Implications of recent genomics developments for breeding

Matthew Hudson

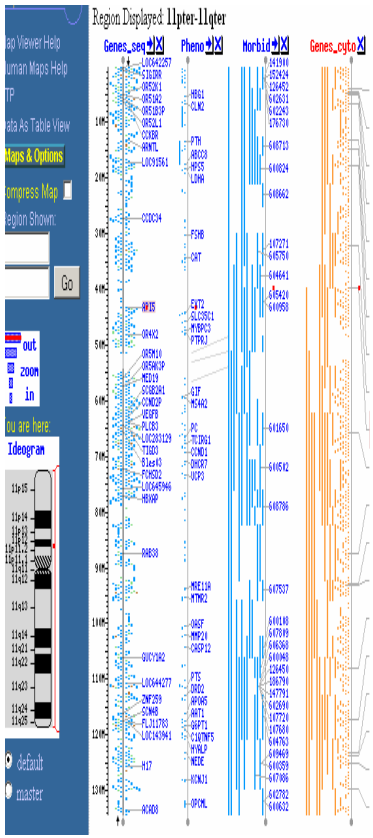
Dept of Crop Sciences

UIUC

Genome projects

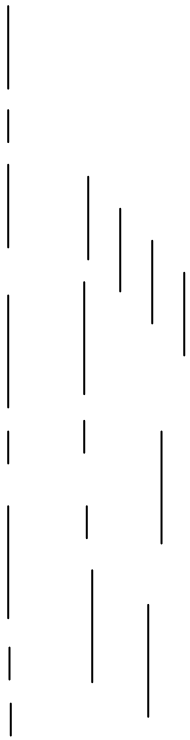
- Almost every crop now has a genome project
- Two plants, one crop are really “done” – Arabidopsis and rice
- Shotgun genomes available for poplar, Chlamydomonas
- Partial genomes for maize, sugarcane, Medicago, Lotus, etc
- What good is this to breeders?

Finished genome



Whole chromosome sequences
Done clone by clone
e.g. human, Arabidopsis

Shotgun genome



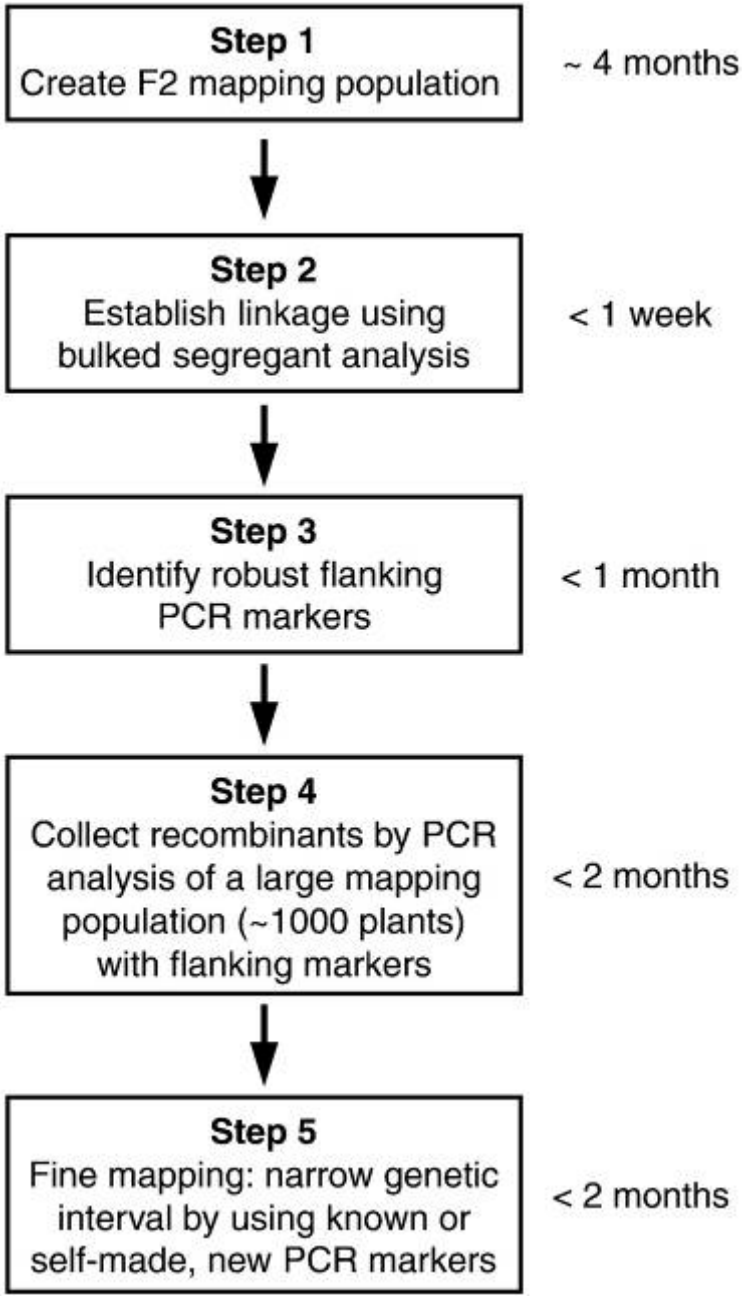
100kb average chunks
Need physical map
e.g. poplar

Maize now

Some BAC contigs
MAGIs

What can a *finished* genome do for breeders?

- Molecular markers – when you want them, where you want them, on a “perfect” physical map
- Cloning QTL – no need to make BAC contigs or chromosome walk (although still need to narrow down locus using fine-mapping and high density markers).
- Candidate gene approach – Good annotation can allow educated guesses about what genes might control key phenotypes (although these are often wrong!)
- Whole-genome resequencing – Know all the genetic differences between any two lines



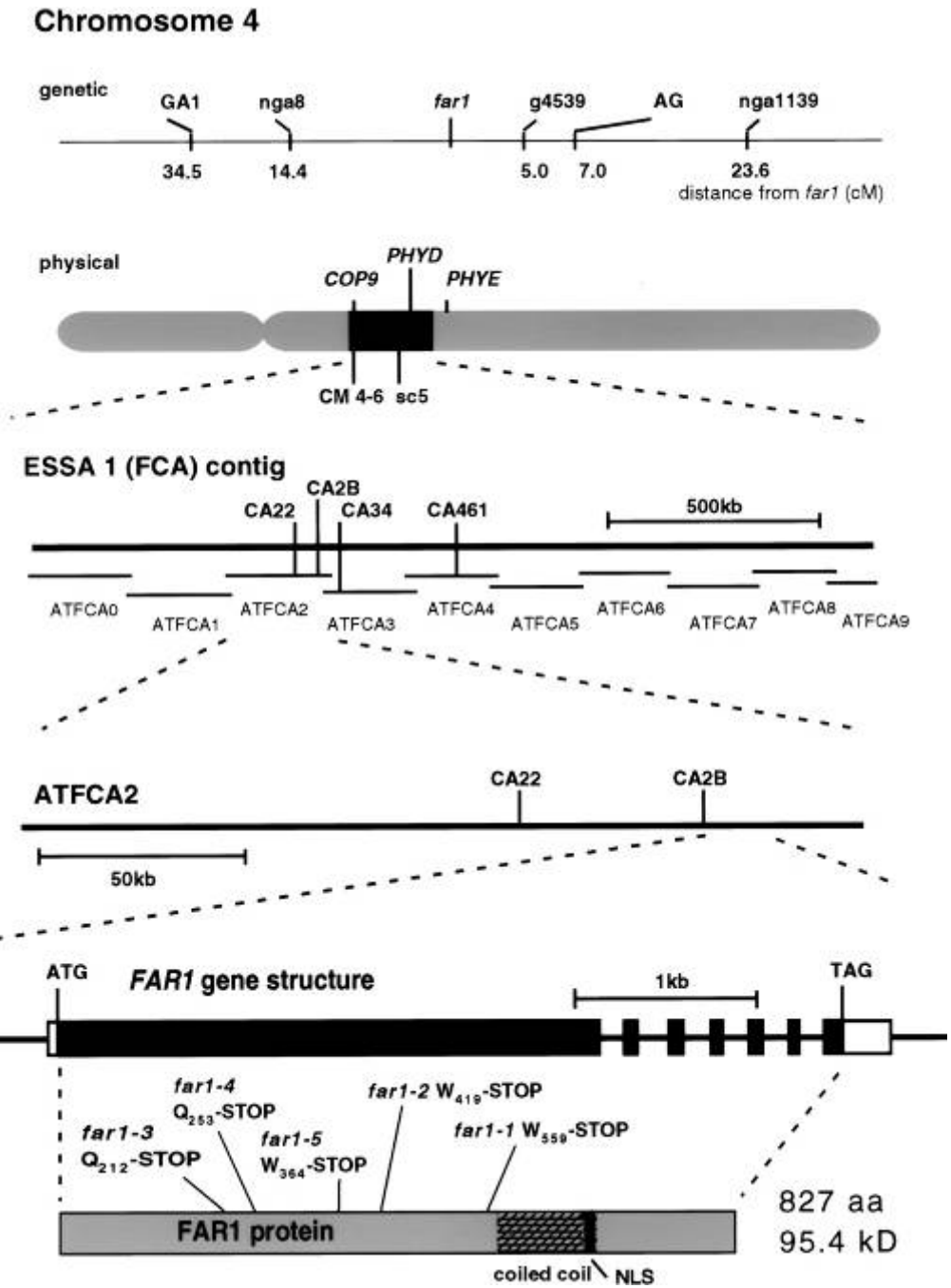
In Arabidopsis, Mendelian loci and QTL are now routinely cloned using the genome sequence

Genome sequence eliminates the contig building of positional cloning:

It's now purely a problem of genetics and computational biology

That doesn't mean that it's always straightforward

Lukowitz, W., et al. Plant Physiol. 2000;123:795-806



1. Coarse map with existing markers
   Generate large mapping population

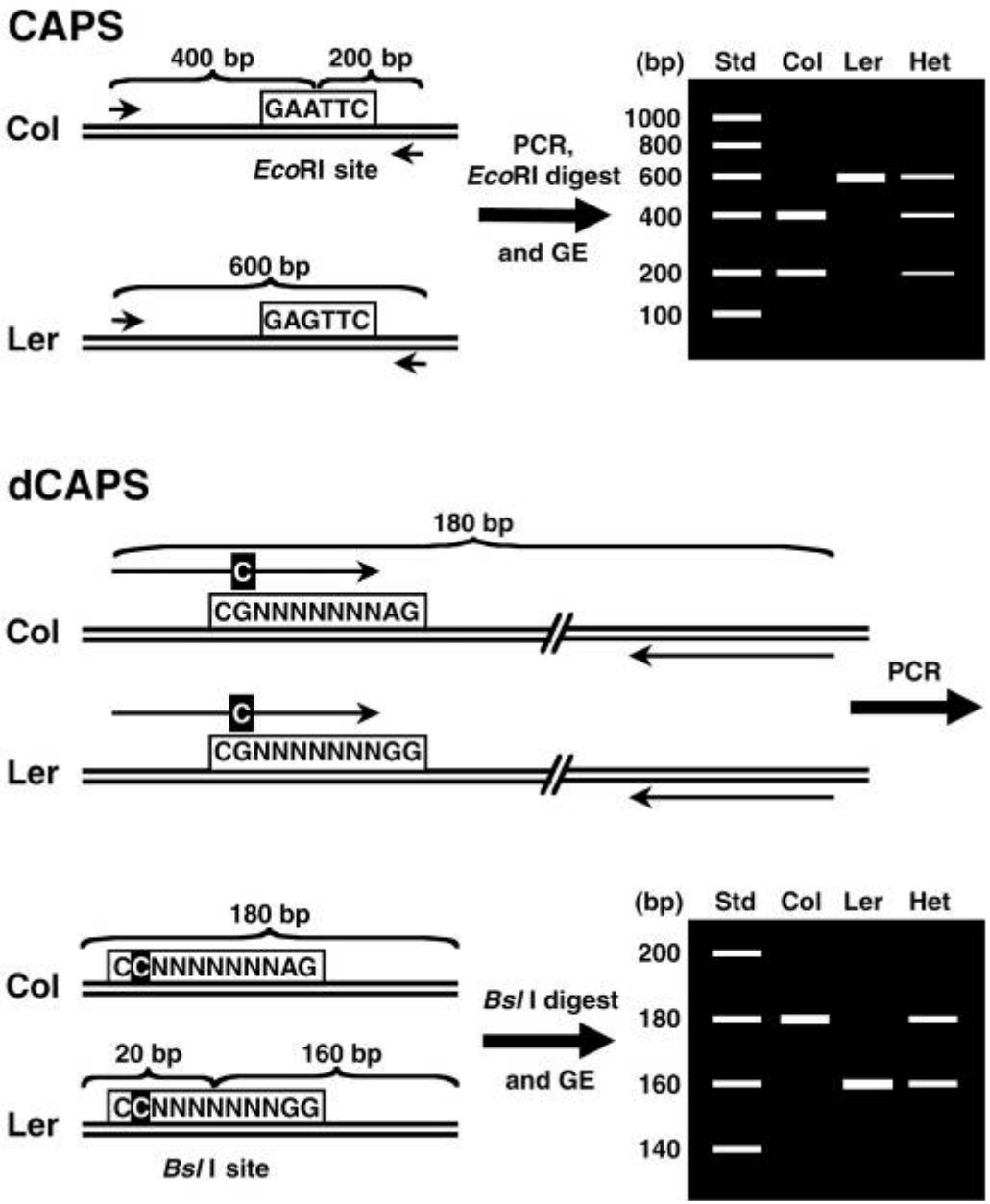
2. Localize within sequenced interval (in this case, ~2Mb)

3. Resequence loci every 100kb within interval in two genotypes. Use polymorphisms to develop new markers

4. Map to 100kb interval. Resequence and develop markers every 20kb

5. Locate multiple alleles to single protein coding locus

Positional cloning of the *FAR1* locus
Matthew Hudson et al.; Genes Dev. 1999; 13: 2017-2027



Markers are often required for only a few individuals

Arabidopsis geneticists prefer CAPS and dCAPS

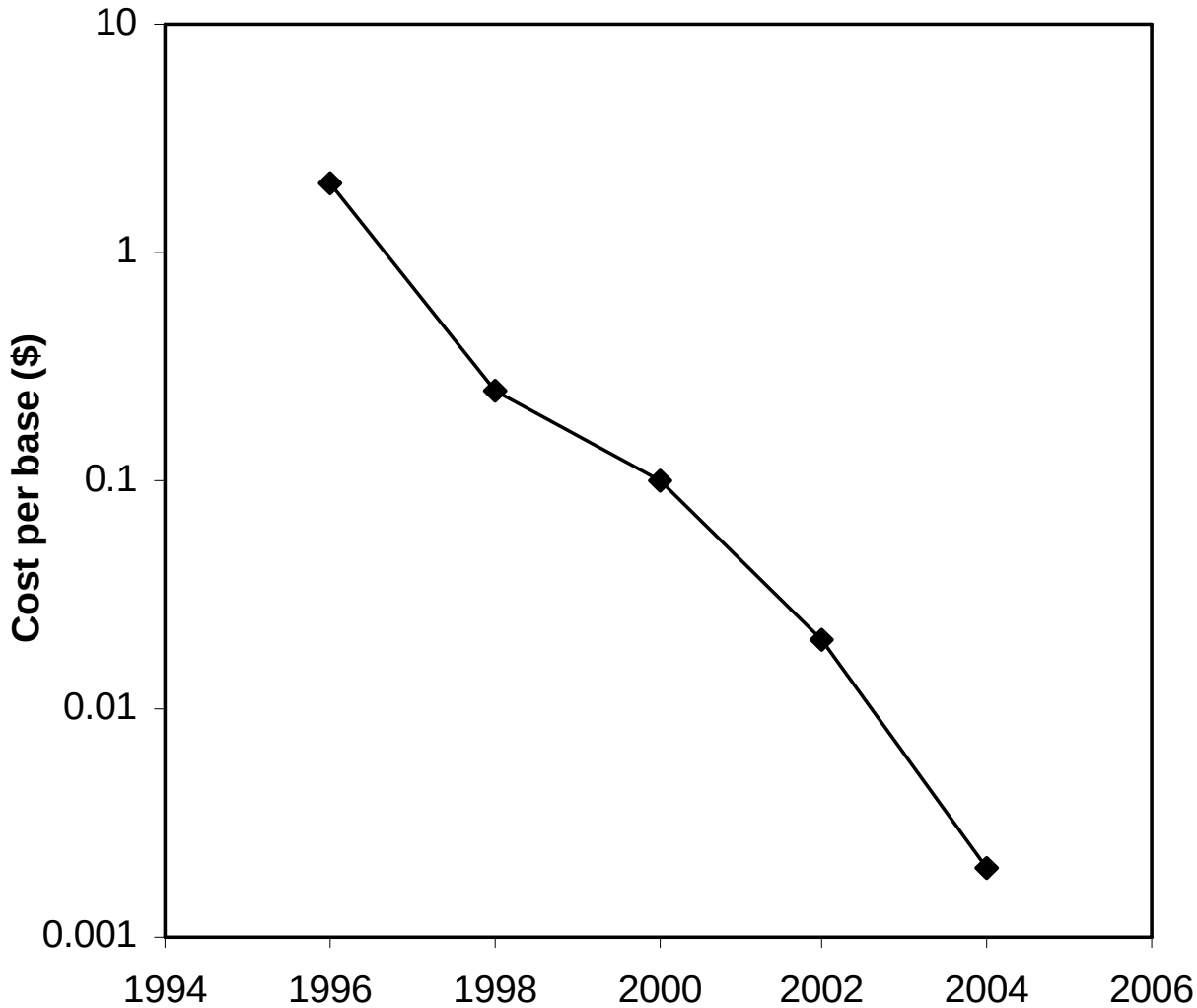
They are cheap, reliable and fast to create from any sequence polymorphism

Lukowitz, W., et al. Plant Physiol. 2000;123:795-806

Whole-genome resequencing

- To create markers, need to do a lot of PCR and sequencing
- Wouldn't it be great to have the whole genome of each line you work with? Then the whole genome would be haplotyped.
- Whole plant genomes still cost $40-50m
- NIH have target for human genome to cost $100,000 in 2010
- $1,000 in 2020
- This is likely to be achieved ahead of schedule

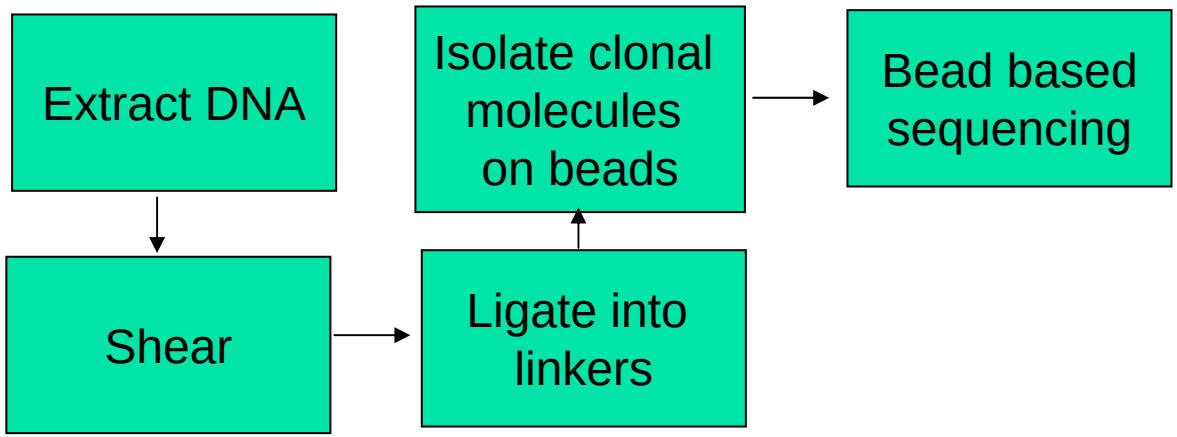
Cost of sequencing is falling exponentially



Next-generation sequencing

- A number of proprietary technologies, most based on the manipulation of microbeads and/or nanobeads where sequencing is performed without gels or capillaries
- First on the market was a company called “454”, the technology is now licensed to Roche
- Now have a major competitor in Solexa
- Recently ABI announced its own next-generation platform, SOLiD.

Next-generation approach

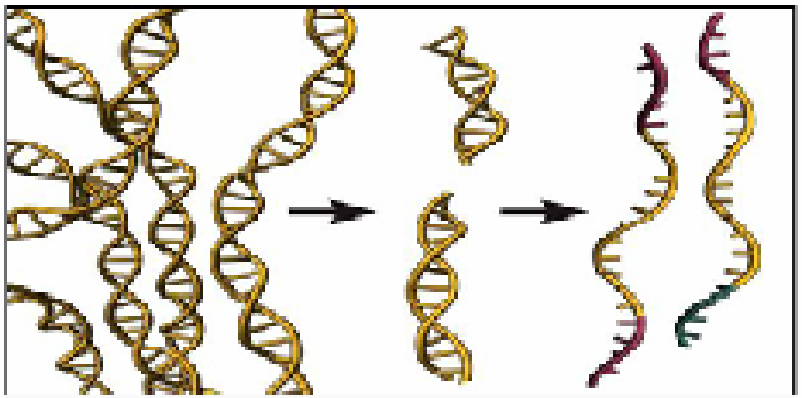


No colonies to pick

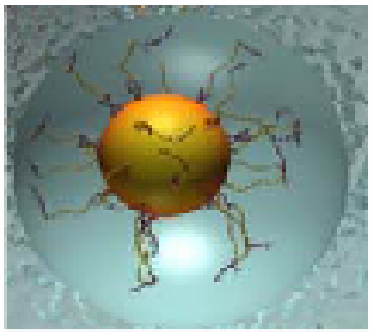
No minipreps or bacteria

Much higher throughput (millions vs. 96 or 384)

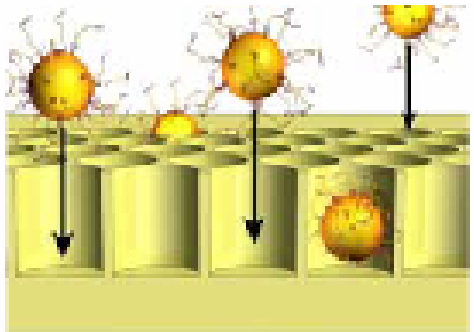
454 Sequencing technology



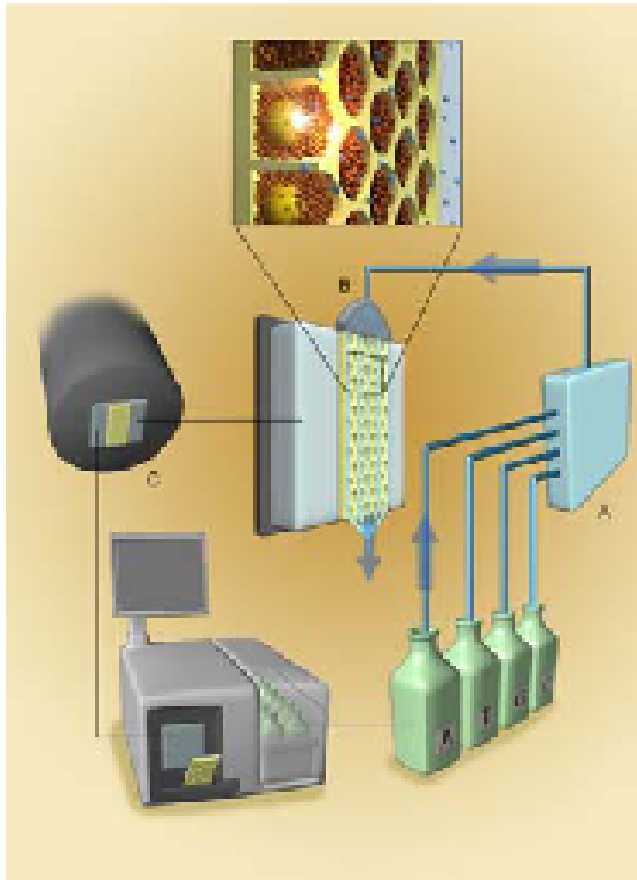
1) Prepare Adapter Ligated ssDNA Library



2) Clonal Amplification on 28 µ beads



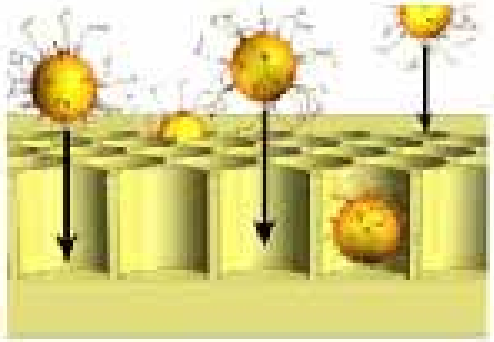
3) Load beads and enzymes in PicoTiterPlate™



4) Perform Sequencing by synthesis on the 454 Instrument

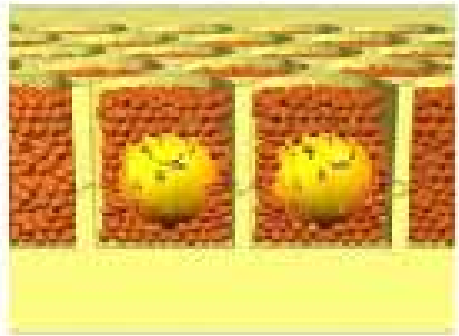
Picowell (50nm) technology

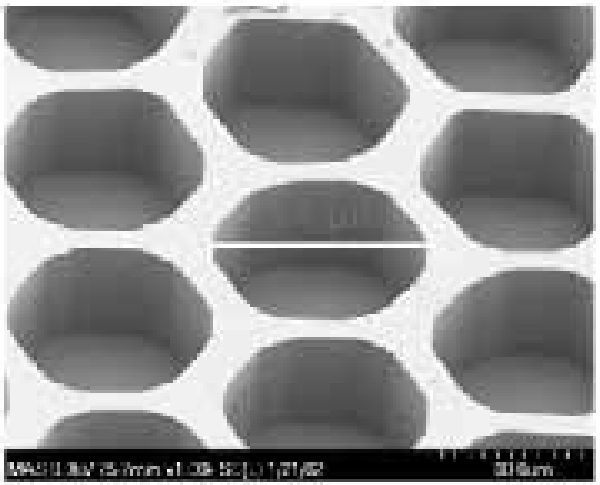
Load beads into PicoTiterPlate™

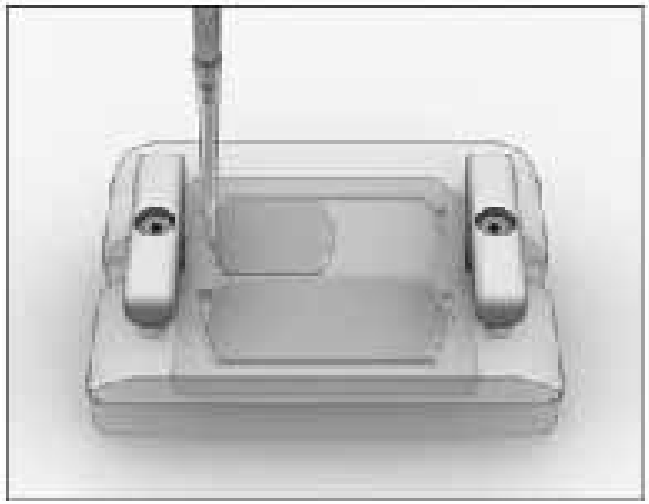


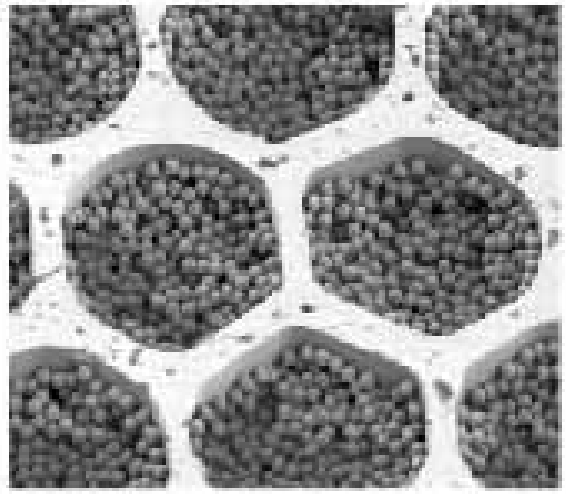
Load Enzyme Beads

Centrifugation



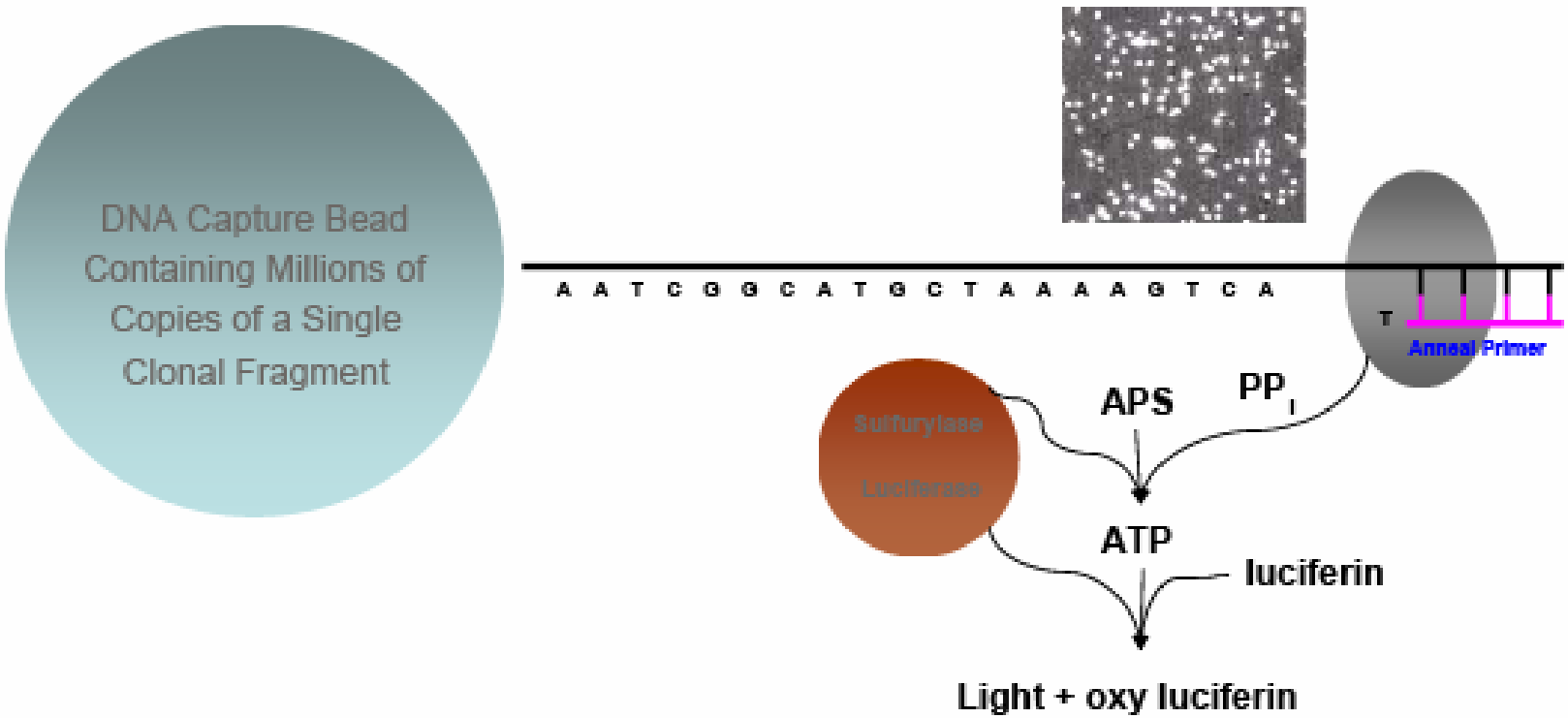






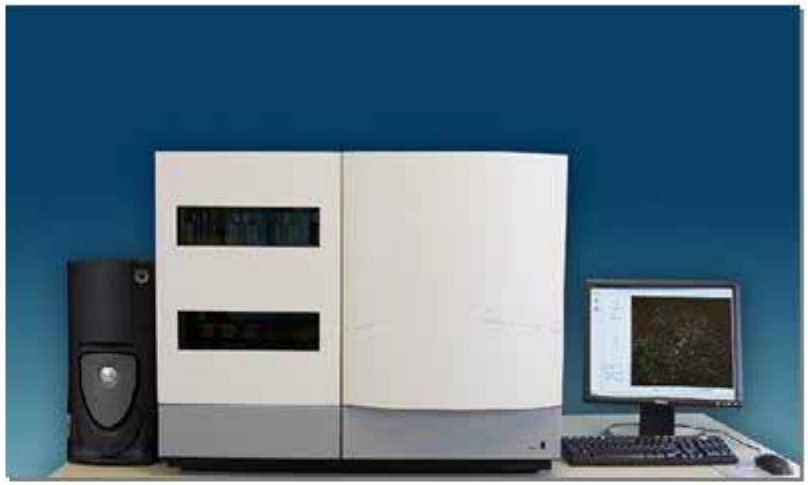
Sequencing by synthesis using chemiluminescence

- 20Mb of sequence for ~$5,000 in running costs
- Quality is similar to early ESTs (97-98% at best)
- We have no clone information, so no read pairings
- Homopolymer…

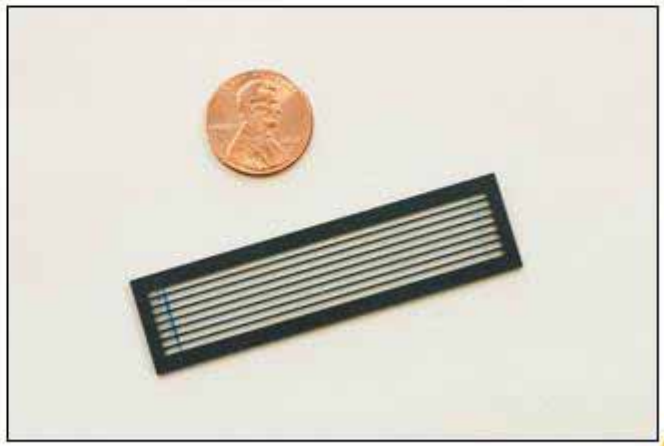


The Solexa Genome Analysis System

- **System Components:**
  - Solexa 1G Genetic Analyzer
  - Cluster Station Instrument
  - Consumables
  - Reagents
  - Software



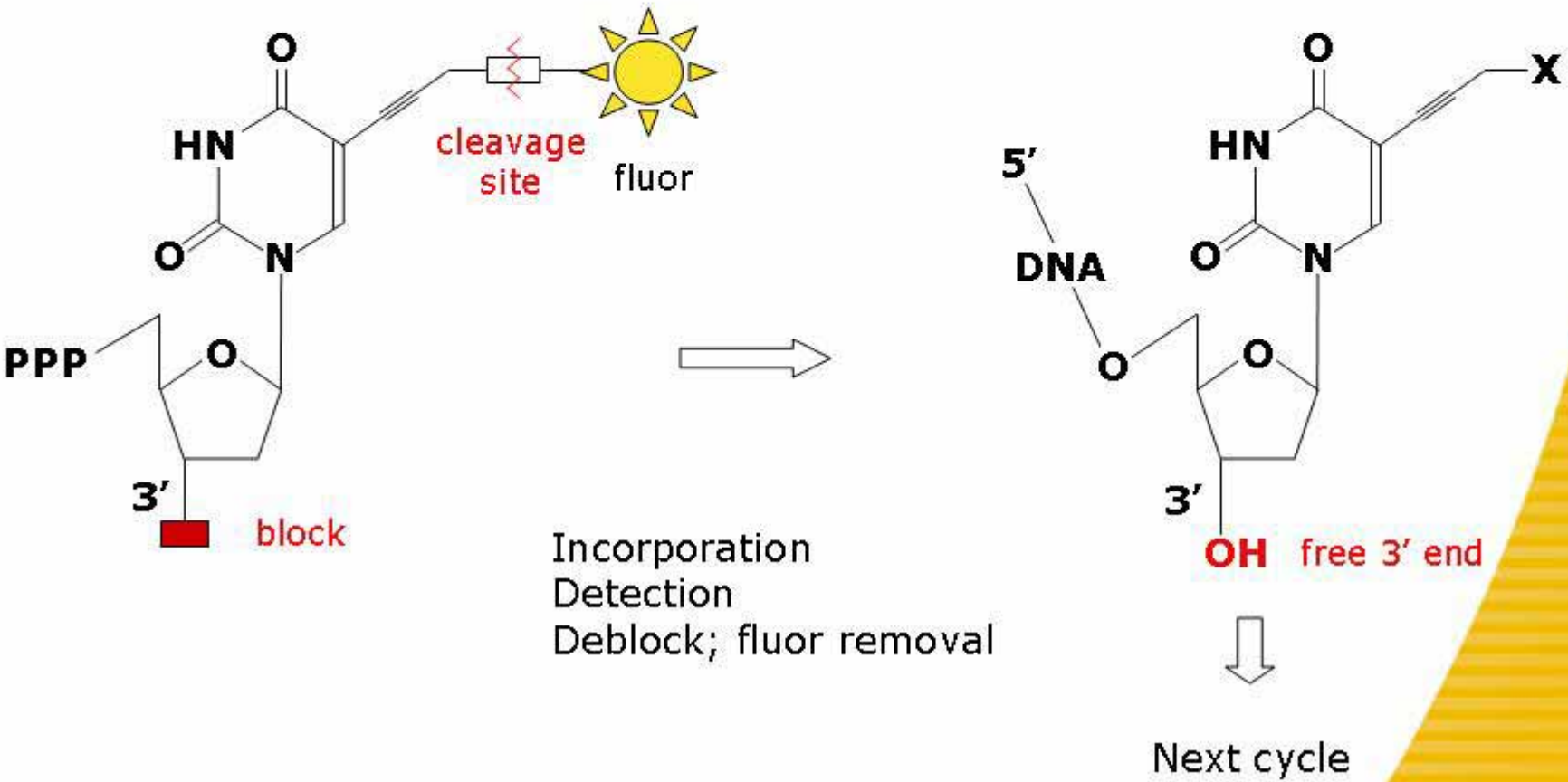
- **Applications:**
  - DNA Sequencing
  - Expression Profiling
  - miRNA Discovery & Analysis
  - Other applications in 2007

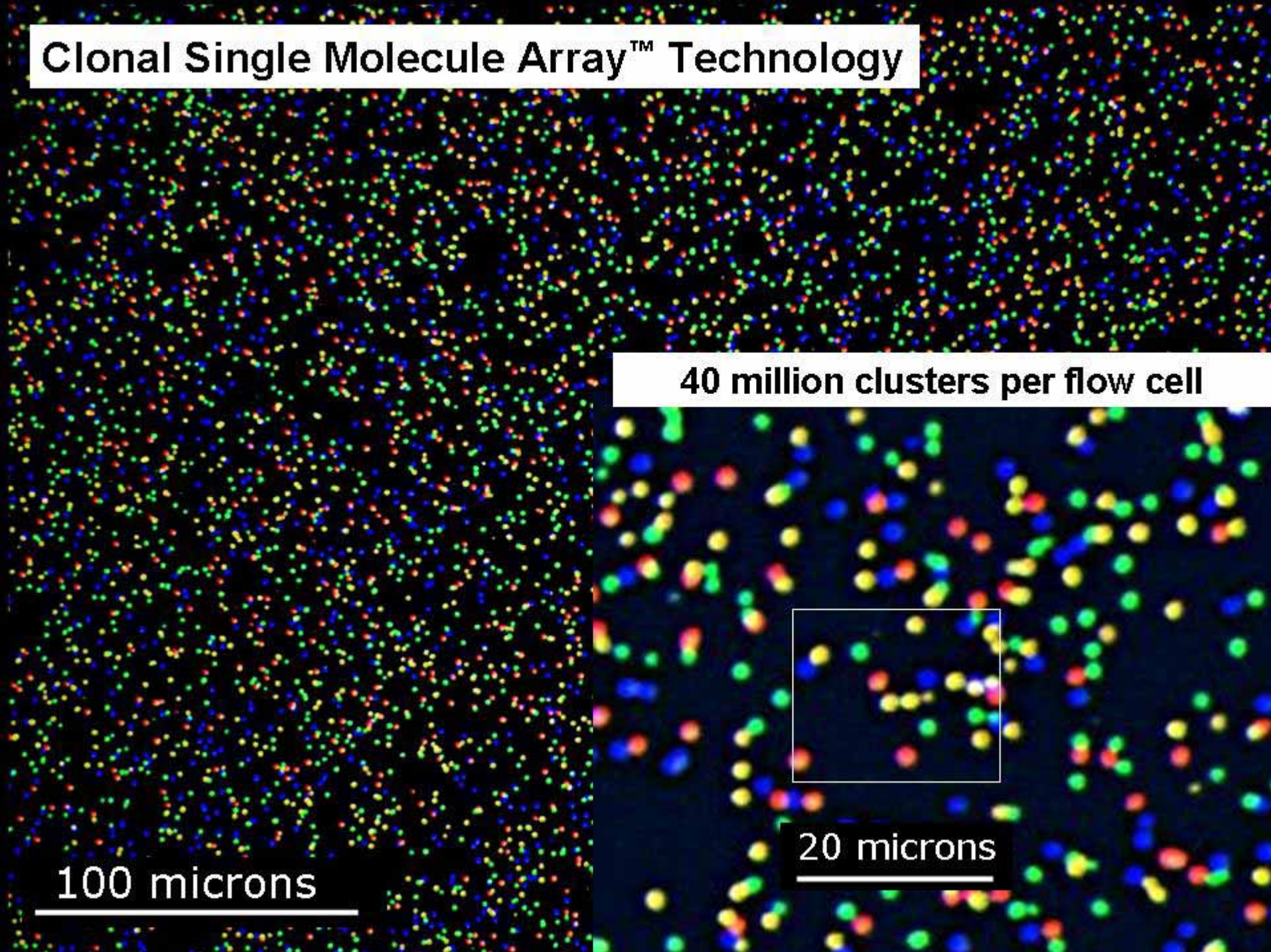


1Gb of sequence for < $3,000 in running costs

Reversible Terminator Chemistry

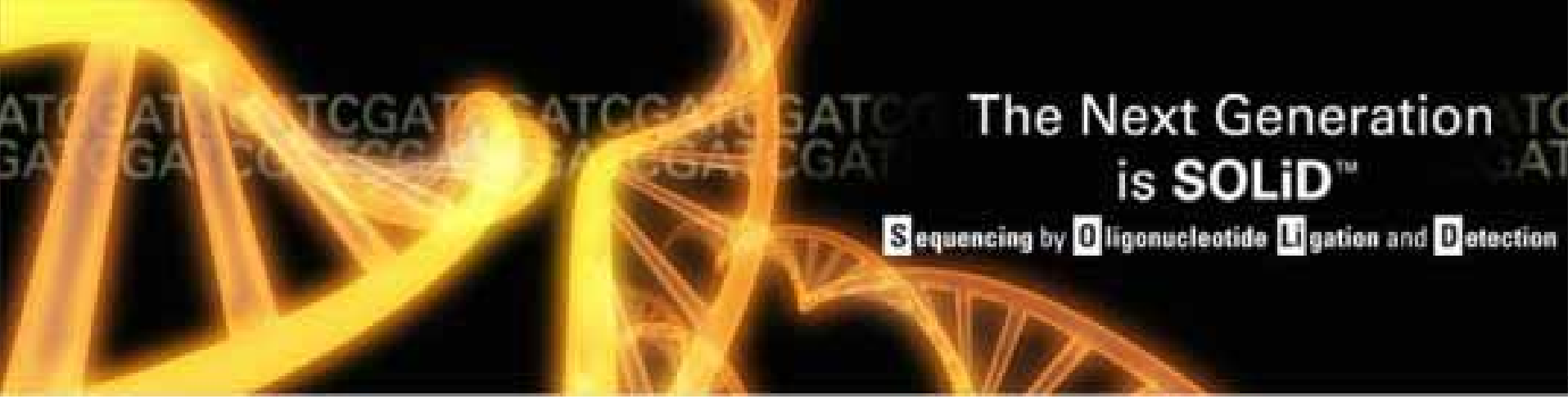
- All 4 labelled nucleotides in 1 reaction
- Higher accuracy
- No problems with homopolymer repeats

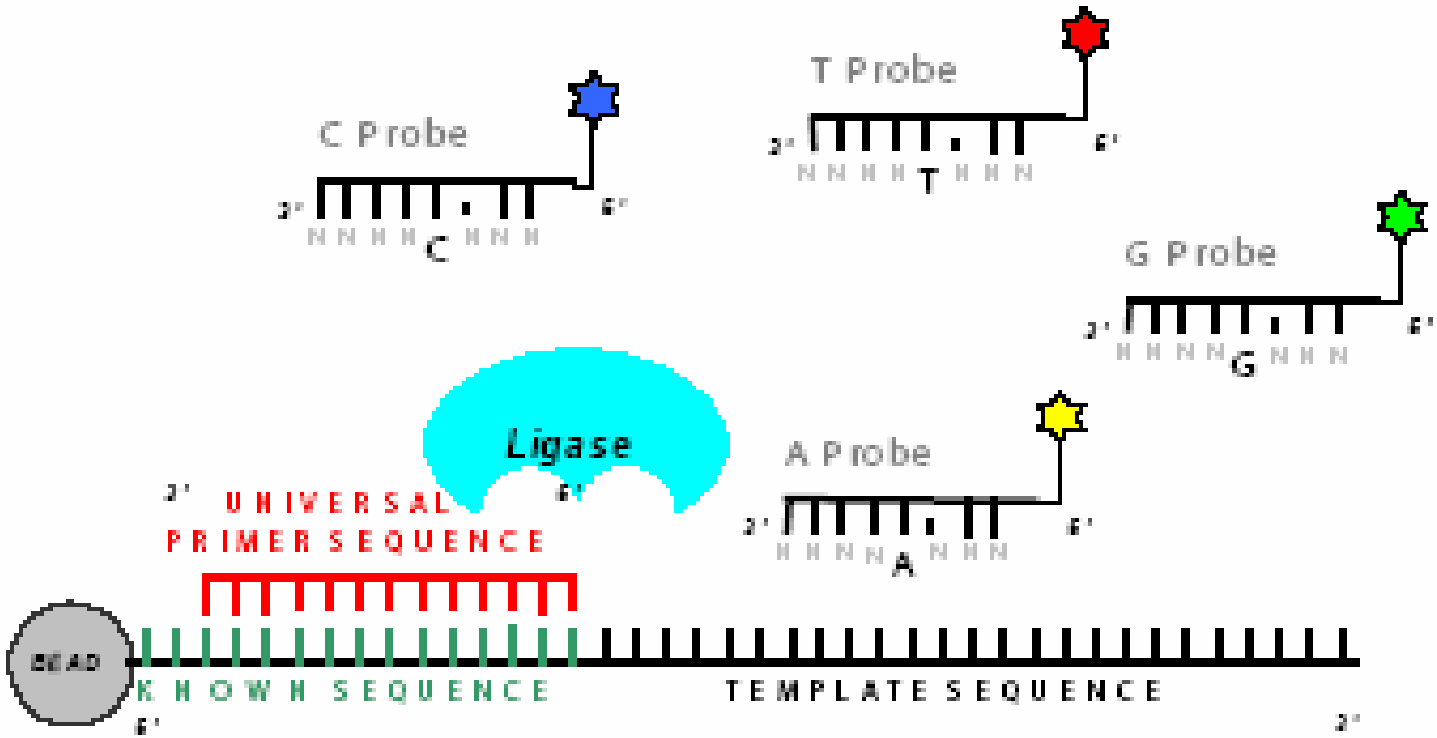



Clonal Single Molecule Array™ Technology
40 million clusters per flow cell
100 microns
20 microns

Polony sequencing / ABI SOLiD

- George Church's group invented "polony" method
- Since developed by Agencourt
- Now bought by ABI
- Similar to Solexa – no wells, small beads, 4-color fluorescent detection, about 1G per run, about $3,000 per run
- Uses ligation of nucleotide-specific probes rather than reversible terminators


The Next Generation
is SOLiD™
Sequencing by Oligonucleotide Ligation and Detection


C Probe
3'
5'
N N N N C N N N
T Probe
3'
5'
N N N N T N N N
G Probe
3'
5'
N N N N G N N N
A Probe
3'
5'
N N N N A N N N
Ligase
3'
5'
UNIVERSAL
PRIMER SEQUENCE
BEAD
KNOWN SEQUENCE
TEMPLATE SEQUENCE
5'
3'

Proof of concept experiments with 454 technology:

- Soybean genome (1.2GB): 2 “454 runs”

- 717,383 successful reads
- 80,176,681 base pairs sequenced
- 112 base pairs average read length

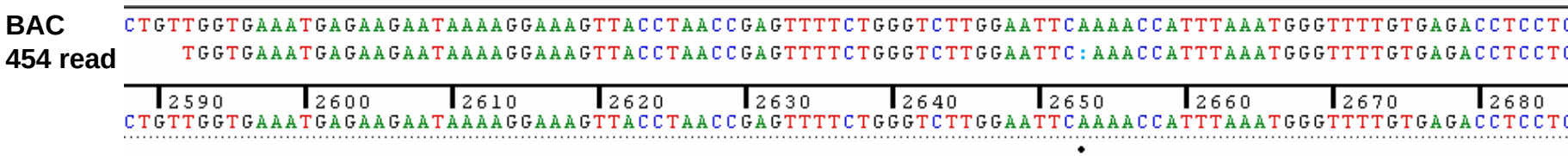
- A genomic survey with ~7% coverage.

- Soybean cyst nematode genome (100MB): 10 “454 runs”
- 3,277,846 reads
- 379,047,339 base pairs
- 116bp average length

- An agmagenomic sequence with ~80% coverage

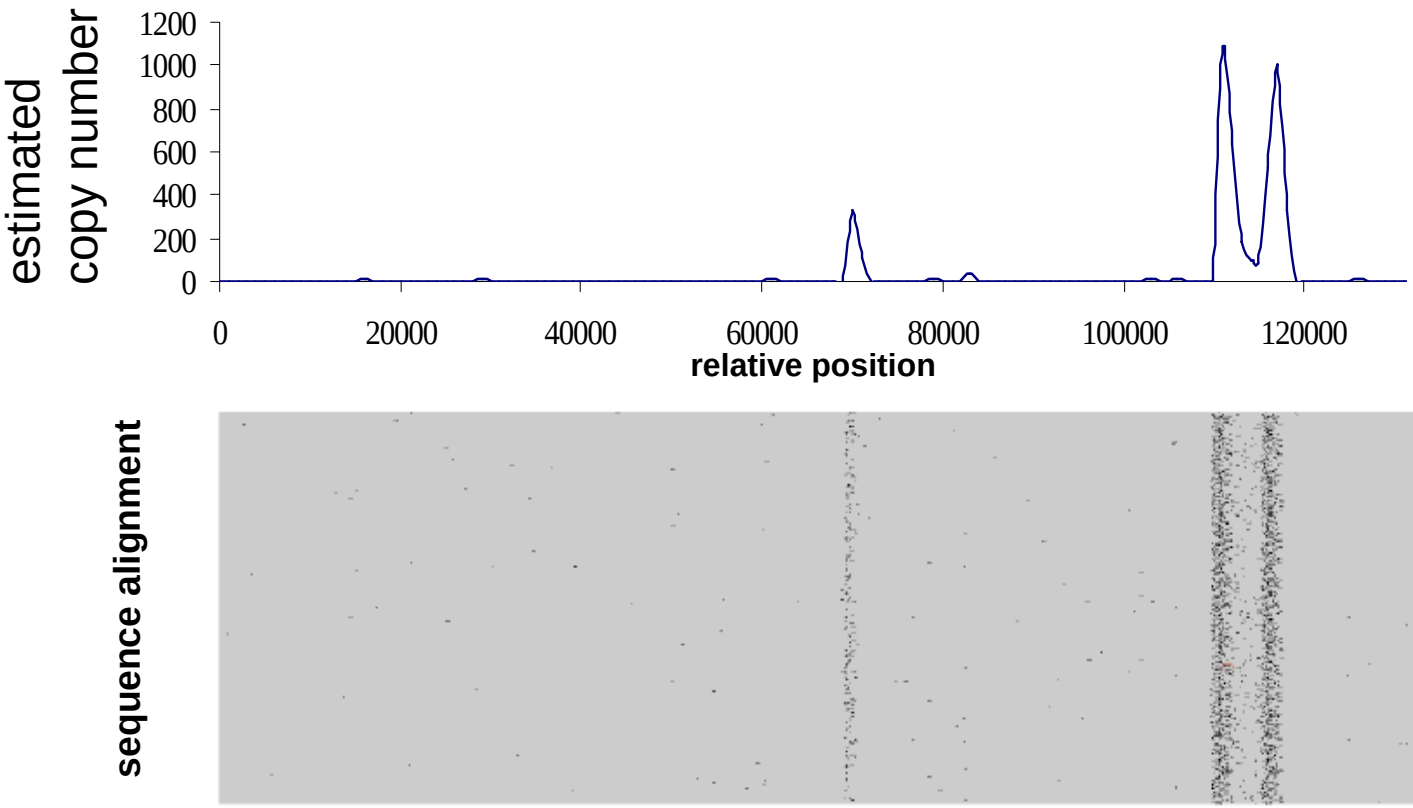
Read quality and genomic match

- We matched 160 genomic reads to the *chs* BAC sequenced by the Clough and Vodkin labs using BLAT.

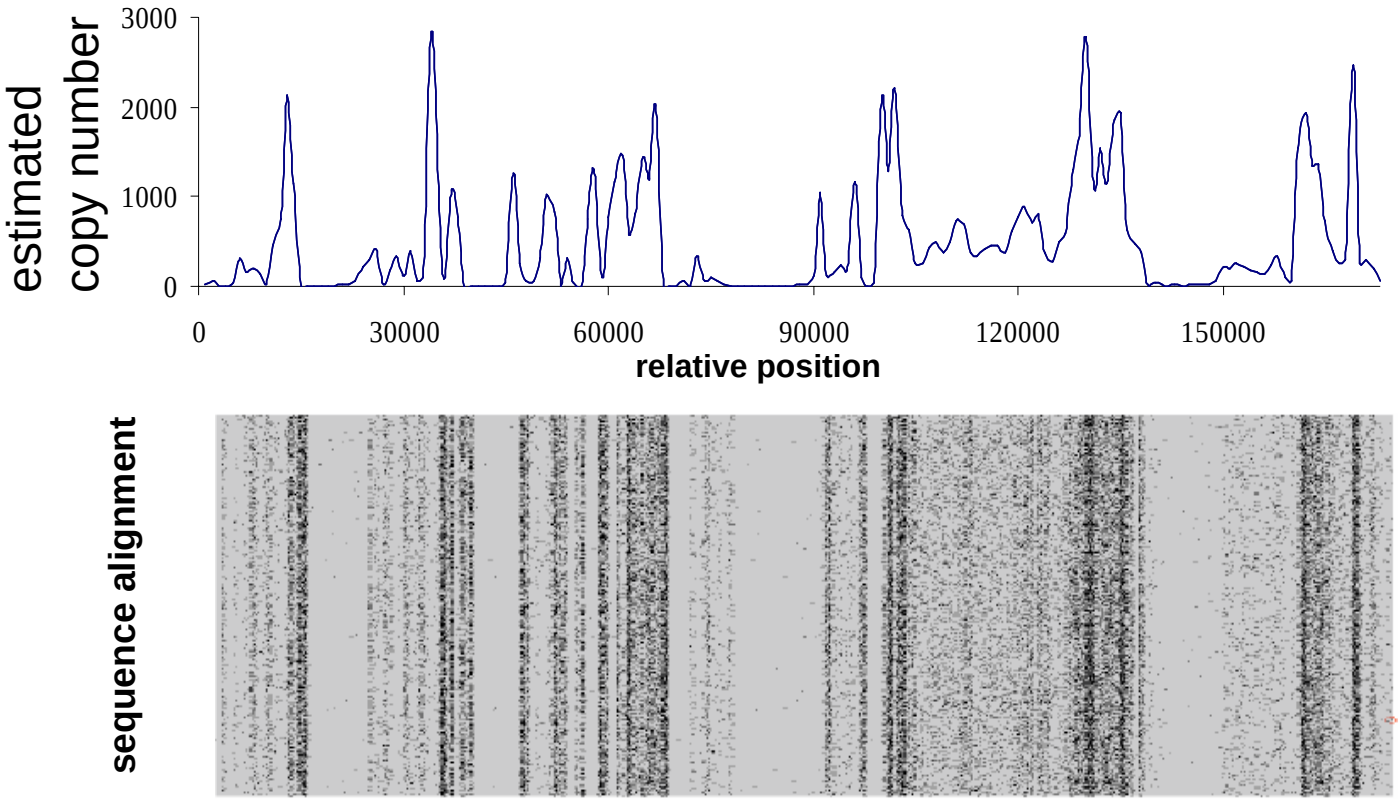


- There are an average of 6 disagreements per read, or about 95% sequence accuracy.
- Mismatches are more common at the ends, as with Sanger sequencing

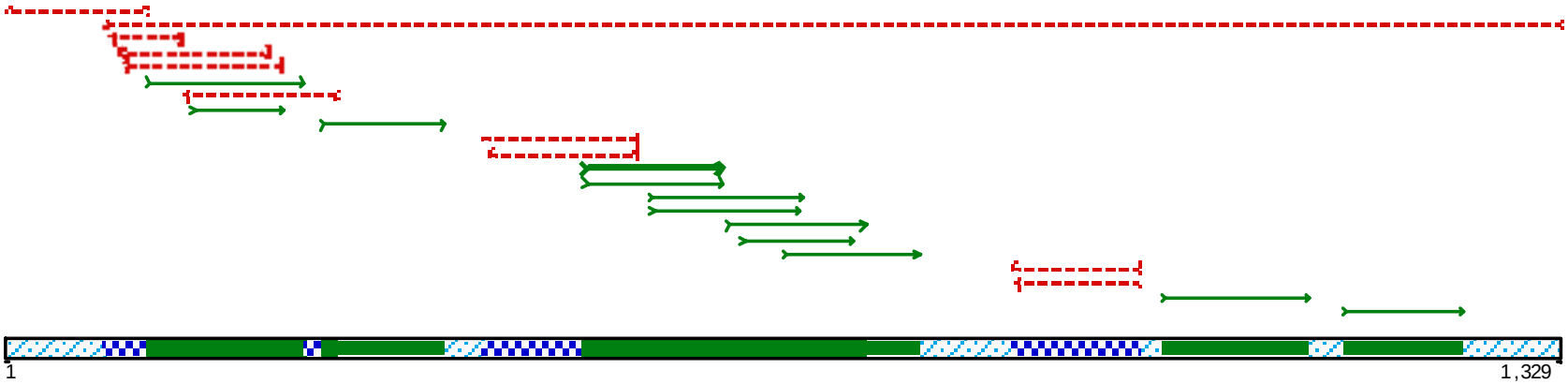
Euchromatic BAC clone (*CHS* locus)



Pericentromeric clone (GM_WBb0078A23)



High coverage SCN sequencing



Acknowledgements

UIUC

Kranthi Varala

Kankshita Swaminathan

Ying Li

Gene Robinson

Amy Toth

Lila Vodkin

Dave Neece

Steve Clough

Adam Thomas

454

Kent Lohman

Lei Du

Gerry Irzyk

Others

Scott Jackson