

MAPPING FOR THE FUTURE OF OUR FOOD



National C-FAR
1302 Longworth
House Office Bldg
Washington DC
25 July 2011



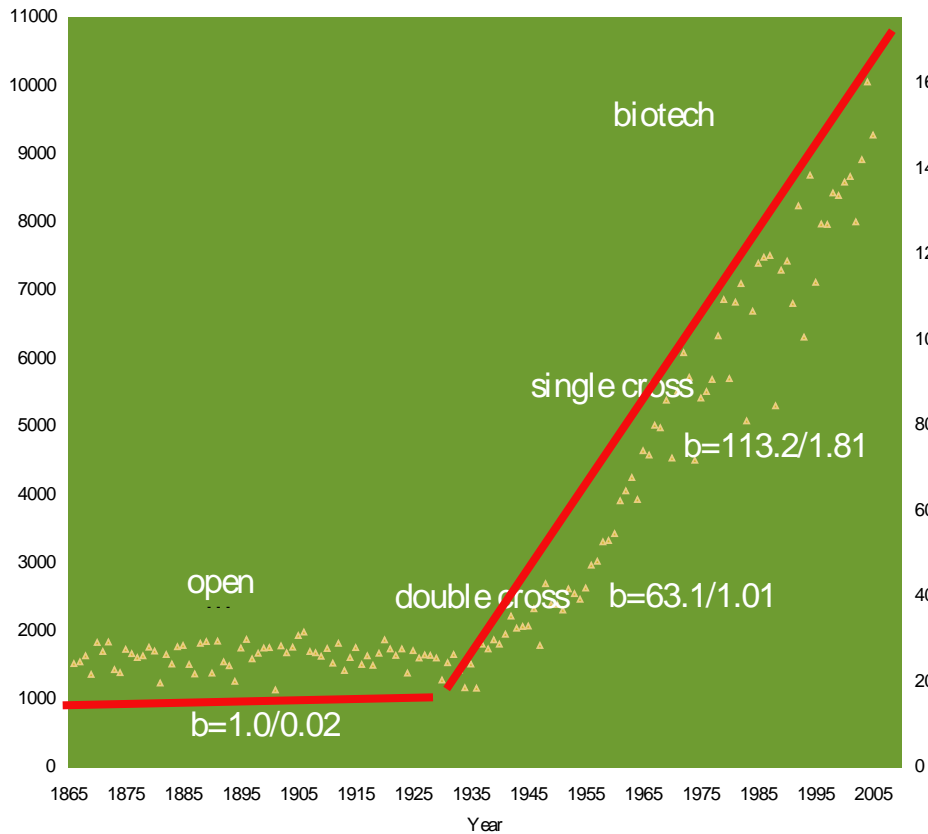
Patrick S. Schnable
Iowa State University
Data2Bio, LLC

Some Definitions

- **Genes:** composed of DNA (command lines within a computer program)
- **Genome:** collection of all genes that together control growth and development (computer program-collection of all command lines)
- **Genetics/Functional Genomics:** determining which genes control traits (debugging a computer program)

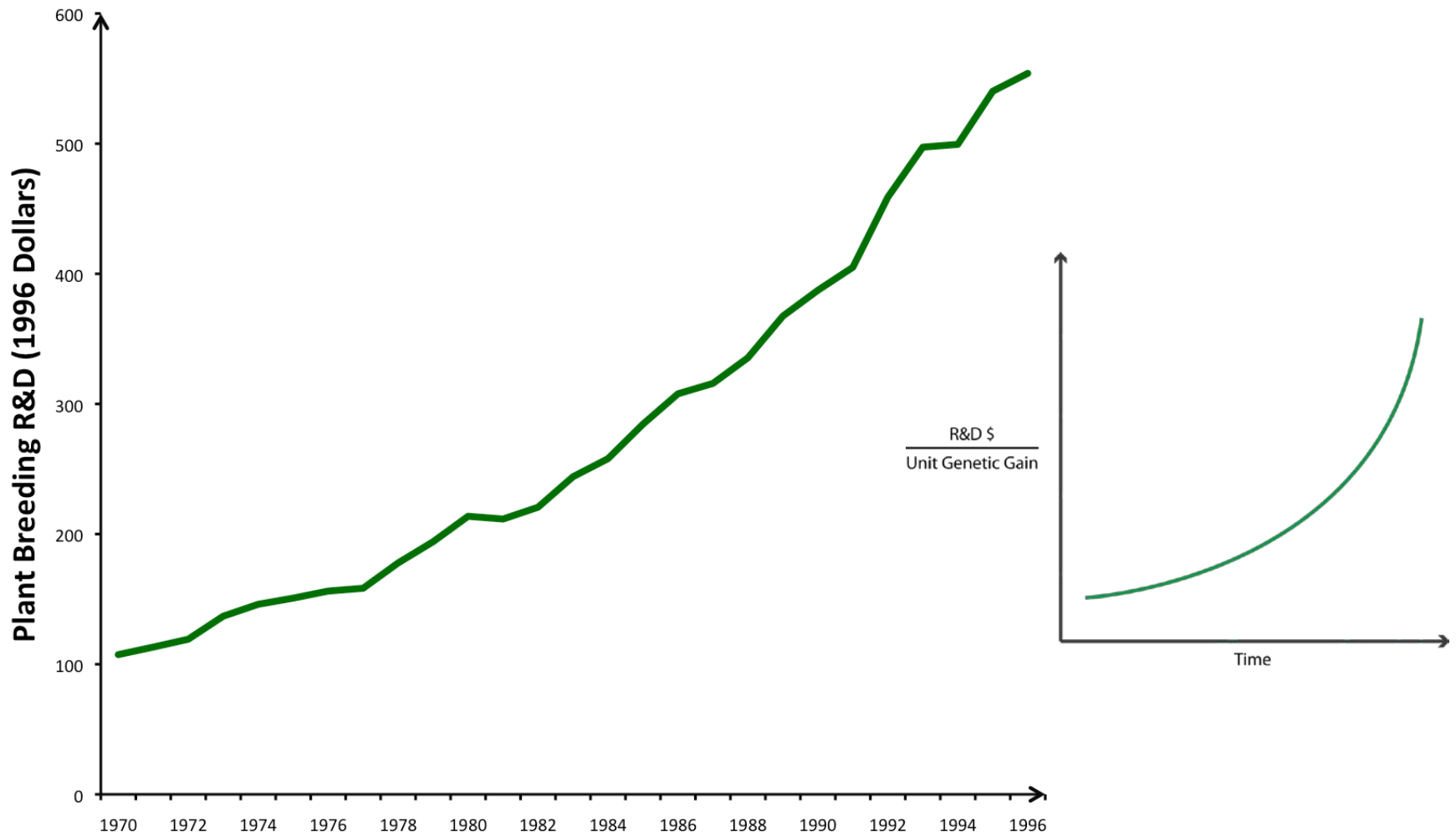
Plant breeders have been successful

U.S. Corn Grain Yields



Troyer (2006) Crop Sci. 46:528-543

Private-Sector R&D Investment in Plant Breeding is Increasing (constant dollars)



USDA Report: The Seed Industry in U.S. Agriculture

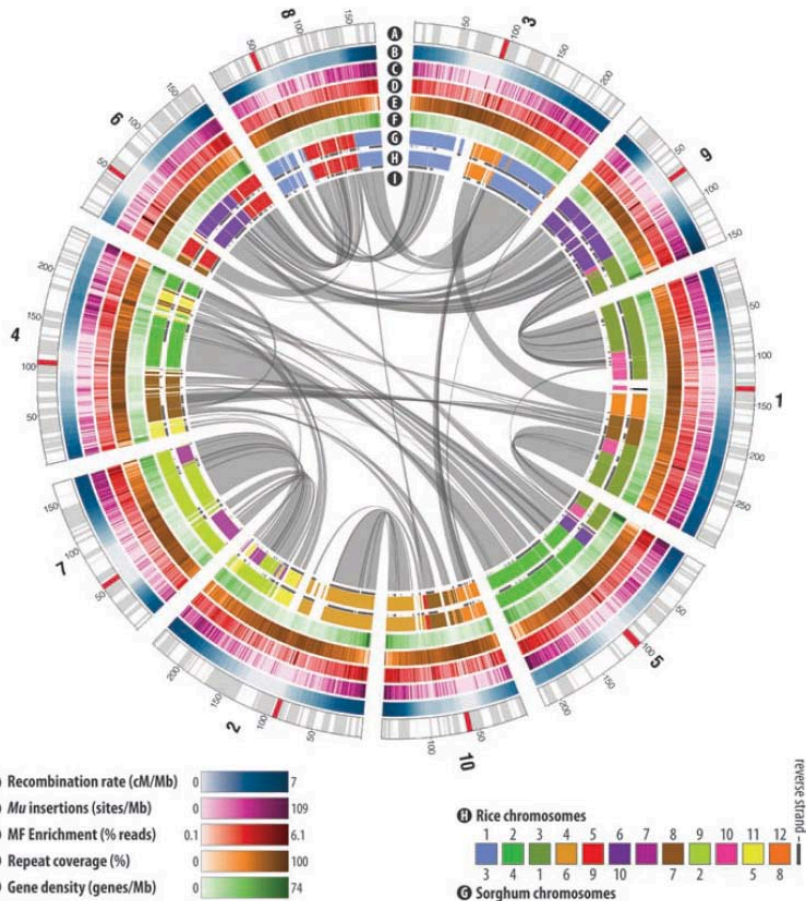
and new challenges await us...

- Increasing demands for food, feed, fiber and fuel
- Decreasing amounts of arable land
- Agricultural inputs:
 - Increasing costs (e.g., nitrogen)
 - Reduced availability (e.g., water)
 - Undesirable ecological impacts
- Increasing climate variability
 - Droughts/floods
 - Temperature extremes
 - Pests & diseases
- Need to be innovative

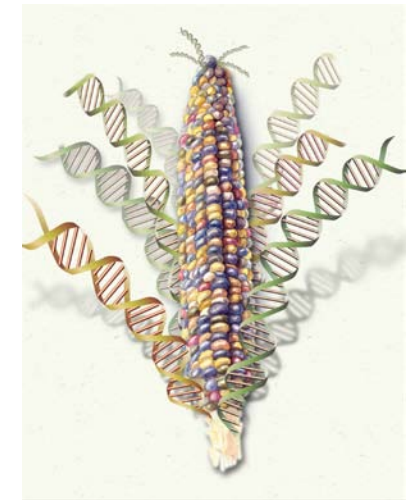


NYT, 1/23/06

The \$30M B73 Maize Genome Sequencing Project



Schnable, Ware et al., 2009



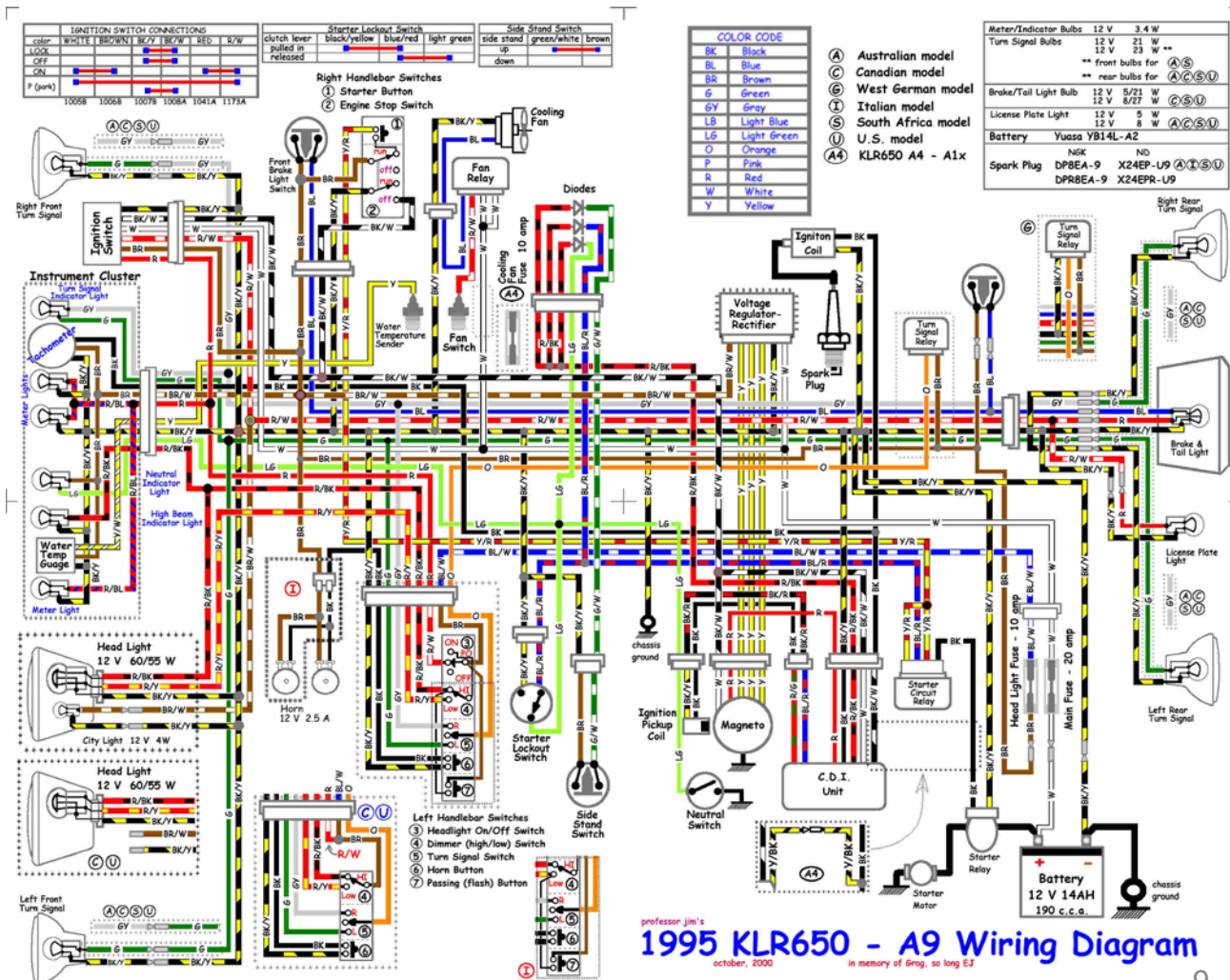
The Maize Genome Sequencing Project, Rick Wilson, PI

Genome Projects are Analogous to the Lewis & Clark Expedition



- Expensive and require extensive planning/coordination
- Exploration of the unknown; expect surprises
- Generates lots of information *that requires subsequent analysis*

A Genome Sequence is Analogous to...



NCGA Workshop on Functionality and the Maize Genome, March 2009

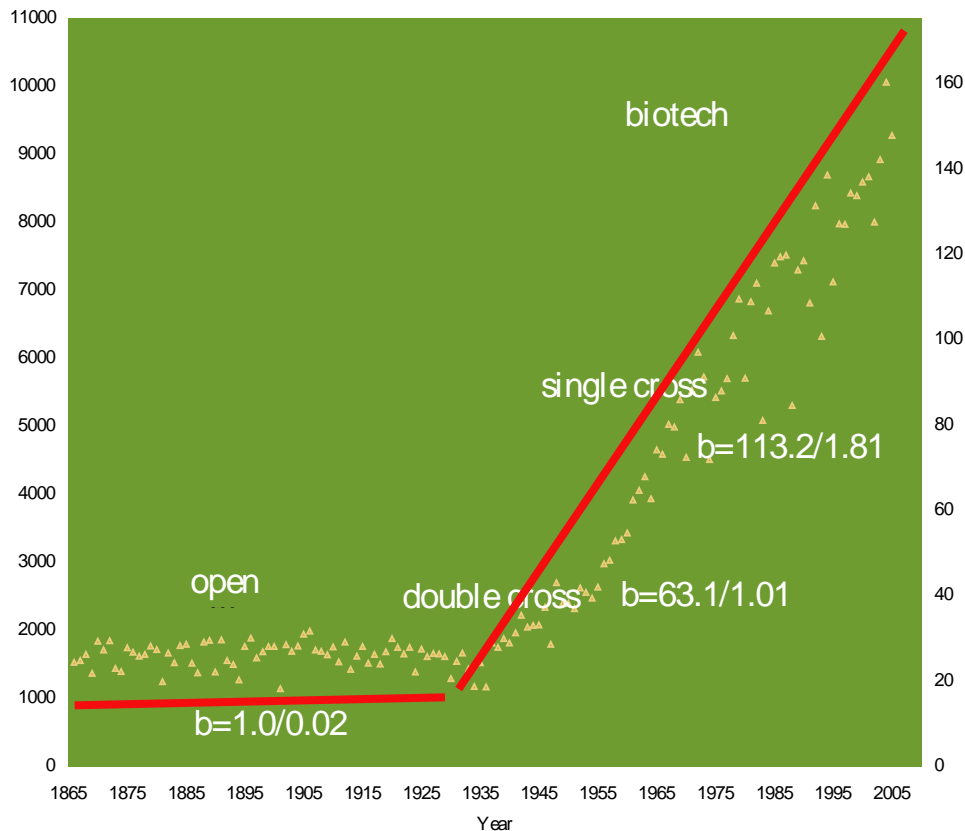
- Participants: Growers, public-sector and private-sector researchers
- Goal: To understand the genetic basis of traits in maize -- traits that are the foundation for improving food, fuel, and fiber crop yields and to translate the answers into crop improvement.

A Major Recommendation from the NCGA Workshop

- In most species (including maize), 1/3-1/2 of genes do not have even a predicted function and the relationships of most others to traits is unknown
- Continued public-sector R&D investment to link the ~50,000 maize genes with “traits”

What does society get for this investment?

Corn Grain Yields

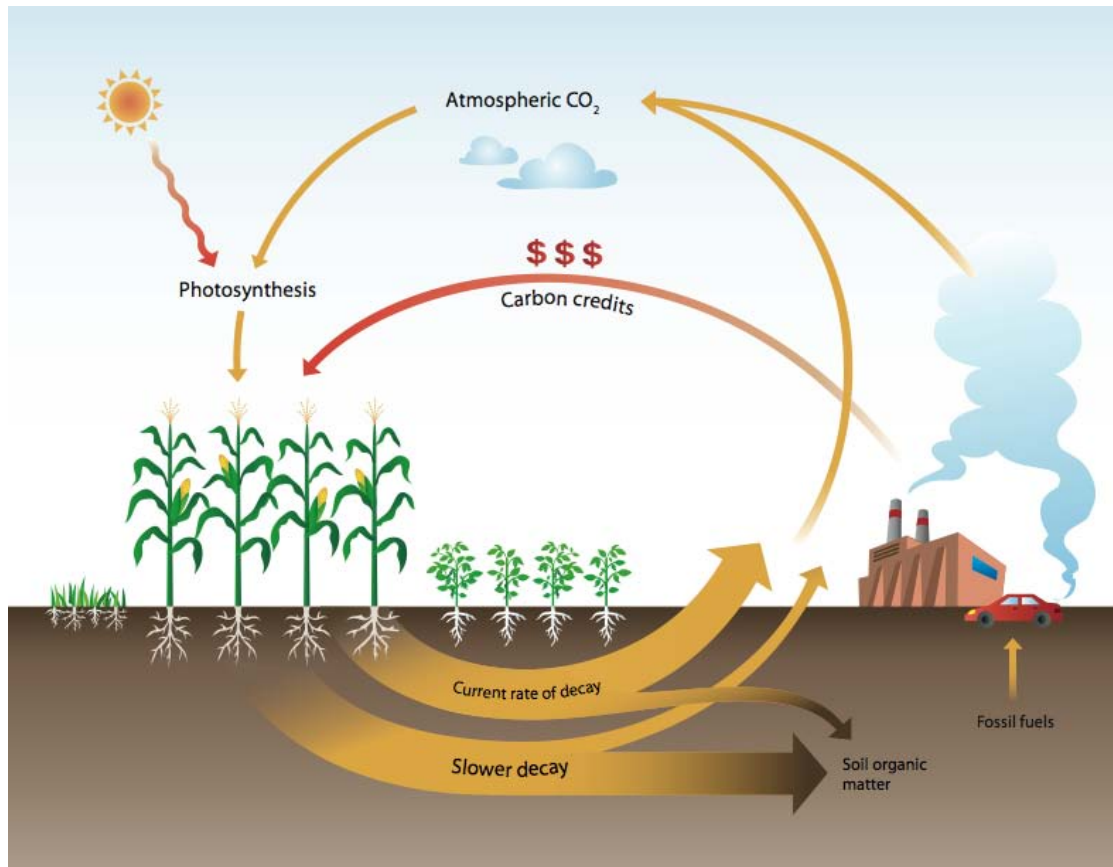


- Increasing demands for food, feed, fiber and fuel
- Decreasing amounts of arable land
- Agricultural inputs:
 - Increasing costs (e.g., nitrogen)
 - Reduced availability (e.g., water)
 - Undesirable ecological impacts
- Increasing climate variability
 - Droughts/floods
 - Temperature extremes
 - Pests & diseases
- Innovation (new traits)

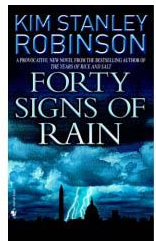
Center for Carbon-Capturing Crops

Rising atmospheric CO₂ concentrations/global climate change

Produce crops with cell walls that are “resistant” to microbial degradation



- Sequester atmospheric carbon in soil
- Increase soil organic matter
- Increase farm income



Maize is Genetically Diverse

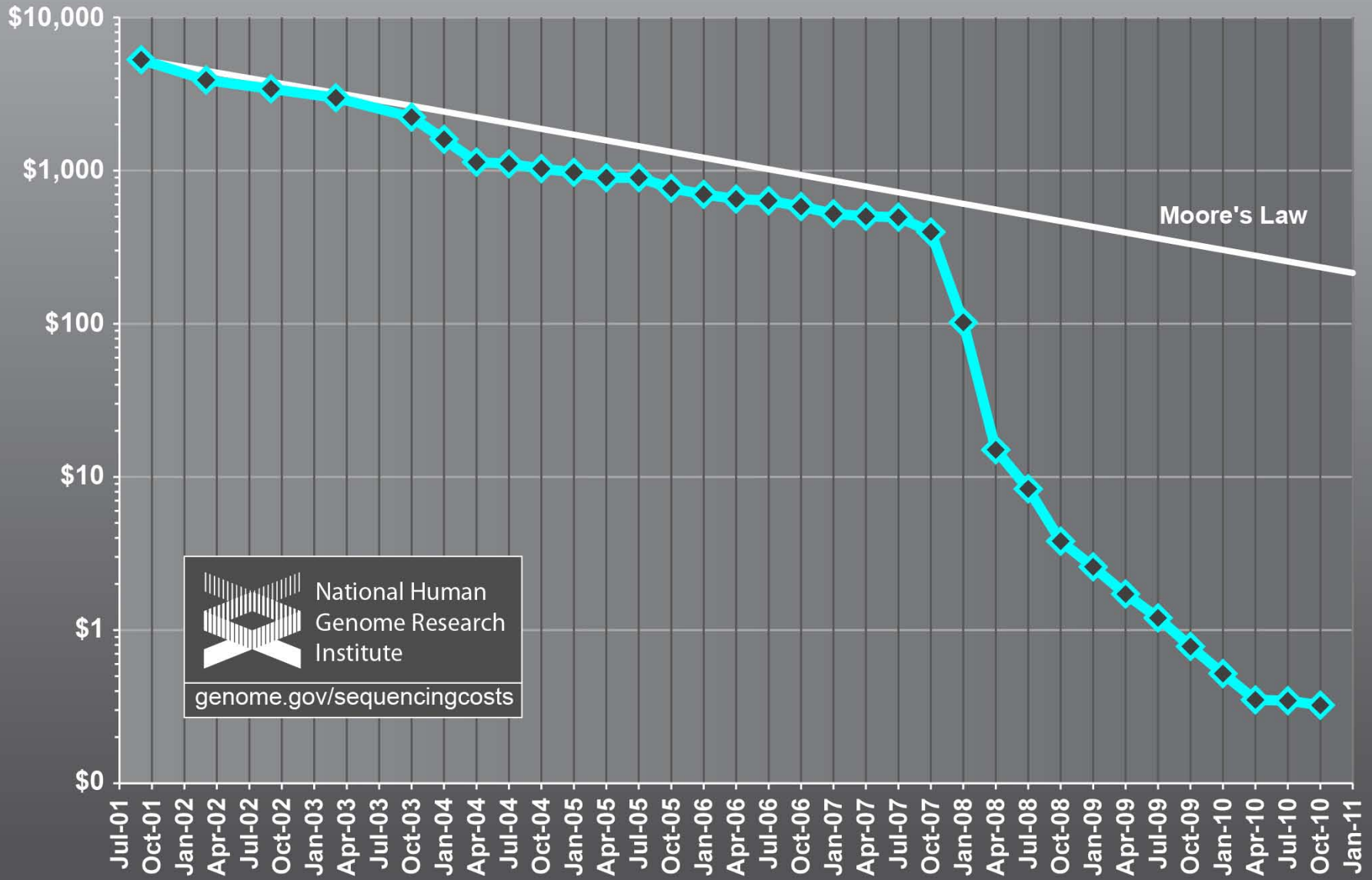


=



Similar Amounts of Genetic Diversity

Cost per Megabase of DNA Sequence



 National Human
Genome Research
Institute
genome.gov/sequencingcosts



Model 377
Version 3.4.1
ABI200
Version 3.3.1b2

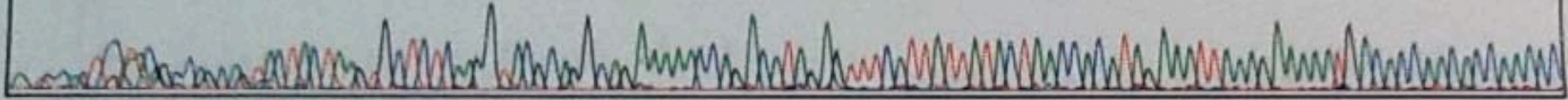
27_a5par_QZ1265_130367
Hong Yao_15497_Run ID: 2015
a5par_QZ1265_130367
Lane 27

15497

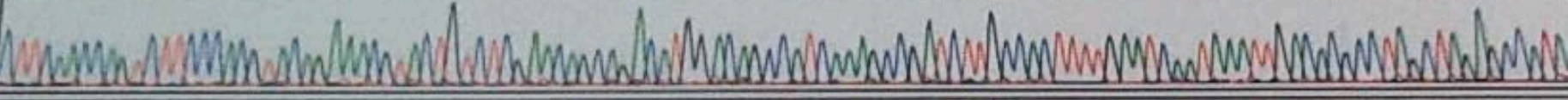
Signal G:39 A:83 T:53 C:36
DT377(BDv3)v1.mob
dR George Matrix v.3 11 June 01
Points 1273 to 9716 Pk 1 Loc: 1273

Page 1 of 1
Thu, Aug 15, 2002 8:23 AM
Wed, Aug 14, 2002 4:52 PM
Spacing: 11.63(11.63)

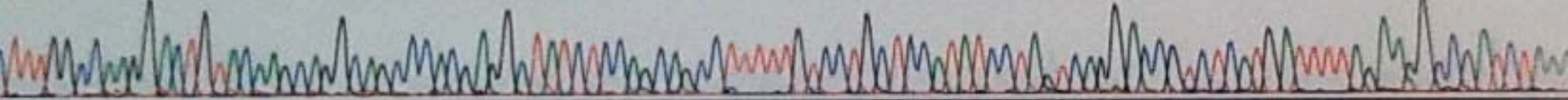
AA TCCACTTC TNC TCCCAO CAO TACTACTAAGTGTCTCTCAAGTGTACGCCAGGCCAGAAAACCCAGTNCATAGAGTTTCAATTAATYATAC TAAACCAC ATAG AATY AAG AAAAAO ACACCAACCCAAACC
10 20 30 40 50 60 70 80 90 100 110 120



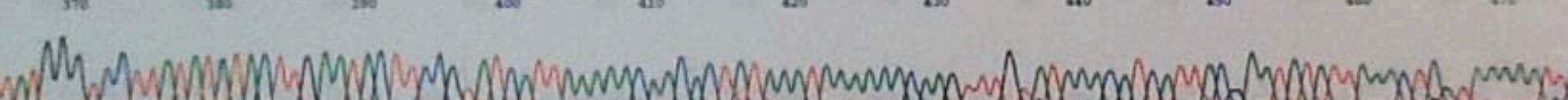
CTTCAACCCAGTCTTCCCAAGTACAGAAACGTACTGTCTCTCAAGTGTACGCCAGGCCAGAAAACCCAGTNCATAGAGTTTCAATTAATYATAC TAAACCAC ATAG AATY AAG AAAAAO ACACCAACCCAAACC
130 140 150 160 170 180 190 200 210 220 230 240



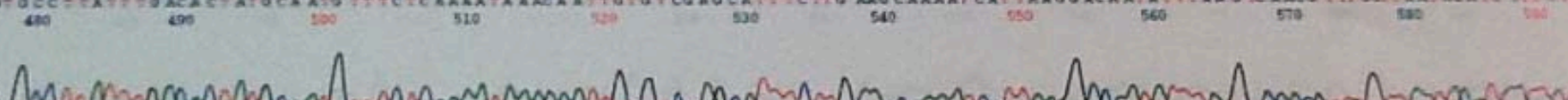
CTTTGCCCCAAGGACTGTACCAAGCAAGGCCAGCCAGGAGTCTCTCAAGTGTACGCCAGGCCAGAAAACCCAGTNCATAGAGTTTCAATTAATYATAC TAAACCAC ATAG AATY AAG AAAAAO ACACCAACCCAAACC
250 260 270 280 290 300 310 320 330 340 350 360



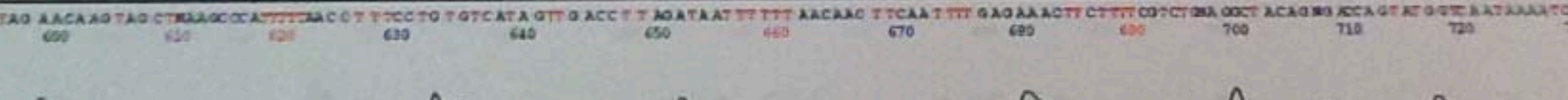
TTATGCGTCTTATACACTTGTATAGTCTCCGTATCATTAAAAACAAACACTATAAATAAAAAACAAATTTGTATAAACATCATTACGAATATATATAAAGGCGGATCTAAAT
370 380 390 400 410 420 430 440 450 460 470



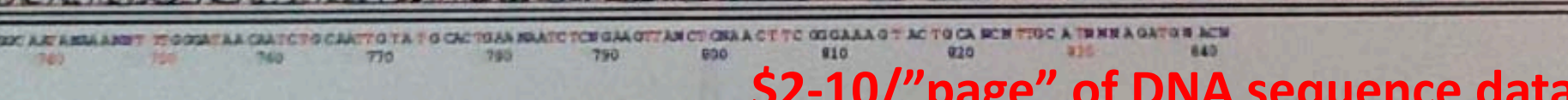
GTGCCCTATTGACACTATGCAATGTTCTCTCAAAATAAACAATTTGTGTCGAGCATTCTTTGAAACAAAATCATTAAGGACAATAATTTAAGGTCAAAGTTTGGTTATACATCCTTCT
480 490 500 510 520 530 540 550 560 570 580 590



TATAGAACAAATAGCTTAAAGCCATTCTTCAACCTTCCCTGTGTCATAATTGACCCTTAAATAATTTTAAACAATTCAAATTTGAGAAACTTCTTTGCTCTGAAAGCTACAGGAGCCAGTATGCTATATAAAT
600 610 620 630 640 650 660 670 680 690 700 710 720



TATAGAACAAATAGCTTAAAGCCATTCTTCAACCTTCCCTGTGTCATAATTGACCCTTAAATAATTTTAAACAATTCAAATTTGAGAAACTTCTTTGCTCTGAAAGCTACAGGAGCCAGTATGCTATATAAAT
730 740 750 760 770 780 790 800 810 820 830 840



\$2-10/"page" of DNA sequence data

A Single Illumina HiSeq 2000 Paired-End Lane

Illumina HiSeq 2000



Sequencing Cost
\$6,000

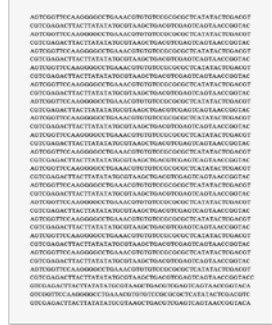
Paired-End Lane:

90 x 10⁶ Reads
2 x 100 bp reads
18 x 10⁹ bp
8 days



18 (50) GB data

Letter Size Paper



↑ 11 inches ↓

← 8.5 inches →

Letter Size Paper:
1" margins
12 pt font size
Times New Roman
~2,500 character
(UPPERCASE)

HP LaserJet P4015 Black & White Printer



Printer:
~50 pages/min
(single sided)

Cartridge:
~24,000 pages

\$0.0008/"page" of DNA sequence data

How much paper is needed to print data from 1 paired-end lane of HiSeq?



18 x 10⁹ bp / 2.5 x 10³ = **7.2 x 10⁶ pages**

1 ream = 500 pages

Or



7.2 x 10⁶ / 500 = **14,400 reams**

Or



14,400 / 10 = **1,440 cases** (12" x 10" x 18")
A pile of cases 12 ft x 12 ft x 12ft

And



1 printer + 300 cartridges + 100 days

Supply Cost:

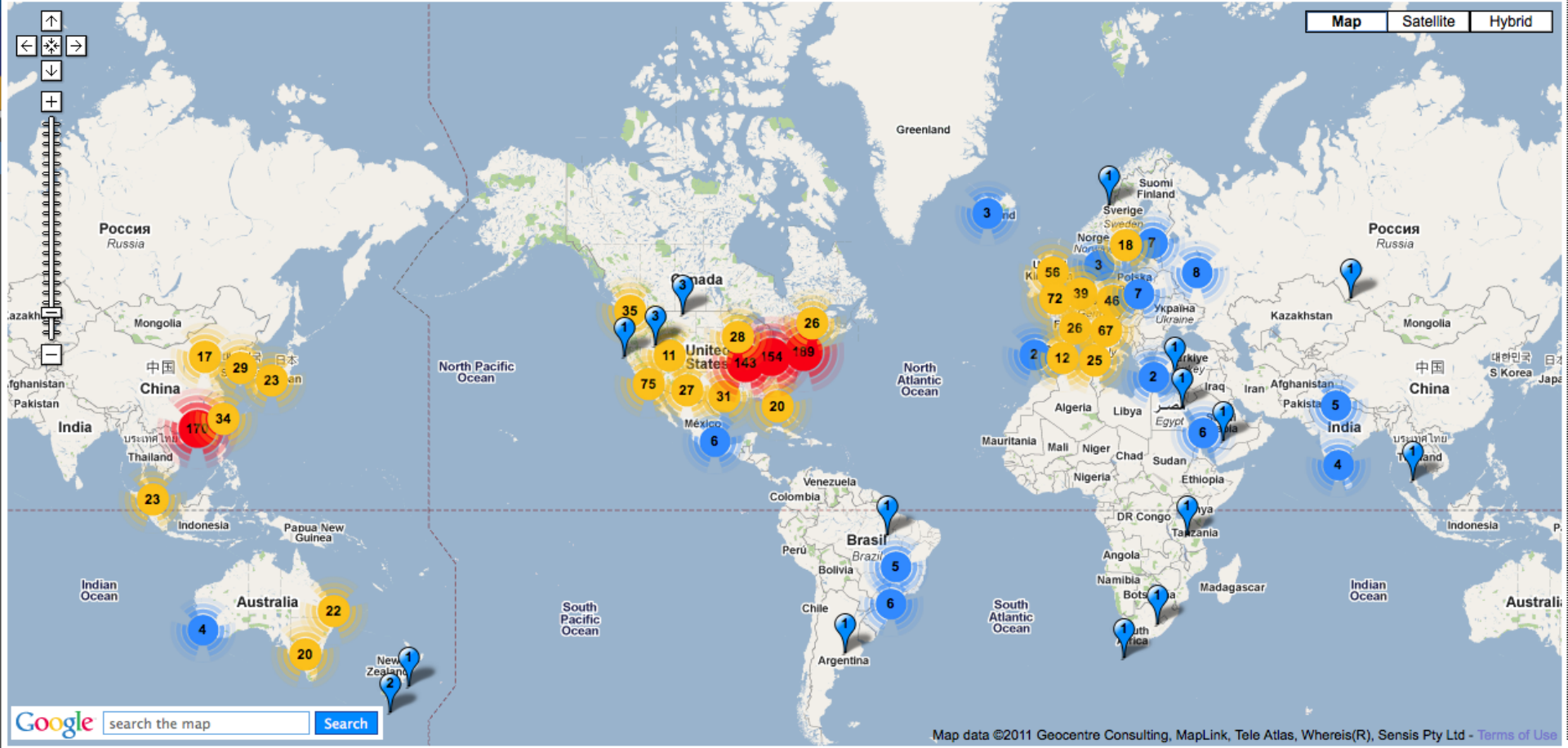
1,440 cases @ \$40/each + 300 cartridges @ \$300/each = **\$147,600**

NGS Technologies are Changing the Game

>1,600 NGS Instruments

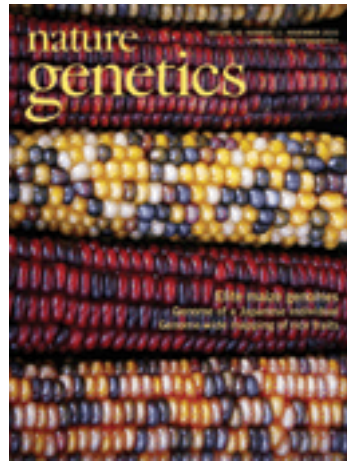
Next Generation Genomics: World Map of High-throughput Sequencers

Show all platforms Illumina GA2 Illumina HiSeq Ion Torrent PacBio Polonator Roche/454 SOLID Service Provider



Map data ©2011 Geocentre Consulting, MapLink, Tele Atlas, Whereis(R), Sensis Pty Ltd - Terms of Use

Sequenced Six Add'l Maize Inbreds



Lai et al., 2010



IOWA STATE UNIVERSITY

UNIVERSITY OF MINNESOTA

Summary

- Applies to all crops
- Plant breeders have been very successful
- But it gets harder each year to maintain rate of yield increase
- Substantial prior Federal investments in plant genomics
- New challenges require continued public-sector investments in agricultural R&D

Travel Support from:

- National Corn Growers Association
- Iowa Corn Growers Association
- International Wheat Genome Sequencing Consortium
- Eversole Associates

Further Information

- USDA Report: The Seed Industry in U.S. Agriculture:
<http://www.ers.usda.gov/publications/aib786/aib786.pdf>
- USDA Report “World Agricultural Supply and Demand Estimates”
<http://www.usda.gov/oce/commodity/wasde/latest.pdf>