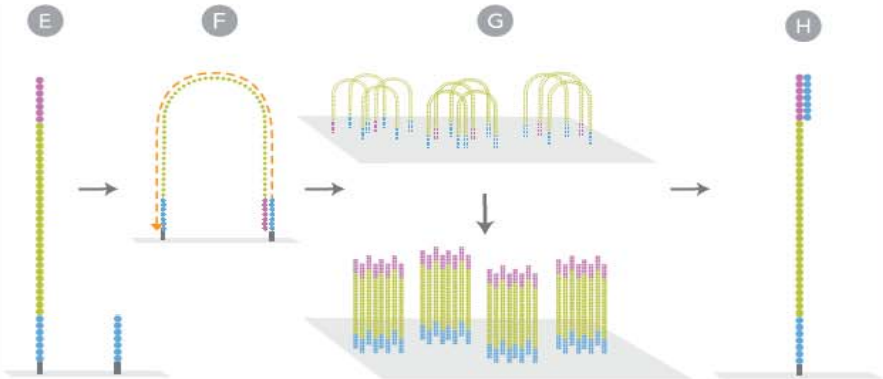
- Sequencing by synthesis (not chain termination)
- Generate up to 100 Gb per run



2 Cluster Generation

5 hours

30 min. hands-on time
(1–8 Samples)

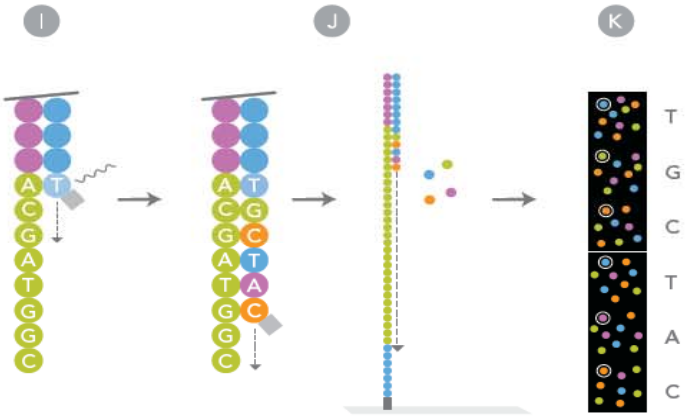


- E Attach DNA to flow cell
- F Perform bridge amplification
- G Generate clusters
- H Anneal sequencing primer

3 Sequencing

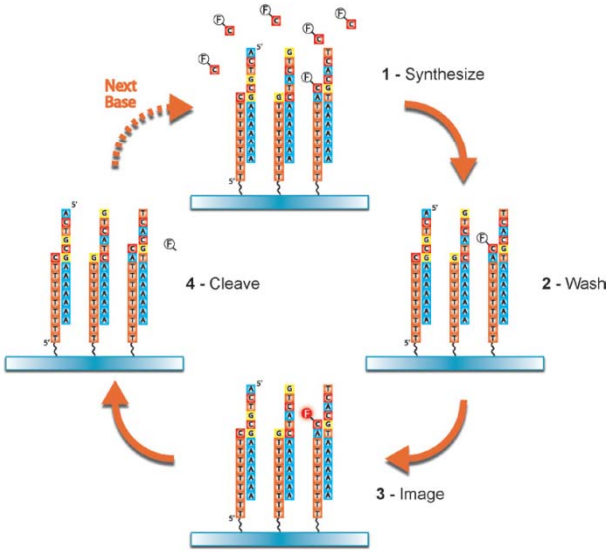
2–3 days (single-read)
4–6 days (paired-end)

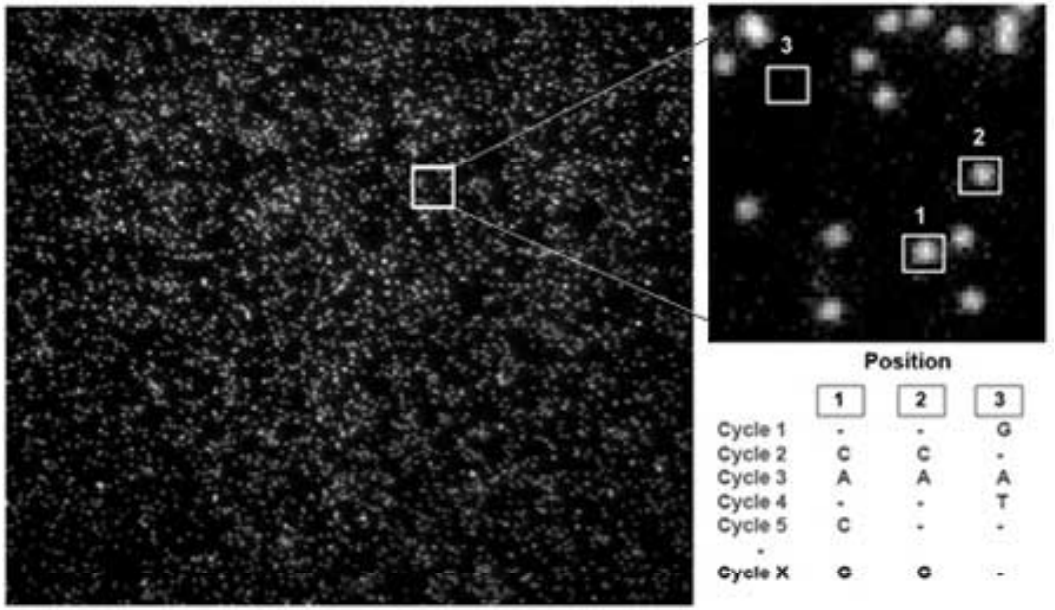
30 min. hands-on time (1–8 Samples)



- I Extend first base, read, and deblock
- J Repeat step above to extend strand
- K Generate base calls

Helicos-True Single Molecule Sequencing (tSMS)™

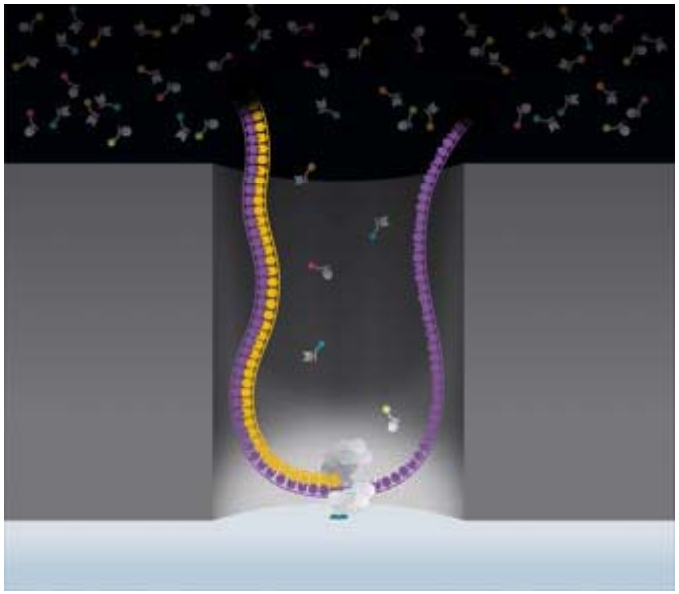




	Sanger Sequencing	"Next Generation" Sequencing	Helicos True Single Molecule Sequencing
Information Capacity:	**100's** of reads per experiment	**100,000,000's** of reads per experiment	**1,000,000,000's of reads per experiment**
Scalability of Sample Preparation:	A few at a time	A few at a time	**Hundreds at a time Easily automated**
Amplification:	Required	Required	**No amplification True direct DNA measurement**
Accuracy:	*Analog base calls* Sequencing biases Not quantitative	*Analog base calls* Sequencing biases Not quantitative	***Digital base call*** **No amplification biases Digital quantitation**

Single Molecule Real Time sequencing

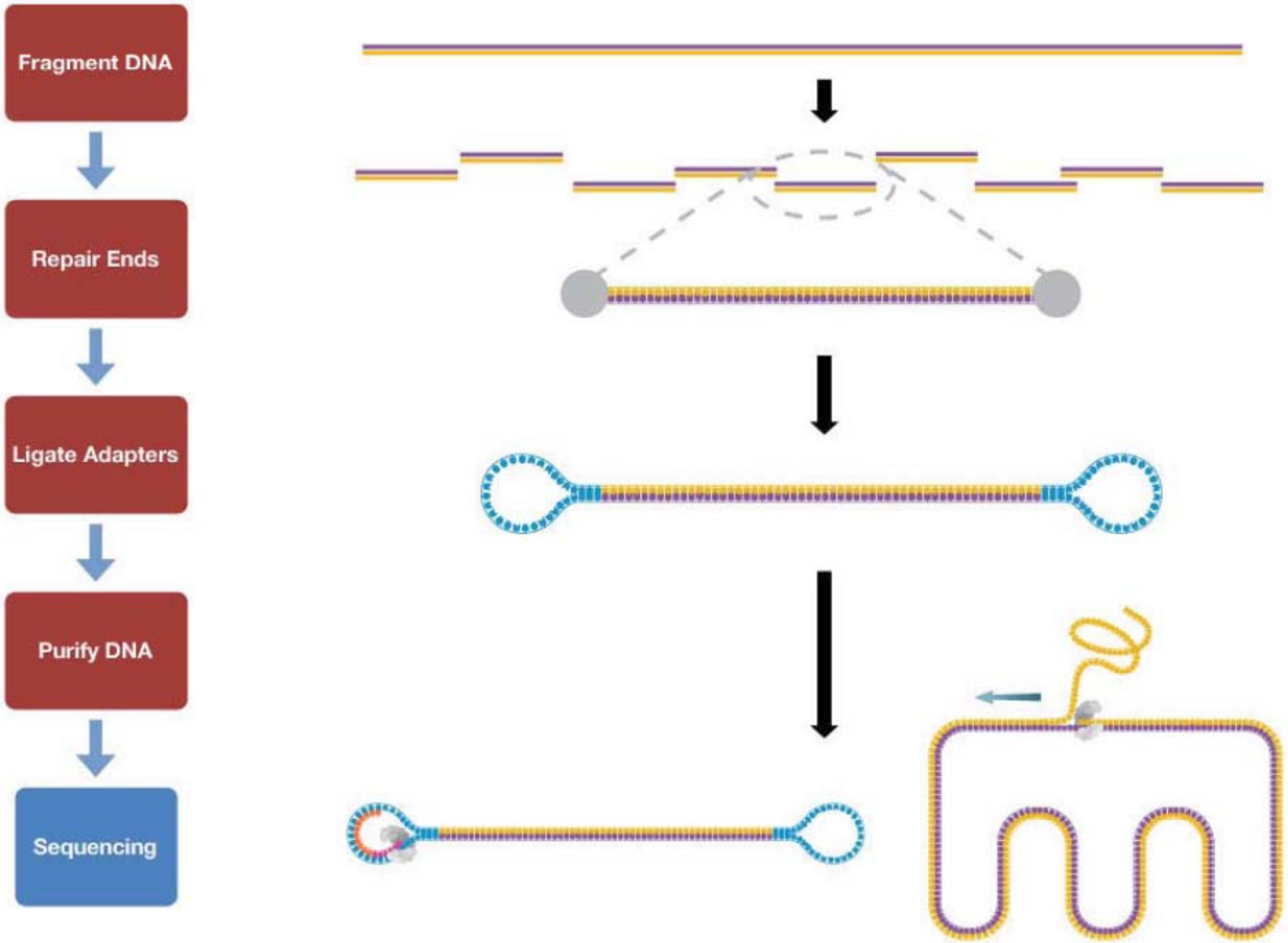
PACIFIC BIOSCIENCES™



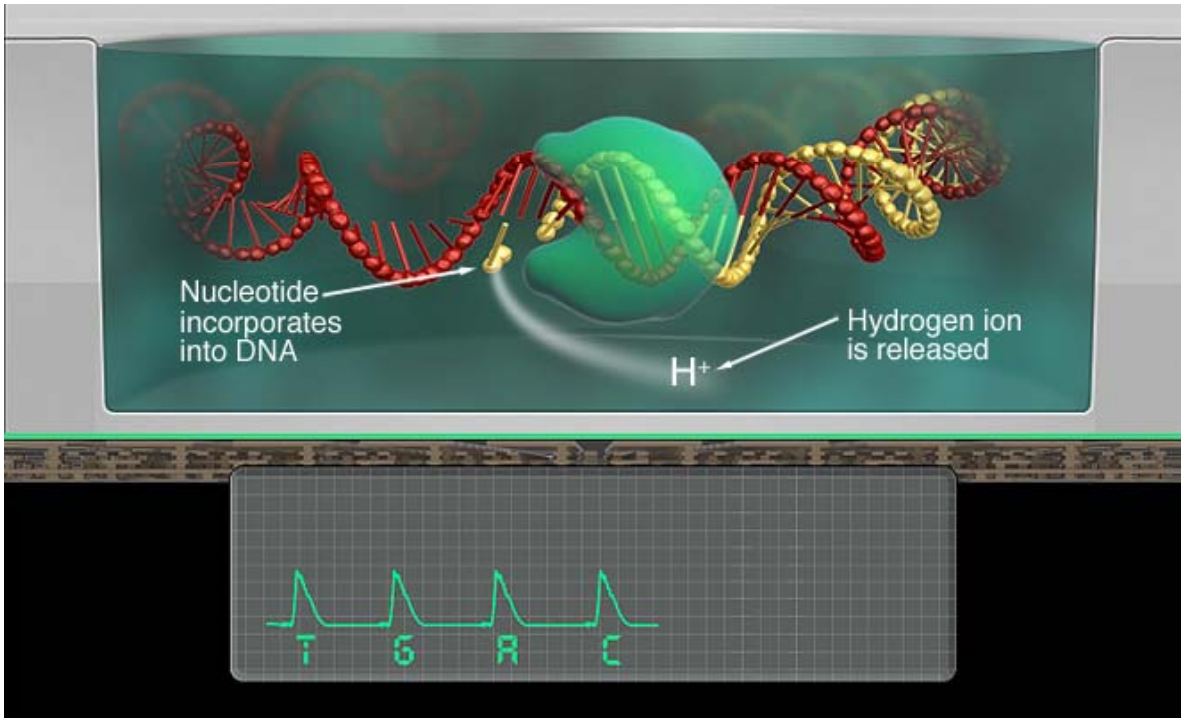
10 bp/sec

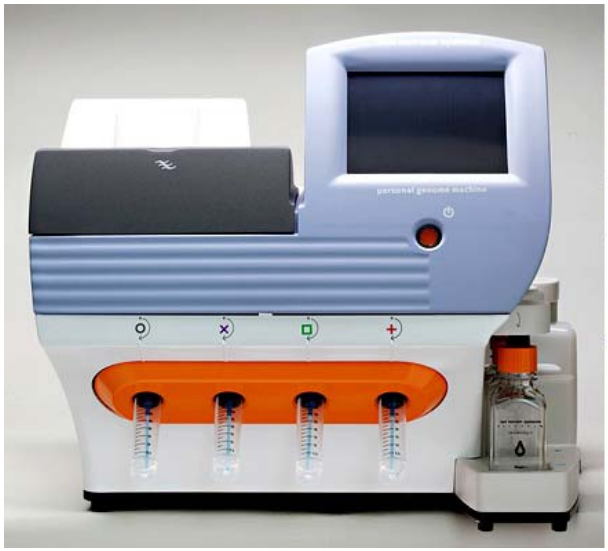
Step 1: Fluorescent phospholinked labeled nucleotides are introduced into the ZMW.
Step 2: The base being incorporated is held in the detection volume for tens of milliseconds, producing a bright flash of light.
Step 3: The phosphate chain is cleaved, releasing the attached dye molecule.
Step 4-5: The process repeats.

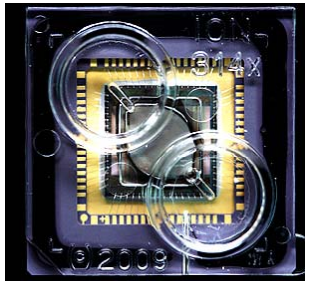
SMRT™ sequencing sample preparation workflow



Ion Torrent







In nature, when a nucleotide is incorporated into a strand of DNA by a polymerase, a hydrogen ion is released as a byproduct.

Sequencing technology 2010

	MB/run	Cost/MB	Length of reads (bp)
Sanger	0.29	4,333	700
Roche 454	180	55.56	175-450
Illumina	20,000	0.50	30-125
Illumina 2010	100,000	0.10	85-100
Helicos	500,000	0.02?	25
Ion Torrent	?	?	25
Pacific Bio	1,000	?	>1000

2010- 10-50 faster..and cheaper

Solanaceae Coordinated Agricultural Project

Next generation sequencing

Allen Van Deynze

UC Davis

November 16th, 2010

USDA United States Department of Agriculture National Institute of Food and Agriculture

Sol CAP

Marker development considerations

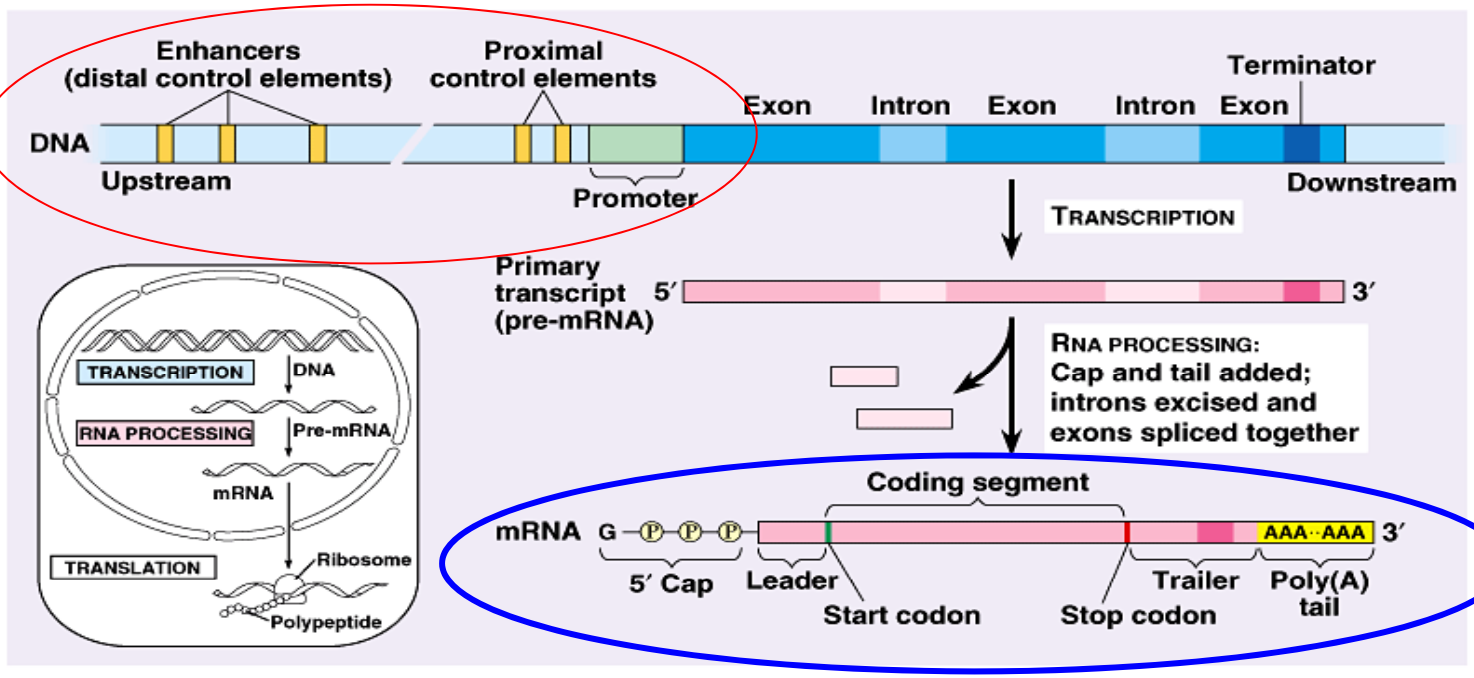
- How to sequence?
- What part of the DNA to sequence?

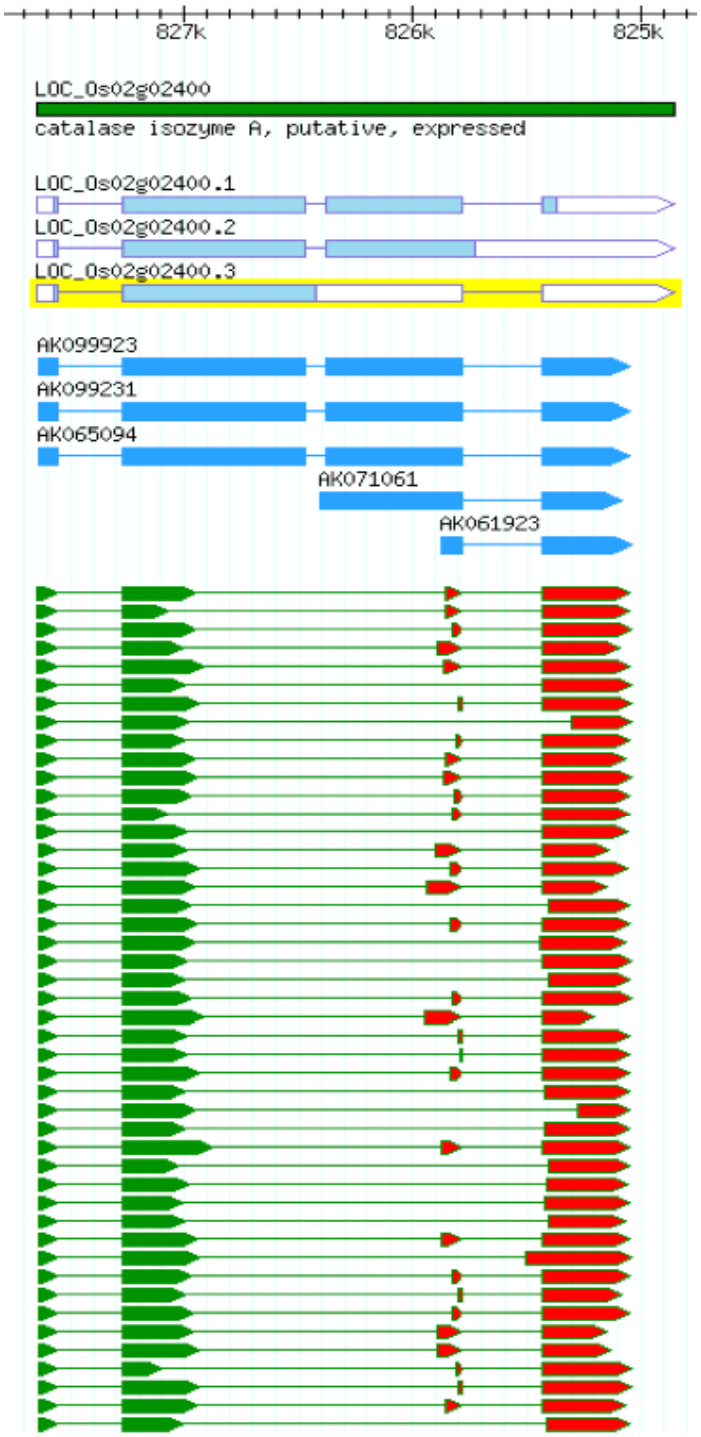
Talk 2

- What lines to sequence?
- How many lines to sequence?

Where to sequence?
Eukaryotic Genomes and Gene Structures

Gene | Intergenic Region | Gene | Intergenic Region | Gene





Locus/Gene

Gene models

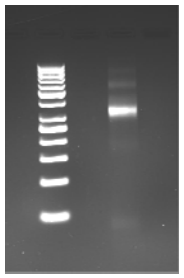
Full length cDNAs

Expressed Sequence Tags

Transcriptome sequencing Illumina

Library creation/QC

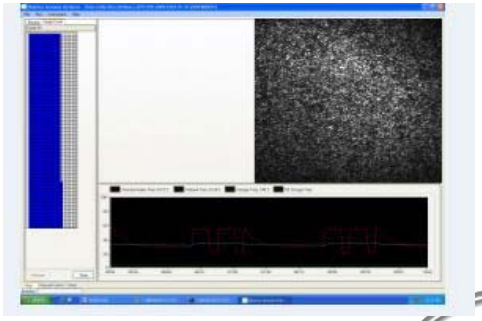
Root
Leaf
Flower
Fruit



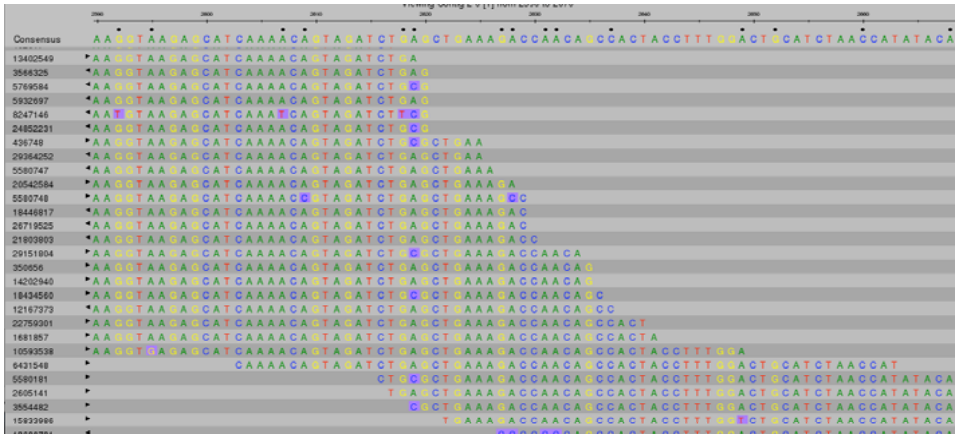
350

GAII sequencing
(single and paired end)

Data Collection

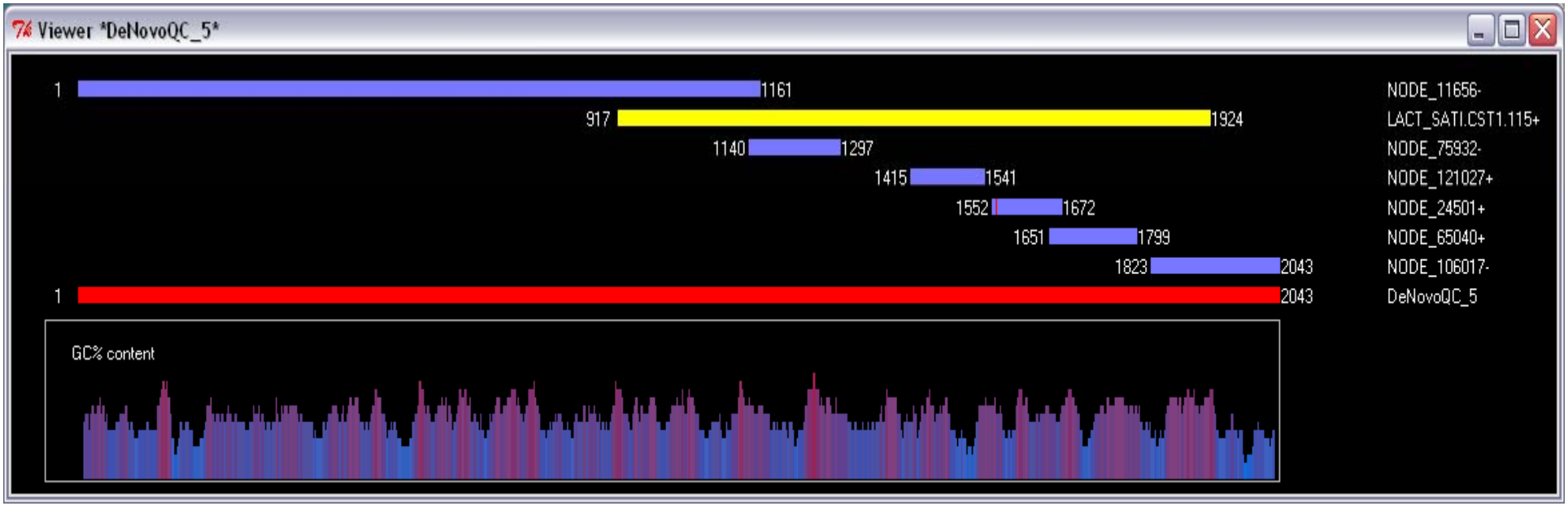


Assembly



Analysis: transcriptome complexity
SNP calling/validation

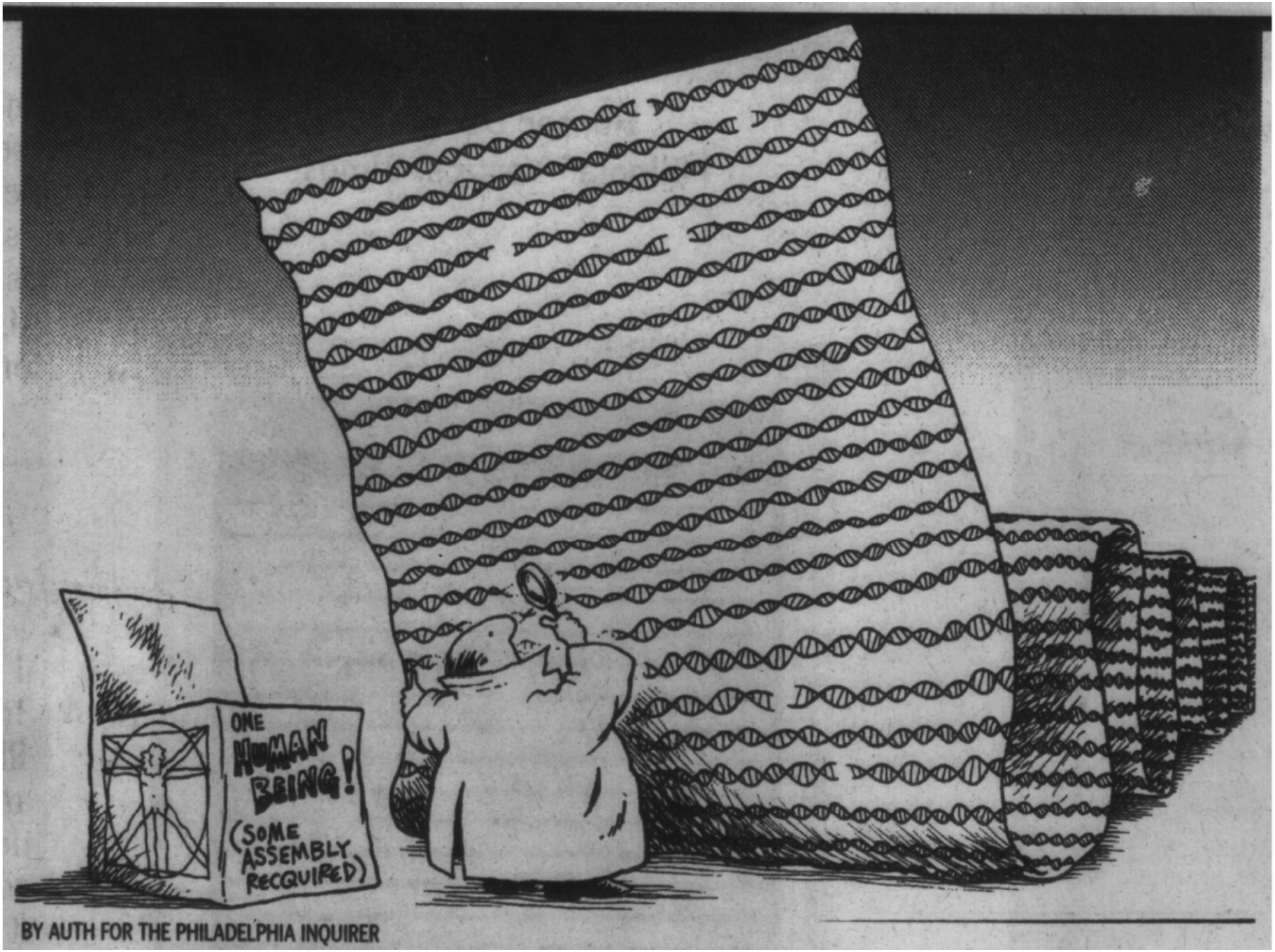
Sequencing all of the EST



Sequencing beyond ESTs

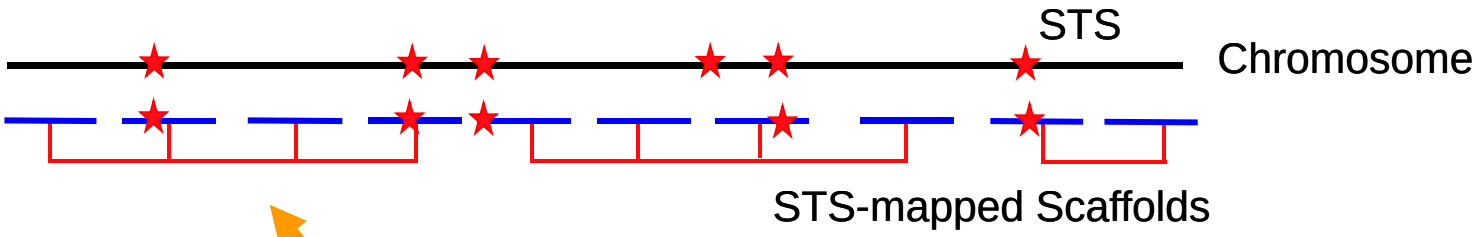
Whole Genome Shotgun Sequencing

- Start with a whole genome
- Shear the DNA into many different, random segments.
- Sequence each of the random segments.
- Then, put the pieces back together again in their original order using a computer

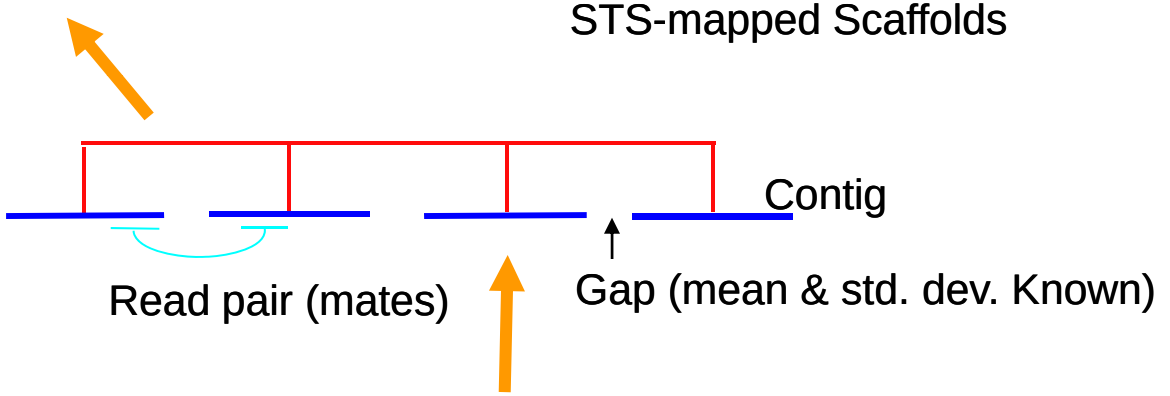

ONE
HUMAN
BEING!
(SOME
ASSEMBLY
RECQUIRED)
BY AUTH FOR THE PHILADELPHIA INQUIRER

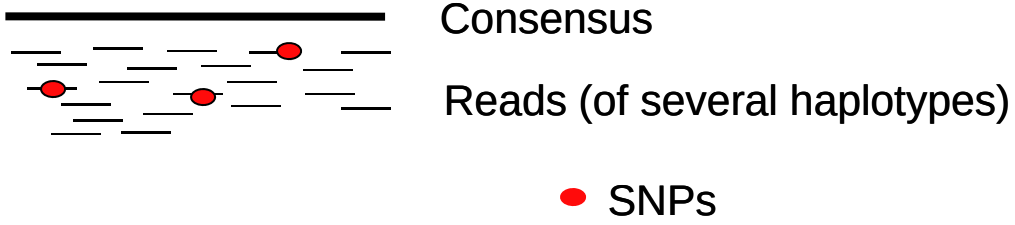
Anatomy of a WGS Assembly

Genetic and physical map



Pac Bio
Sanger
454
Illumina
Ion Torrent
Helicos





So what?

Anchored Genome Assembly

- Gene function
- Gene order
- Gene model
- Allele
- Functional mutation

Genome Browser