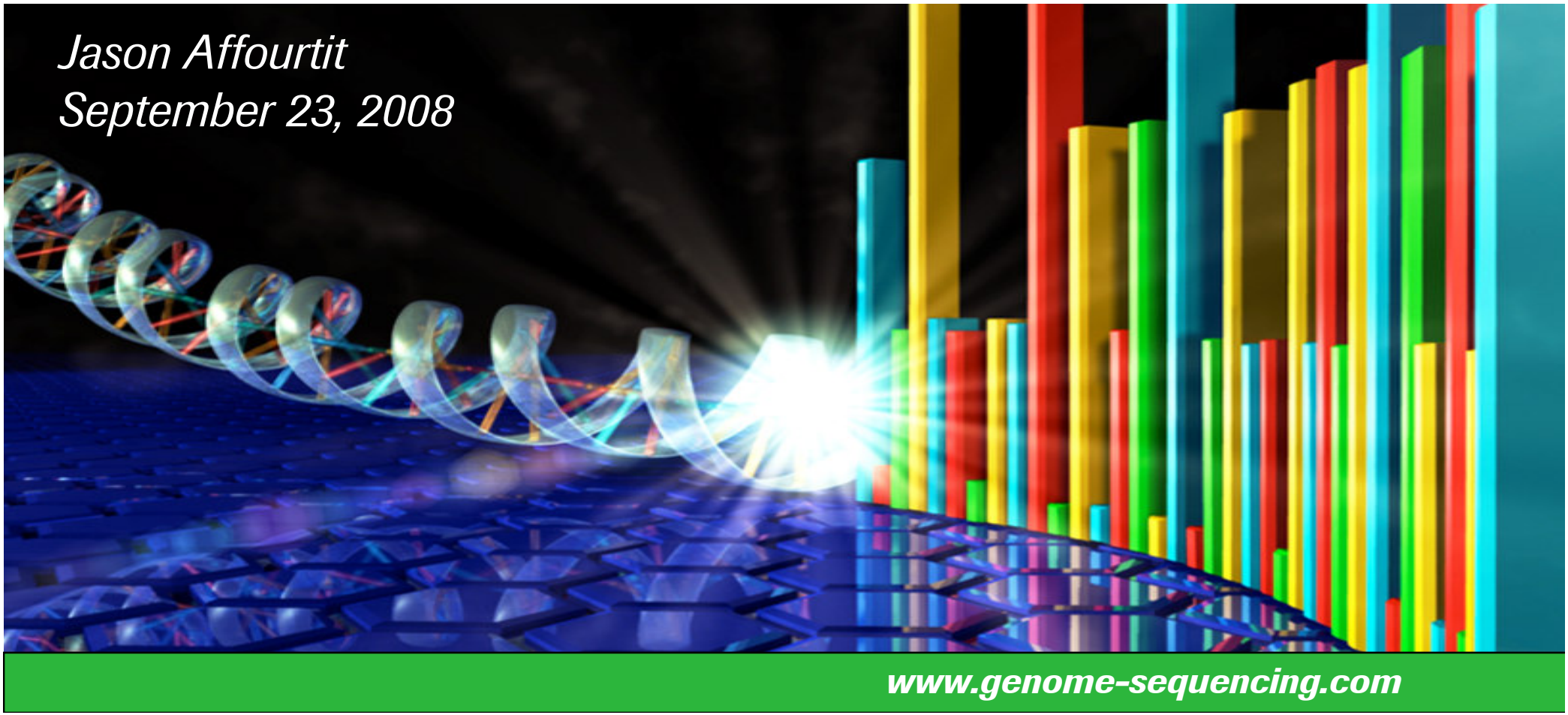
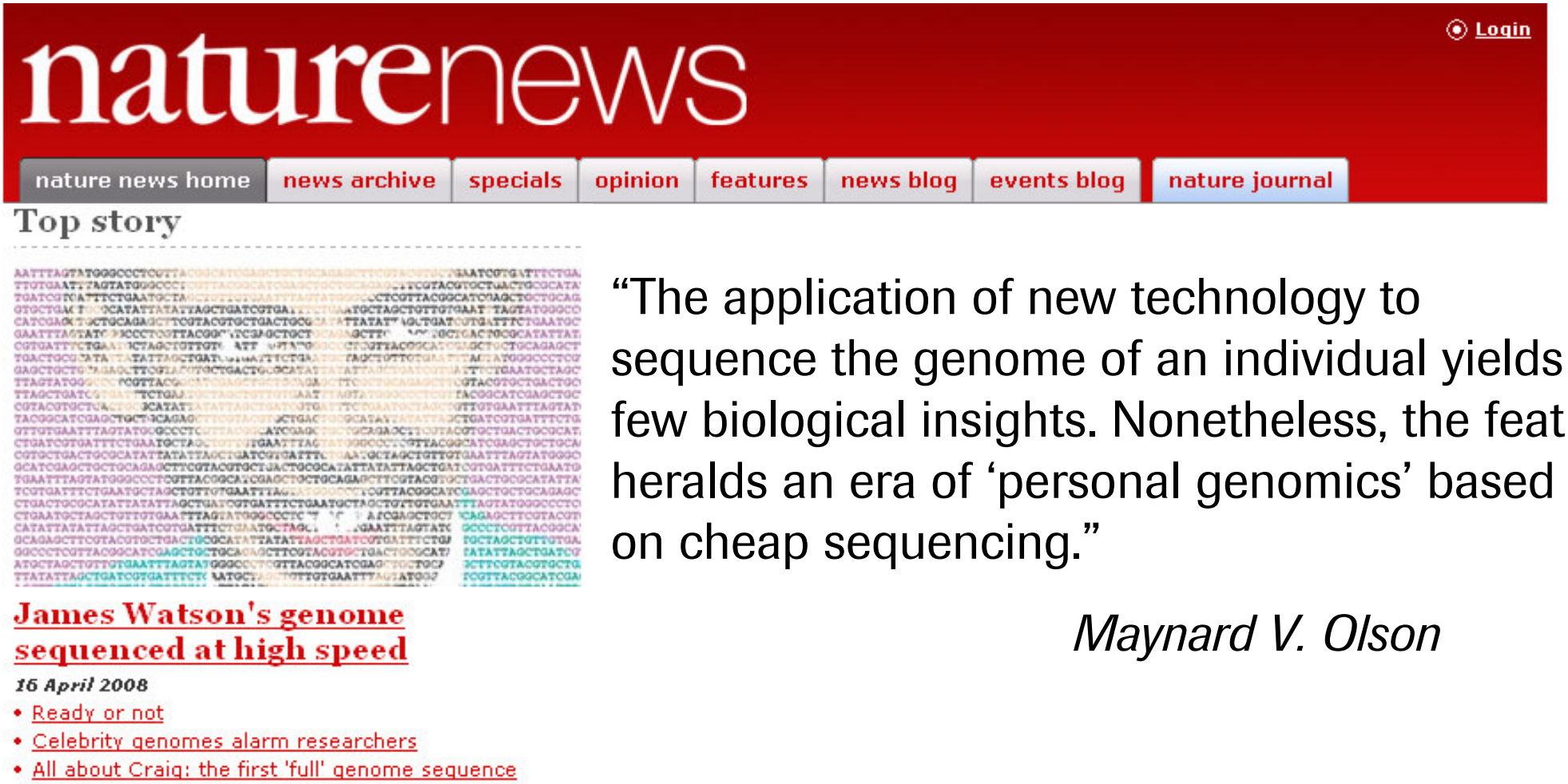
Titanium Series for the Genome Sequencer FLX
Read Length *Really* Matters



Which Factors are Critical When Considering Next Generation Sequencing?

- Ease of use / reliability
- Accuracy
- Read length
- Bioinformatics / analysis capabilities

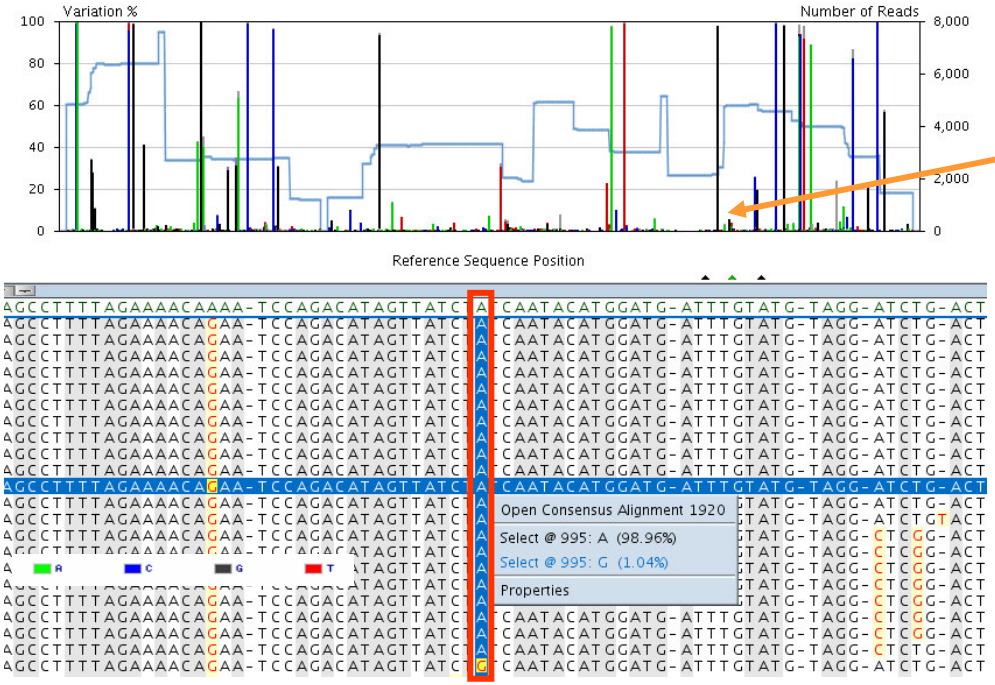
The First of The Rest of Us



"The application of new technology to sequence the genome of an individual yields few biological insights. Nonetheless, the feat heralds an era of 'personal genomics' based on cheap sequencing."

Maynard V. Olson

Detection of Rare Drug Resistance Mutations in HIV by Ultra Deep Sequencing



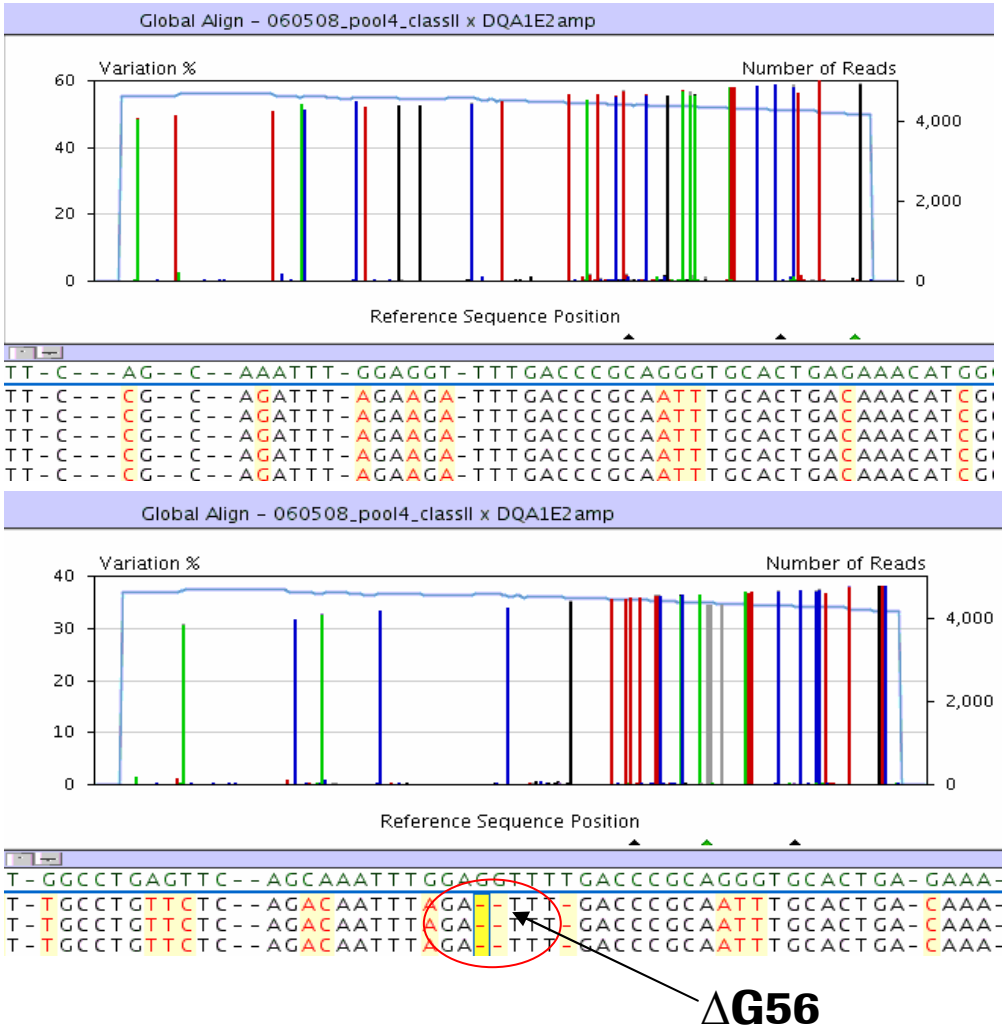
Y181C:

a major RT inhibitor

mutation at 1% prevalence

- Minor variants <20% are rarely detected with population sequencing
- In a retrospective clinical study, detected resistance mutations more than doubled with UDS
- Significant association of baseline resistance mutations with virological failure

Tissue Typing via FLX Titanium Sequencing

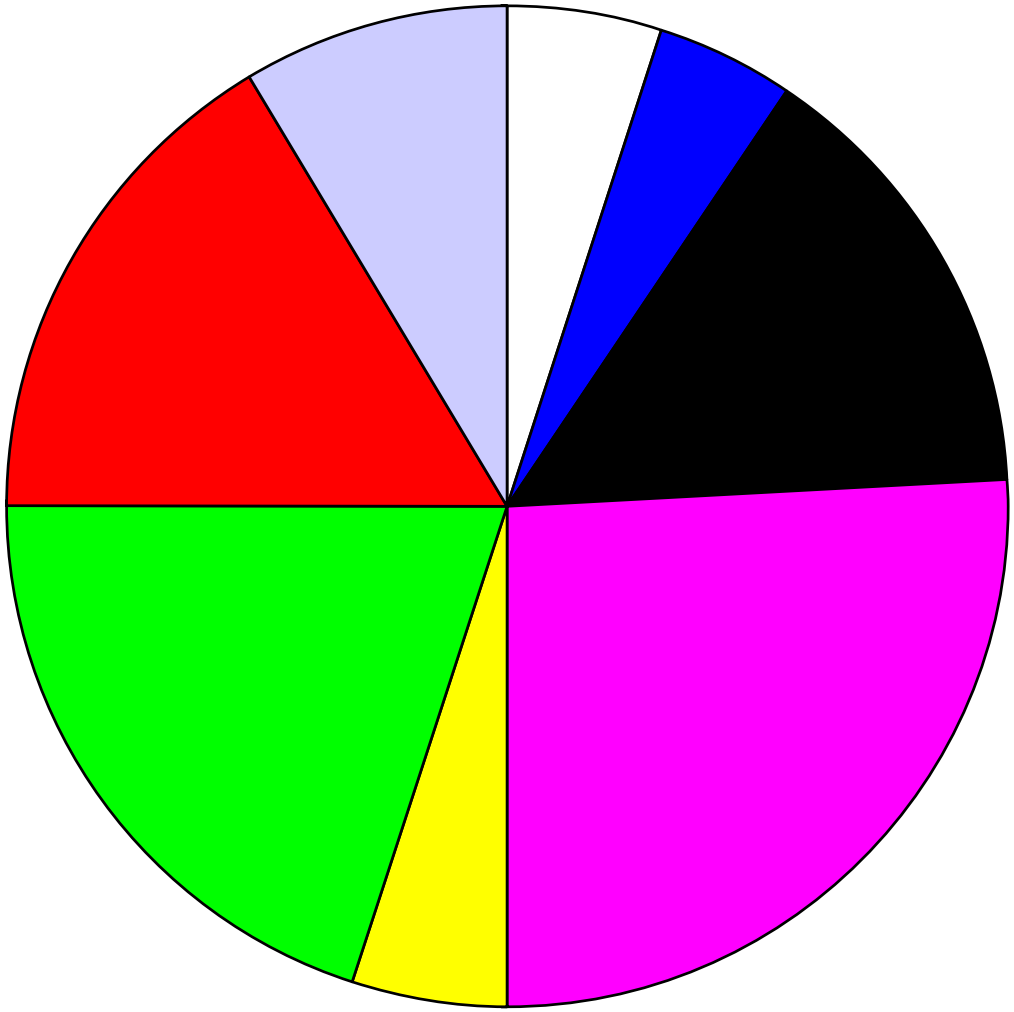


allele 1: HLA DQA1*03

allele 2: HLA DQA1*05

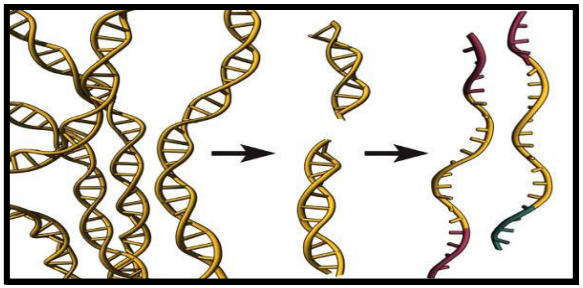
- Full high resolution HLA haplotypes in a single Titanium run!
- No laborious cloning, allele-specific PCR, or family studies
- No manual review of sequencing traces

250+ Publications: Scientific Validation

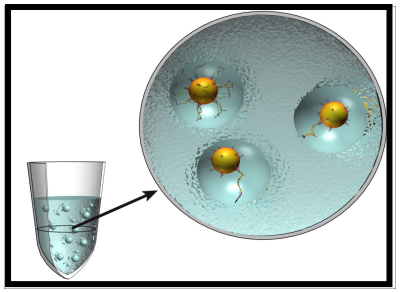


- Evolution & Ecology
- Human genetics & genomics
- Metagenomics & Microbial Diversity
- Microbes, Viruses & Infectious Diseases
- Cancer Research
- Plants & Agricultural Biotechnology
- Model & Non-model Organisms, Systems Biology
- Technology & Informatics

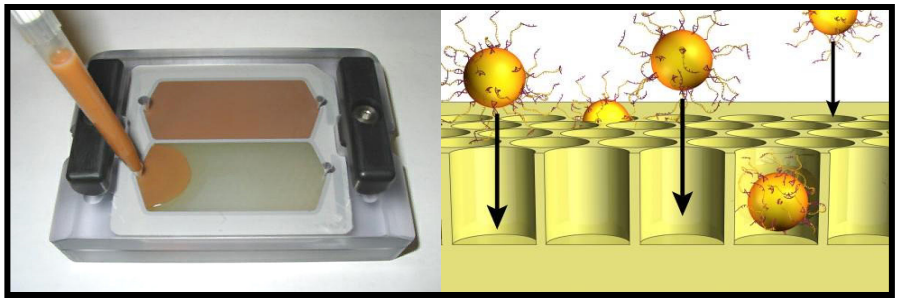
454 Sequencing Overview



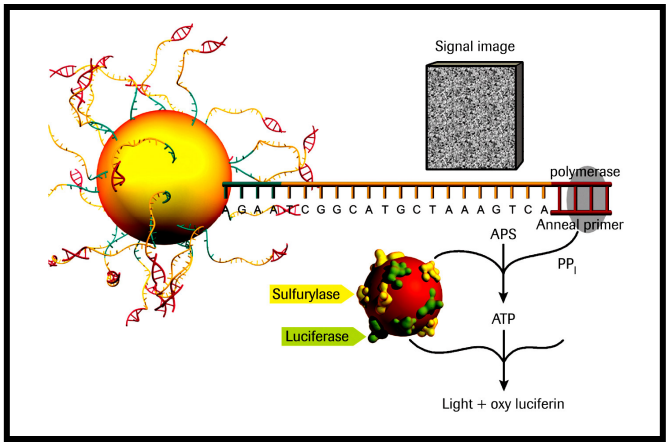
Shear DNA and add linkers



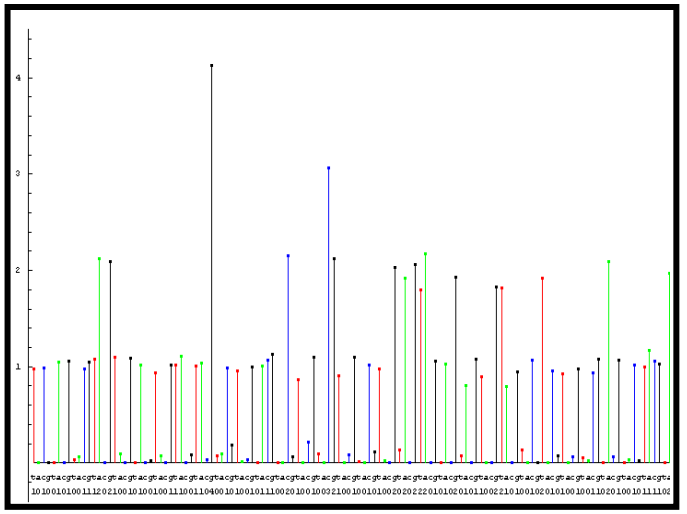
'Emulsion PCR'
Clonal amplification



Deposition of beads into wells of PTP device



Sequencing-by-synthesis
Detection of PPi release



454 Sequencing Overview

One fragment

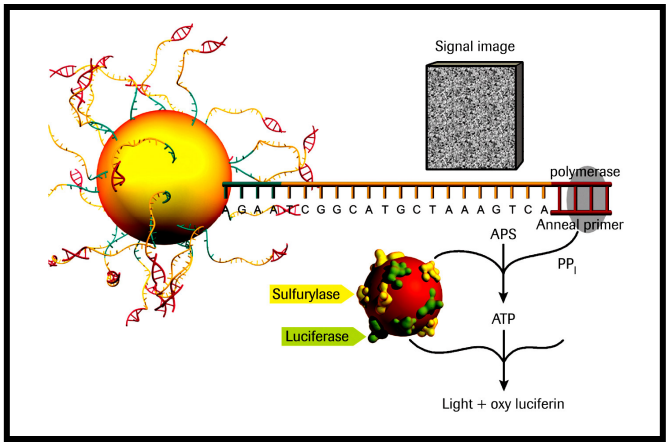
Shear DNA and add linkers

One bead

'Emulsion PCR'
Clonal amplification

One well

Deposition of beads into wells of PTP device

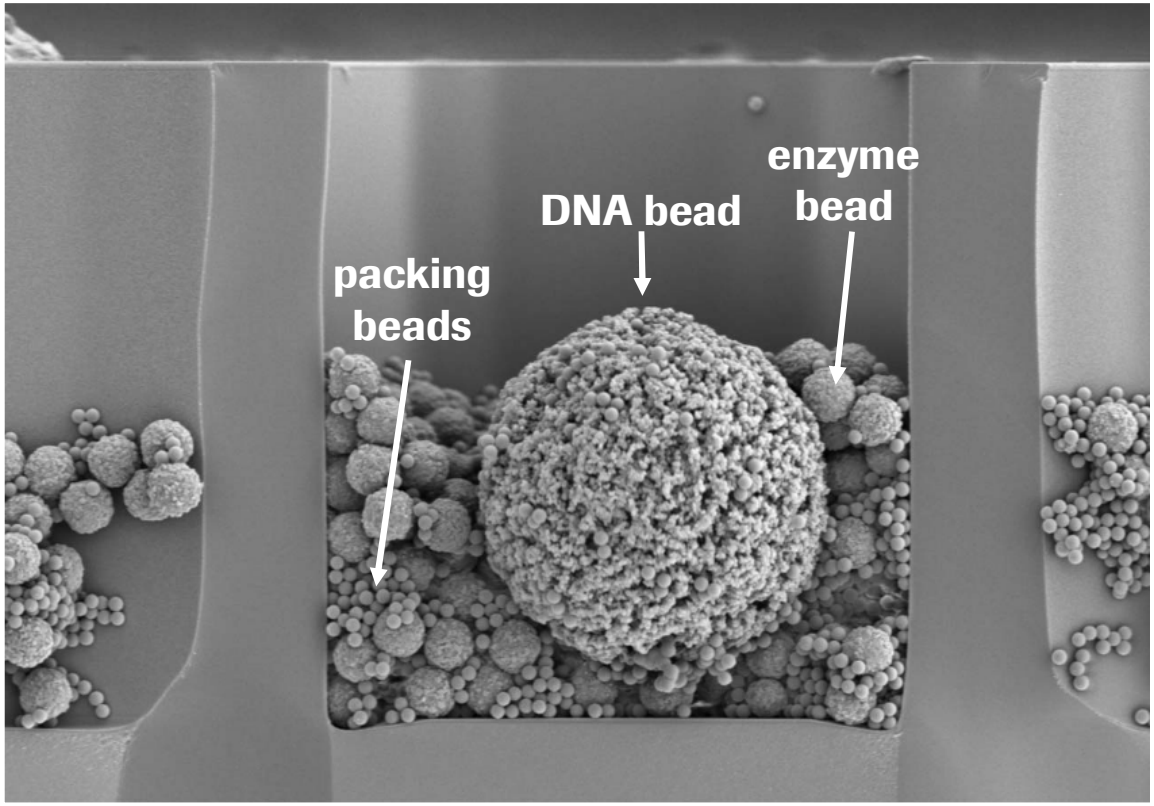


Sequencing-by-synthesis
Detection of PPi release

One read

Sequencing Workflow

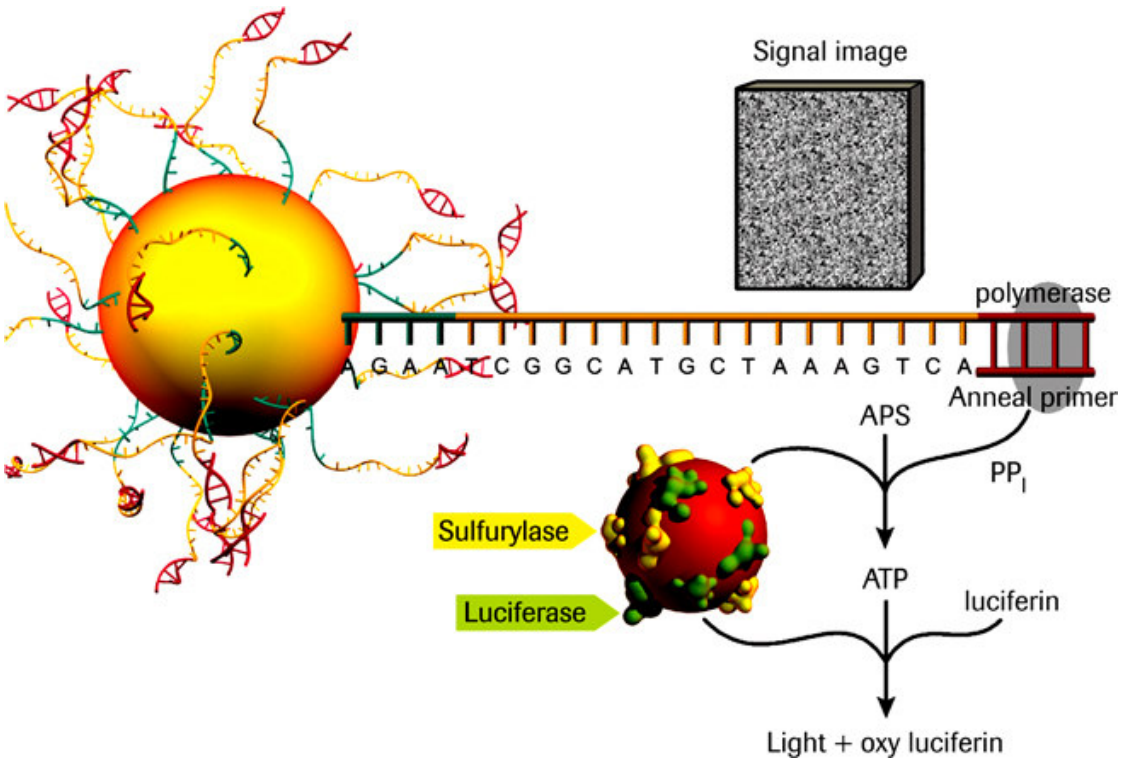
Sequencing by Synthesis



- Bases (TACG) are flowed sequentially and always in the same order across the PicoTiterPlate device during a sequencing run.
- A nucleotide complementary to the template strand generates a light signal.
- The light signal is recorded by the CCD camera.
- The signal strength is proportional to the number of nucleotides incorporated.

Sequencing Workflow

Sequencing by Synthesis



- Bases (TACG) are flowed sequentially and always in the same order across the PicoTiterPlate device during a sequencing run.
- A nucleotide complementary to the template strand generates a light signal.
- The light signal is recorded by the CCD camera.
- The signal strength is proportional to the number of nucleotides incorporated.

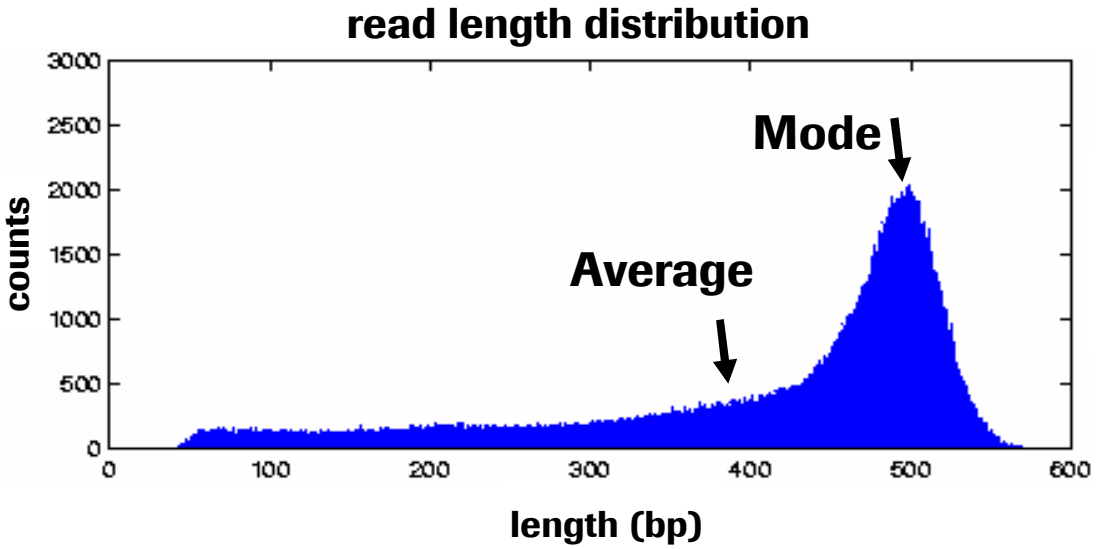
We need longer reads while making sequencing faster, easier, and cheaper!

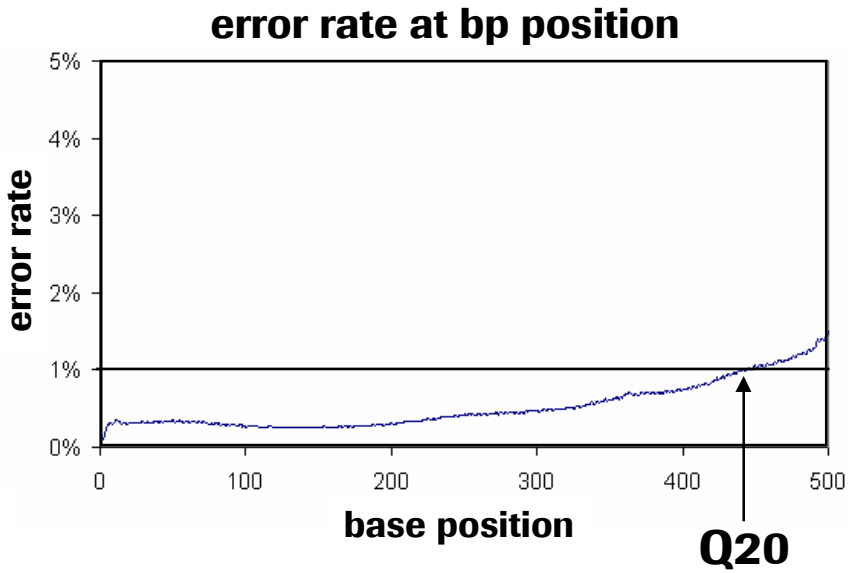
GS FLX *Titanium* Series Kits

- New kits for library prep., emPCR, and sequencing
  - Workflow simplified
  - Polymerase mix included in emPCR kits
- 200 cycles instead of 100 cycles
  - Read length > 400 bp
  - Reduced time per cycle
- New higher density PTP with 3.6M wells (> 2X current 1.6M wells)
  - Resolved cross talk
  - Initially > 1M reads per run, potential for > 2M

Same GS FLX Sequencer!

GS FLX Titanium Series Performance

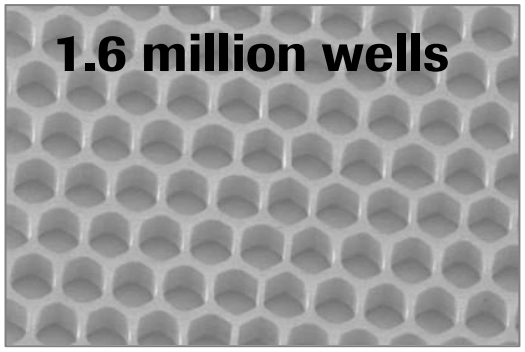
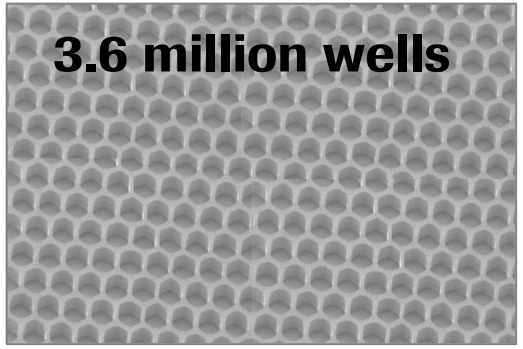
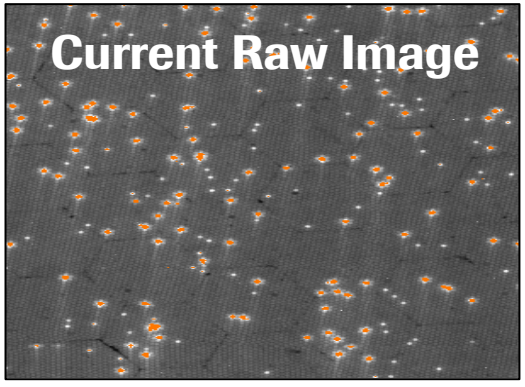
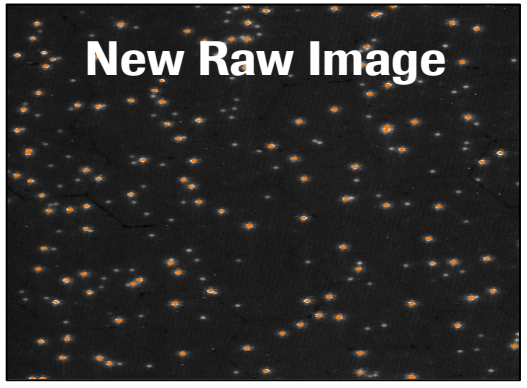




- More than one million Sanger-like reads per run
- Read length mode ~500 bp with an average of 350 – 400 bp
- 400 – 600 Mb average yield depending on organism and experience
- Q20 read length of ~400 bp

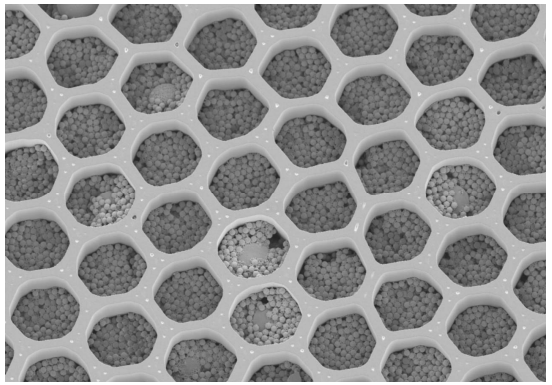
GS FLX Titanium Series PTP

Innovations Improve Yield & Signal Quality

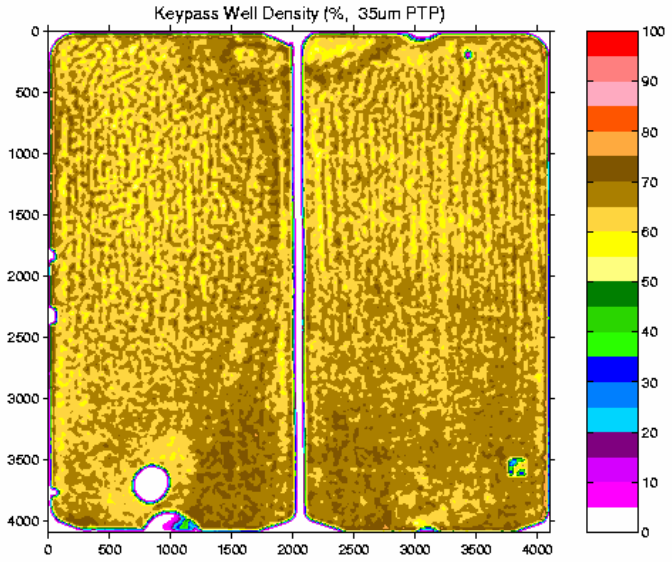
GS FLX Standard		GS FLX Titanium series	
	→		• Higher Density • More Reads • More Experiments
	→		• Longer Reads • Faster Cycle Time • Improved Accuracy

GS FLX Titanium Series in a Nutshell

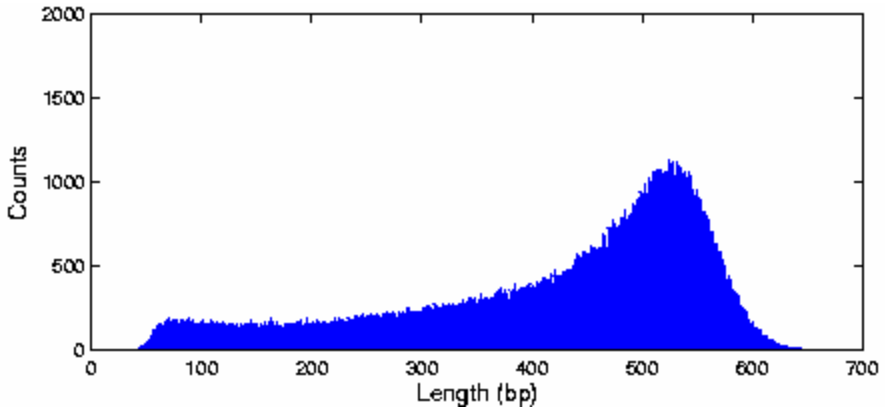
- Technology
  - More, longer, high quality reads
  - Decreased time per cycle & optimized chemistry
- Workflow
  - Streamlined processes
  - Protocol flexibility to meet research needs
- Data analysis
  - Enhanced mapper/assembler software
  - Additional functionality in Amplicon Variant Analysis package

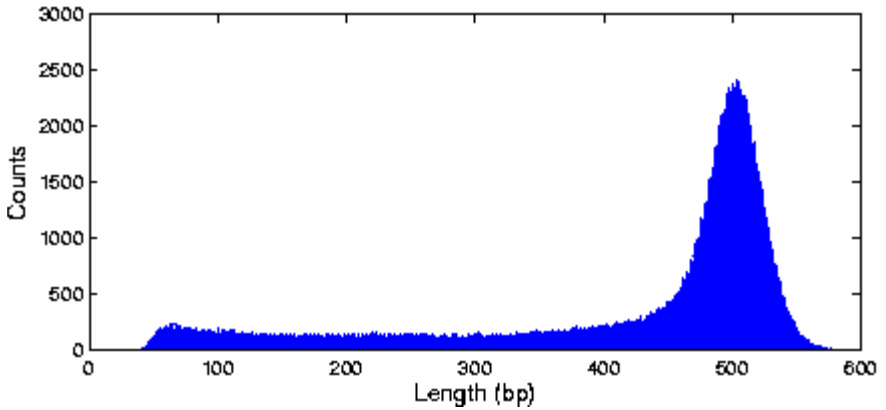


Titanium Results: Microbial Sequencing



- 2 microbial genomes (1 per region)
- 659 M total high quality bases
- 68% of reads are high quality
- 409/413 bp average read length
- Results from external site



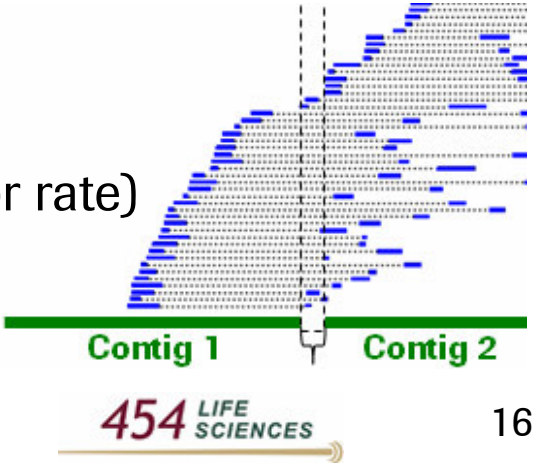


Powerful Assembly Software

- Present GS FLX software
  - Mappable – Any genome, local variations
  - Draft Assembly – Bacterial genomes, yeast-sized eukaryotes, BACs
  - Auto. “Finished” – Adenovirus, an occasional individual BAC

- Titanium release adds further capabilities
  - Mappable – paired-end results, protein translations, known SNPs
  - Draft Assembly – fungal genomes, small to medium eukaryotes

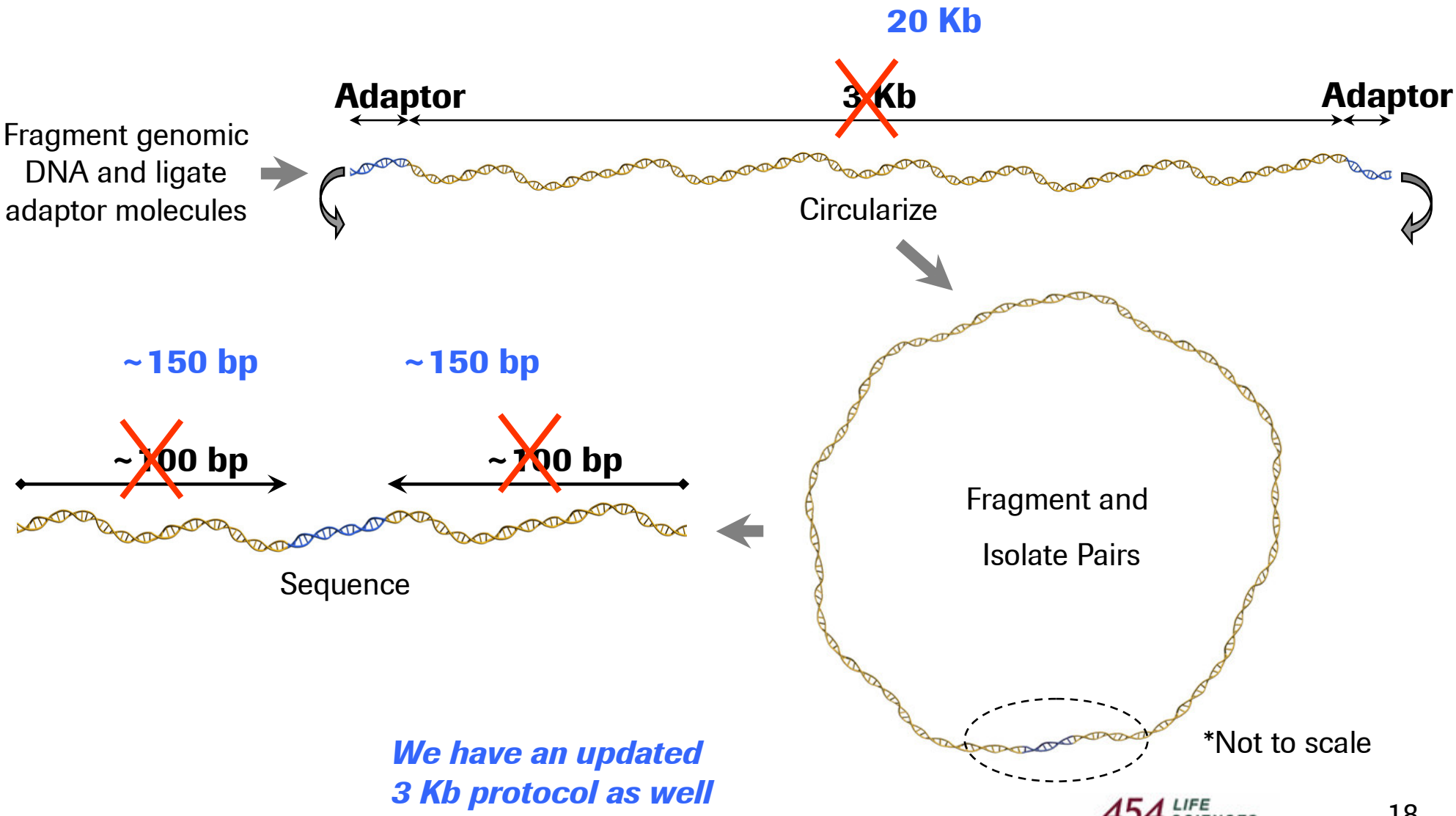
- In R&D today
  - Manually “Finished” – *C. jejuni* (1 contig, 1/50,000 error rate)



One-Button "Draft" *de novo* Assembler Results from GS FLX Titanium Runs

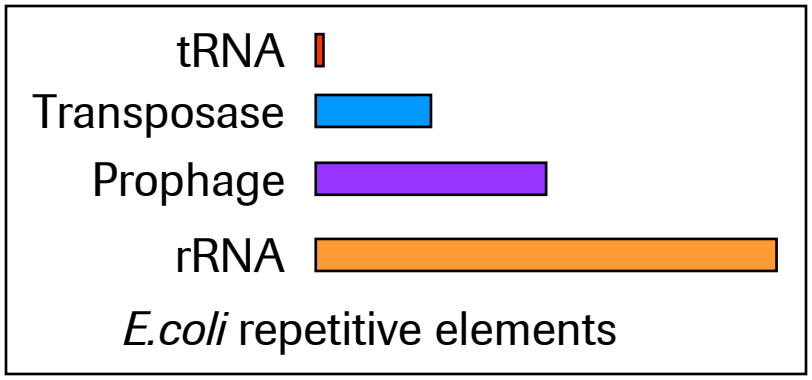
	C. jejuni	*T. thermophilus*	*E. coli*	*E. coli (FLX Std)*
Genome Size:	1,641,481	2,127,575	4,639,675	4,639,675
Number of Runs:	1/8	1/8	1/4	1
Assembly Contigs:	32	145	101	105
Assembly Cover:	98.57%	97.04%	98.11%	97.61%
Overall Accuracy:	99.996%	99.992%	99.995%	99.998%
Avg. Contig Size:	50.6 kb	14.4 kb	45.1 kb	43.3 kb
N50 Contig Size:	129.1 kb	25.0 kb	94.9 kb	105.5 kb
Largest Contig:	339.3 kb	64.2 kb	284.3 kb	204.7 kb

Long-Tag Paired End Reads: Evolution

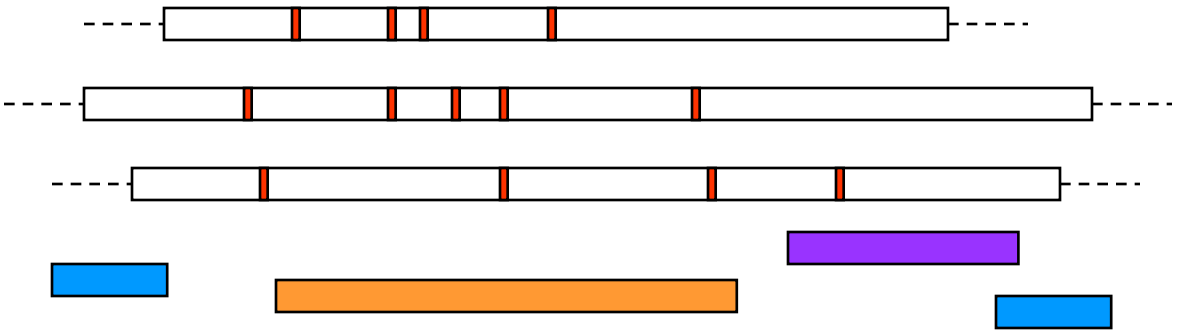


We have an updated 3 Kb protocol as well

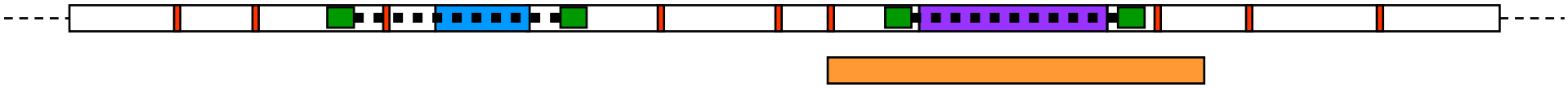
de novo Assembly of *E. coli* Genome into <u>One</u> Scaffold Using Only 454 Sequencing



Shotgun Only: 98 Scaffolds



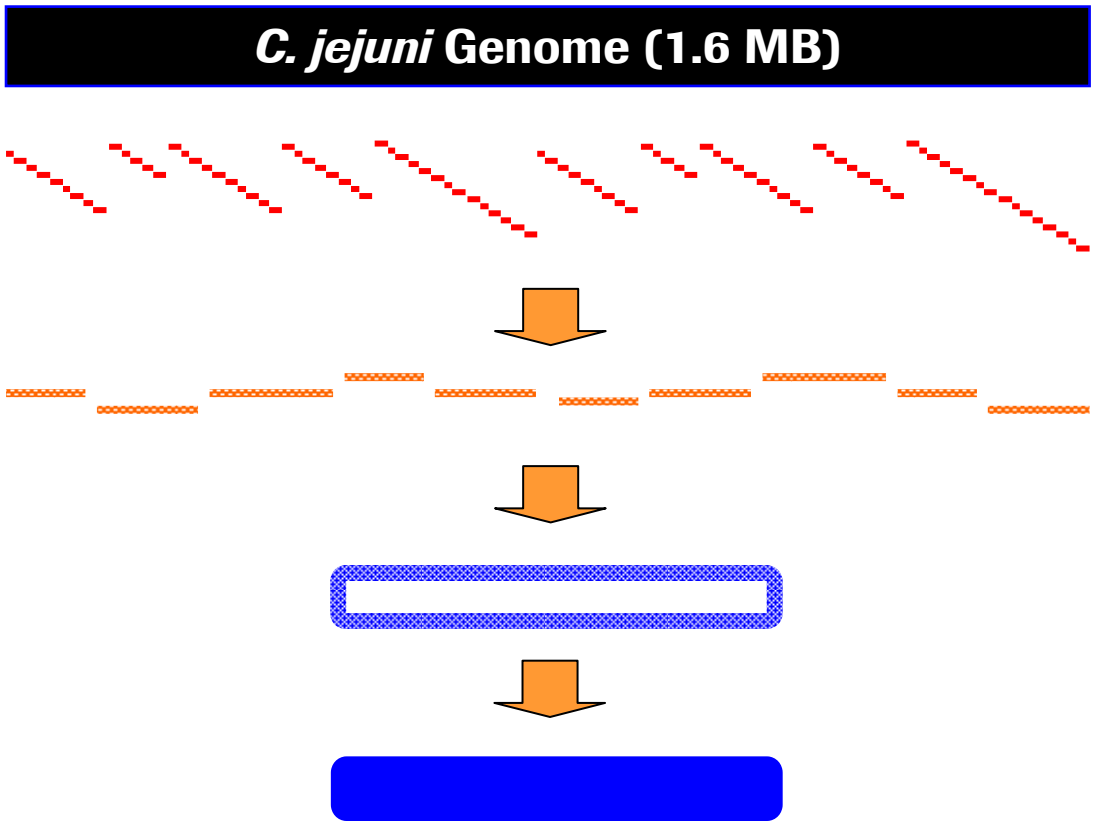
Shotgun + 3 kb Jump: 7 Scaffolds



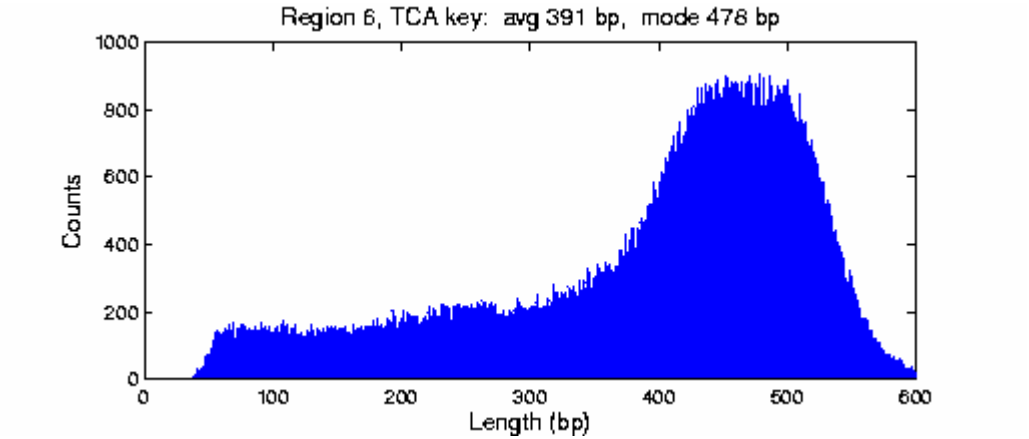
Shotgun + 3 kb Jump + 20 kb Jump: **1 Scaffold**

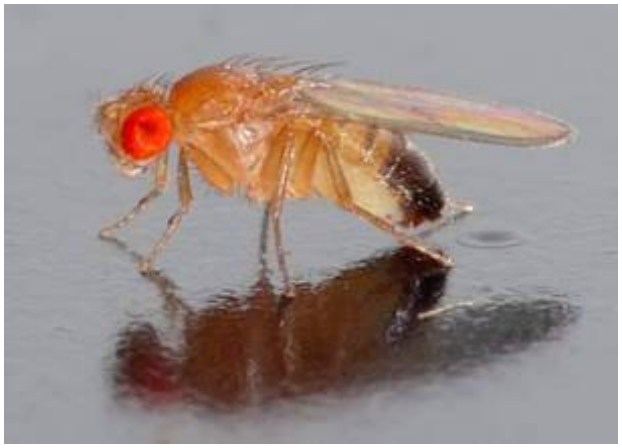
Assembling *C. jejuni* into One Contig

454 Read Type	Genome Coverage	Number of Contigs/ Scaffolds
FLX Shotgun	18×	27
+		
3 Kb Jump	36×	4
+		
20 Kb Jump	38×	1
+		
in silico Finishing		



Larger Genomes





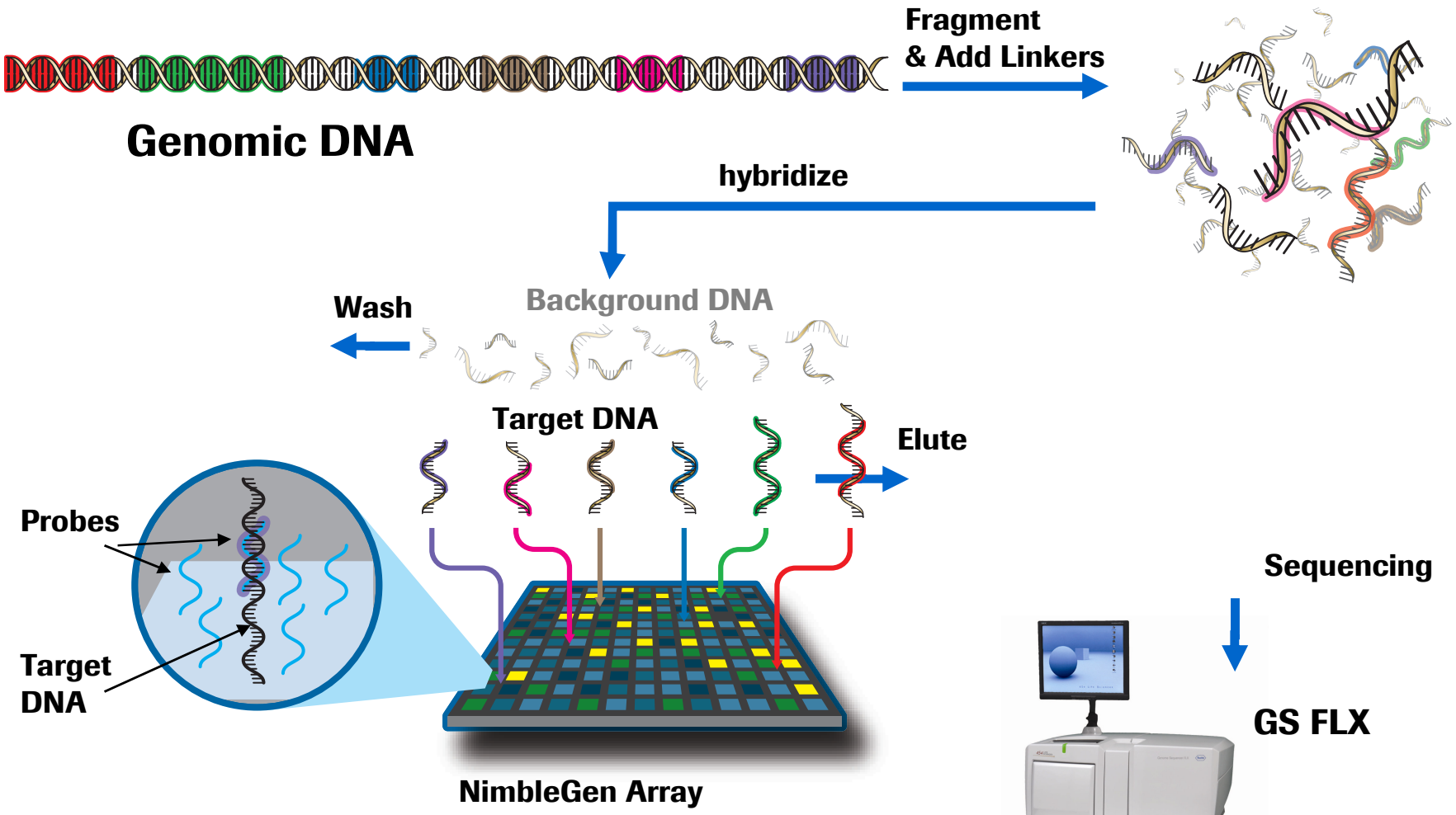
Larger Genomes – *de novo* Assembly

- **No BAC libraries**, just Whole Genome Shotgun (WGS) + Paired-End
- Titanium shotgun reads combined with PE spans up to 20 kb
- Pilots using Titanium on ***Arabidopsis*** (157 Mb) and ***Drosophila*** (180 Mb)
- Improvements to GS Assembler
    - Have completed assemblies of 3x-11x datasets of 200-900 MB genomes

Arabidopsis **Data Set**	**Contigs/Scaffolds**	**Contig/Scaffold Size**
15x shotgun reads (Ti)	~10,300 contigs, 107MB	**21 kb** N50 contig size
plus 3 kb PE (**11x**)	~1,500 scaffolds, 108MB	266 kb N50 scaffold size
plus 10 kb (**11x**) & 15 kb (**14x**)	**238** scaffolds ≥ 7 kb, 109 MB	**4.1 Mb N50 scaffold size**

Drosophila **Data Set**	**Contigs/Scaffolds**	**Contig/Scaffold Size**
12x shotgun reads (Ti)	~13,500 contigs, 116 MB	**27 kb** N50 contig size
plus 3 kb PE (**9x**)	~2,300 scaffolds, 117 MB	346 kb N50 scaffold size
plus 20 kb PE (**68x**)	255 scaffolds ≥ 10 kb, 118MB	**3.3 Mb N50 scaffold size**

NimbleGen Microarray Sequence Capture



454 LIFE SCIENCES

Length *Really* Matters for Sequence Capture Performance

	FLX	Titanium
Total Reads Aligned:	88.72%	90.75%
Aligned Reads On-Target:	61.60%	69.40%
Targets Hit:	98.30%	98.70%
Bases with 10+ Coverage:	74.80%	94.40%

Deeper coverage = more powerful heterozygote calling!

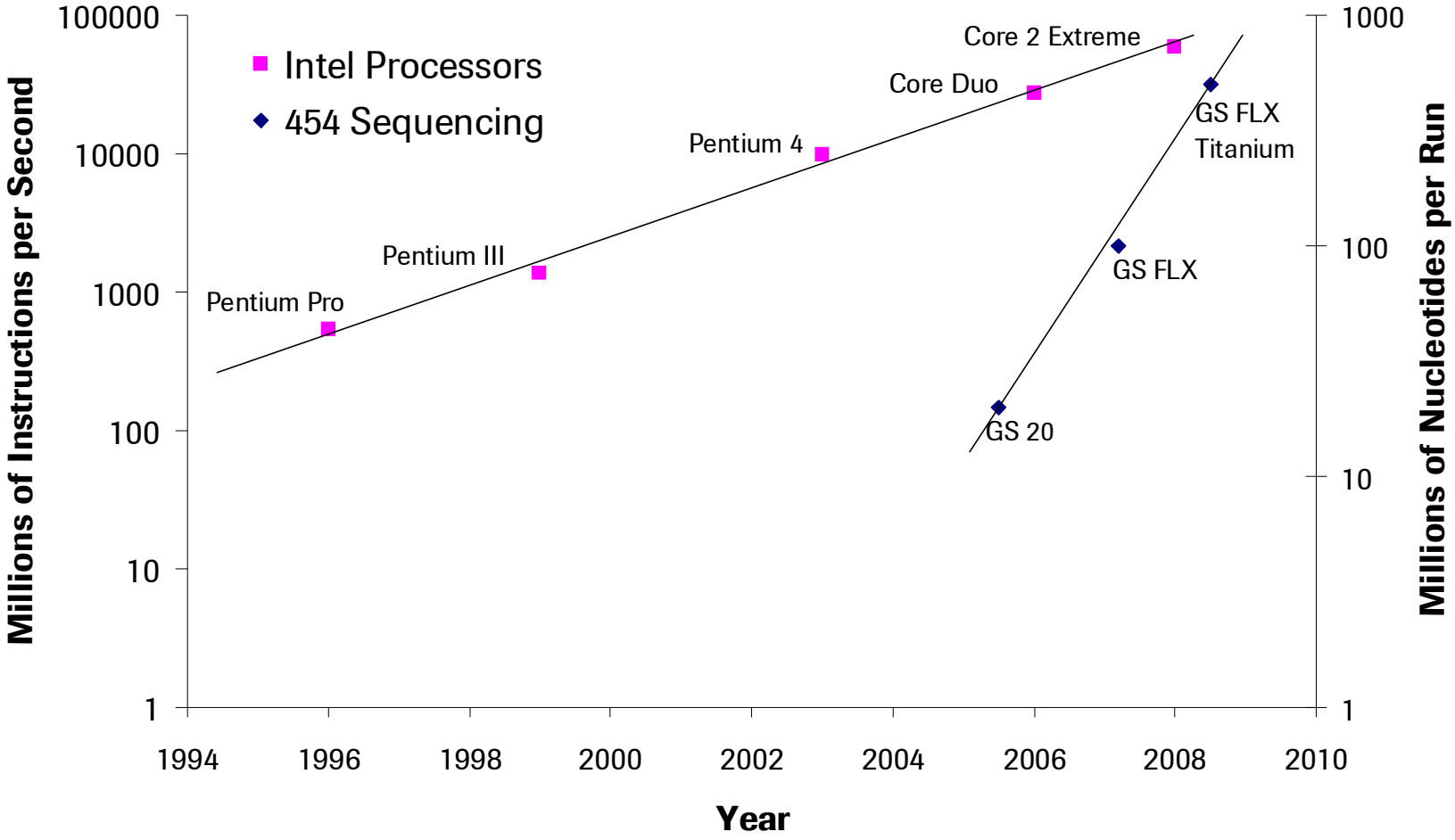
BCM Baylor College of Medicine

HGSC HUMAN GENOME SEQUENCING CENTER

454 LIFE SCIENCES

Commitment to Innovation

"Moore's Law"-Like Performance Improvement



Summary

- GS FLX Titanium: software and reagent upgrades only
- Streamlined workflow and protocols mean more efficient experiments and fast turnaround of projects
- More than 1M reads with a read length mode of 500 bp and Q20 read length of ~400 bp → 400-600 Mb yield per 10 hr run
- New paired end protocols and updated software allow for unparalleled *de novo* assembly and variant analysis options

Read length really *matters!*

Acknowledgements

- 454 Life Sciences R&D
- Roche NimbleGen
- Baylor College of Medicine / HGSC


Roche