Agricultural Genomics: The Rise of the Genomes

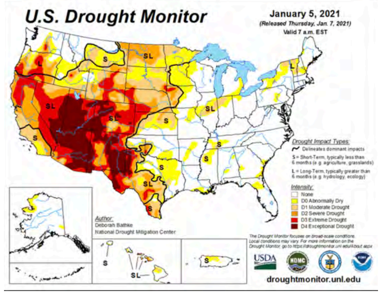
Dr. Doreen Ware, USDA ARS

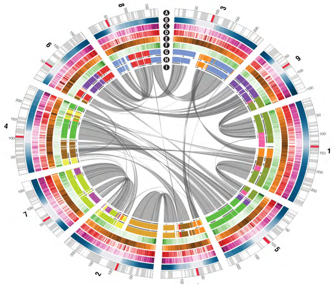
Agricultural Biotechnology: Emerging Technologies and Insights

January, 27, 2021

Advancing Agriculture Through Collaborative Research on Crop & Model Species

Outline

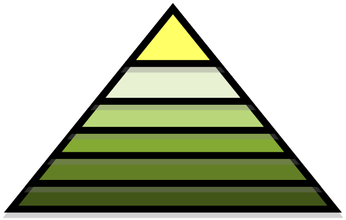




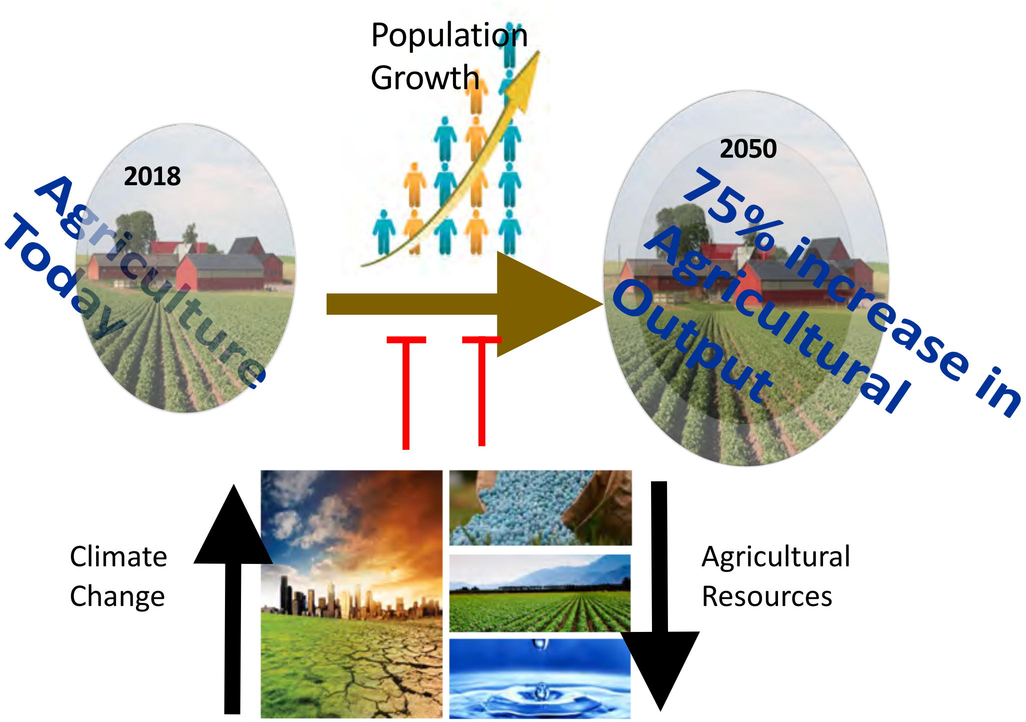
- Agricultural Drivers
- Maize genome 16 years
  - What’s in a genome
  - Improvements in Sequencing technology we can Continue to evaluate approaches to develop reference assembly and annotations
- Genome/Biology enabled agriculture
  - Sorghum EMS population
    - Forward and reverse genetics
  - Breaking down complex trait: Yeild & Quality
    - Plant architecture: flower
    - Response to environment: water, heat, nitrogen, disease
- Biology & “Big Data”
  - collaborative infrastructure
  - Future

CC NB-ARC

LRR



Drivers for Agriculture: Sustainability and Defense



Population Growth

2018

Agriculture Today

2050

75% increase in Agricultural Output

Climate Change

Agricultural Resources

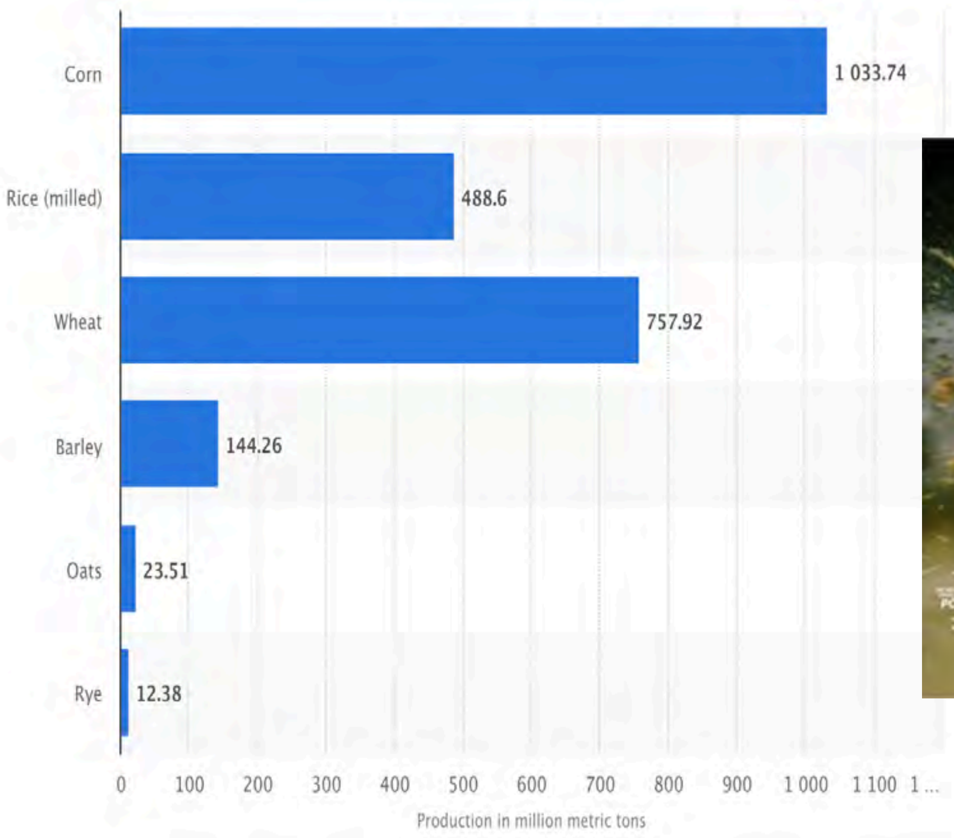
Nature Editorial, *How to feed a hungry world*. 2010

BREEDING FOR 2050 AND BEYOND. Prepare for plant and animal pests and disease while they are still offshore. Design plants for new environments.

CLIMATE CHANGE. Collect and preserve natural diversity.

Maize is the highest world-wide production crop

Reference genomes are foundation tools for ensuring food security & environmental sustainability



Acknowledgements

Maize Diveristy Project (NSF & USDA ARS) ~ 2003- present

USD AARS **Ed Buckler**, Sherry Flint-Garcia, Mike McMullen, Jim Holland, Peter Bradbury, Doreen Ware Cornell QiSun, UCDavis Jeff Ross-Ibarra

Maize Genome Project (NSF, DOE, USDA) ~ 2005-2010

University of Washington, Rick Wilson, CSHL: Rob Martienssen, Dick McCombie, Doreen Ware, University of Arizona, Rod Wing, University of Iowa Pat Schnable

B73 Genome & Annotation Improvements V4 (NSF & USDA ARS) ~ 2015 -2017

CSHL staff: Yinping Jiao, Bo Wang, Mike Campbell, Josh Stein, Sharon Wei, Doreen Ware, Dick McCombie, Eric Antinou
PacBio: David Rank, Paul Peluso, Jason Chin, Tyson Clark, Ting Hong, Elizabeth Tseng
BioNano: Alex Hestie, Tiffany Liang, Jinghua Shi. USDA ARS Mike McMullen, Kate Guill, University of Georgia: Kelly Dawe , Jonathan Gent, University of Hawaii: Gernot Presting, Kevin Schneider, Thomas Wolfgruber
Institutes: Cold Spring Harbor Laboratory, USDA ARS, Pacific BioSciences, BioNano, University of Georgia, University of Hawaii

Transcriptome Variation (USDA ARS) 2016- 2019

CSHL staff: Yinping Jiao, Bo Wang, Mike Campbell, Josh Stein, Sharon Wei, Doreen Ware, Dick McCombie, Sara Goodwin
PacBio: Elizabeth Tseng, Tyson Clark, Ting Hong, Kevin Eng, Primo Baybayan
Institutes: Cold Spring Harbor Laboratory, USDA ARS, Pacific BioSciences

Maize Nested Association Mapping Panel Reference Assemblies (NSF & USDA ARS) 2018- present

University of Georgia: Kelly Dawe, University of Iowa: Matt Hufford University Minnesota Candy Hirsch
CSHL staff: Josh Stein, Kapeel Chougule Sharon Wei, Doreen Ware, Dick McCombie, Sara Goodwin
PacBio: Emily Hatas, Paul Peluso, Jason Chin, Cortiva: Victor Llaca, Kevin Fengler, Greg May, DNA Nexus: Brent Hannigan, Chai Fungtammasan, Brittanny O'Sullivan, NIH: Adam Phillippy, Serge Koren
Institutes: Cold Spring Harbor Laboratory, USDA ARS, University of Georgia, University of Iowa, Pacific BioSciences, DNANexus, NIH

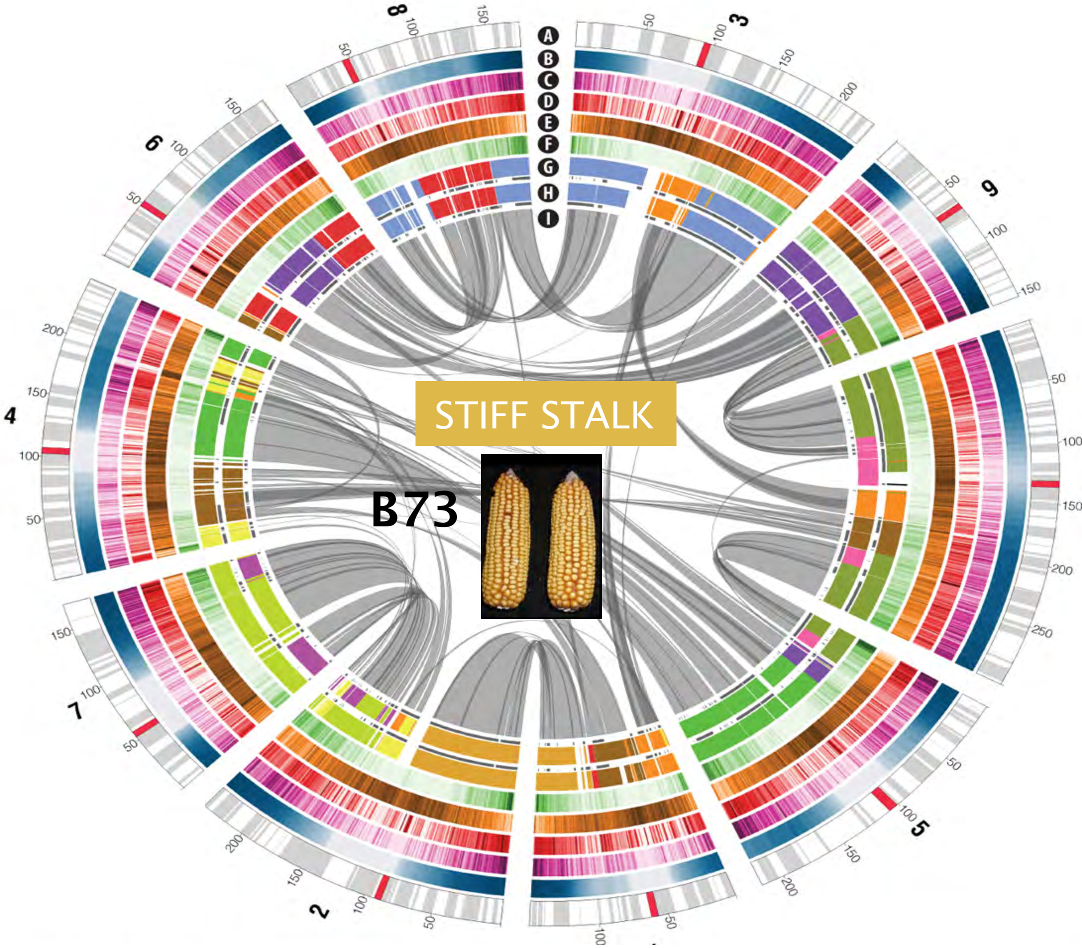
Gramene www.gramene.org (NSF & USDA ARS) 2002- present NSF IOS-1127112

PI: Doreen Ware, PI (USDA ARS, CSHL) & Pankaj Jaiswal, Co-PI (OSU), Paul Kersey (Ensembl Genomes EBI), Helen Parkinson (ATLAS EBI), Lincoln Stein (Reactome OCIR), Crispin Taylor (ASPB)
Gramene @ CSHL Andrew Olson, Joshua Stein, Sharon Wei, : Marcela Karey Monaco,
Institutes: Cold Spring Harbor Laboratory, Oregon State University, EMBL – European Bioinformatics Institute, Ontario Institute for Cancer Research, American Society for Plant Biologists

USDA ARS Sorghum Functional Genomics and Germplasm improvement: 2015- present

USDA Doreen Ware, **Zhanguo Xin**, Chad Hayes, Yinghua Huang, Gloria Burow, Ratan Chopra, John Burke
SorghumBase: Nick Gladman, Yinping Jiao, Bo Wang, Kapeel Chougule, Andrew Olson, Sharon Wei, : Marcela Karey Monaco
Corteva,

Maize Genome is over 12 years old



Washington University in St.Louis

Arizona Genomics Institute AGI

CSH

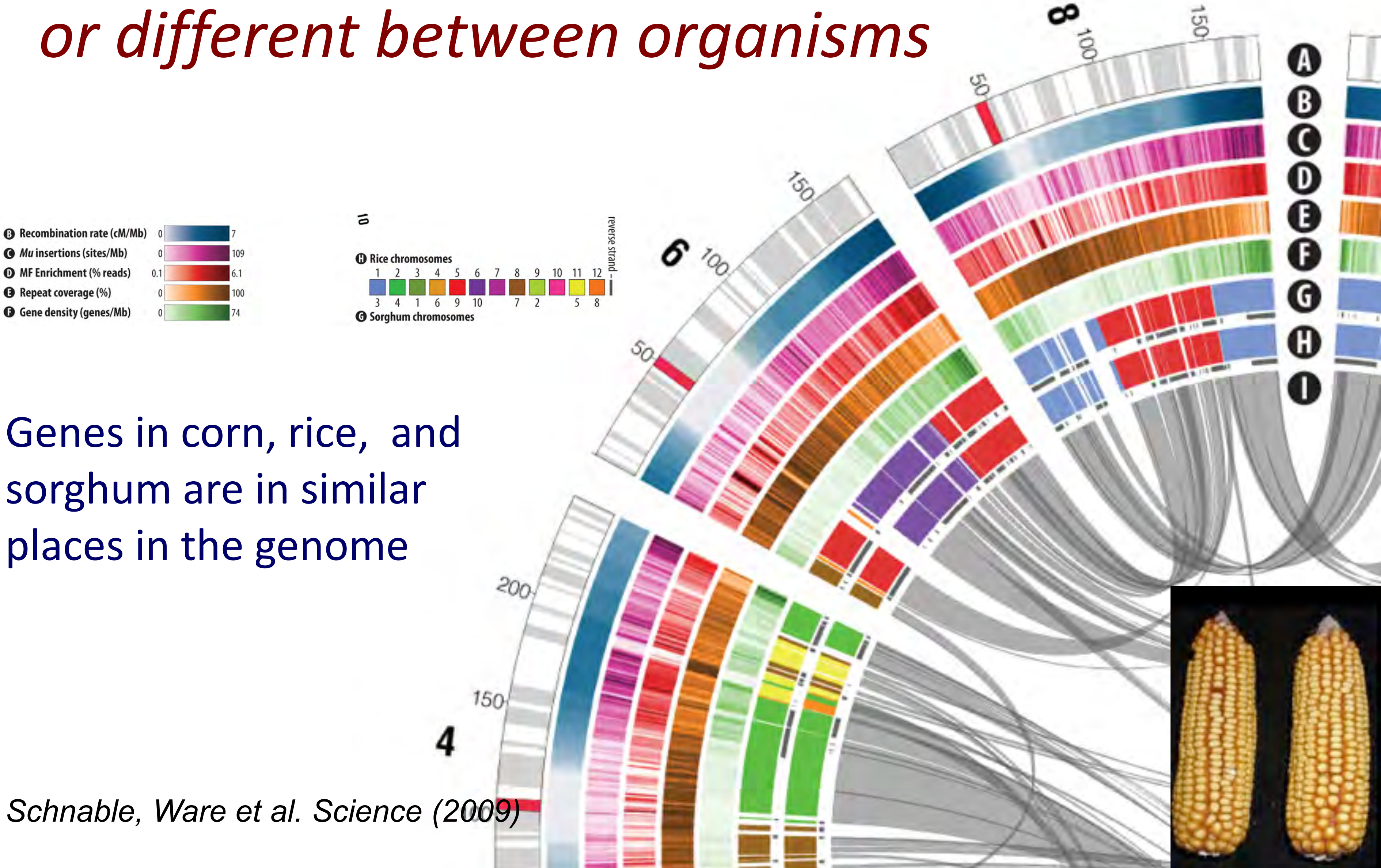
IOWA STATE UNIVERSITY

Maize Sequence Consortium (NSF, DOE, USDA) PI Rick Wilson

Schnable, Ware et al. Science (2009)

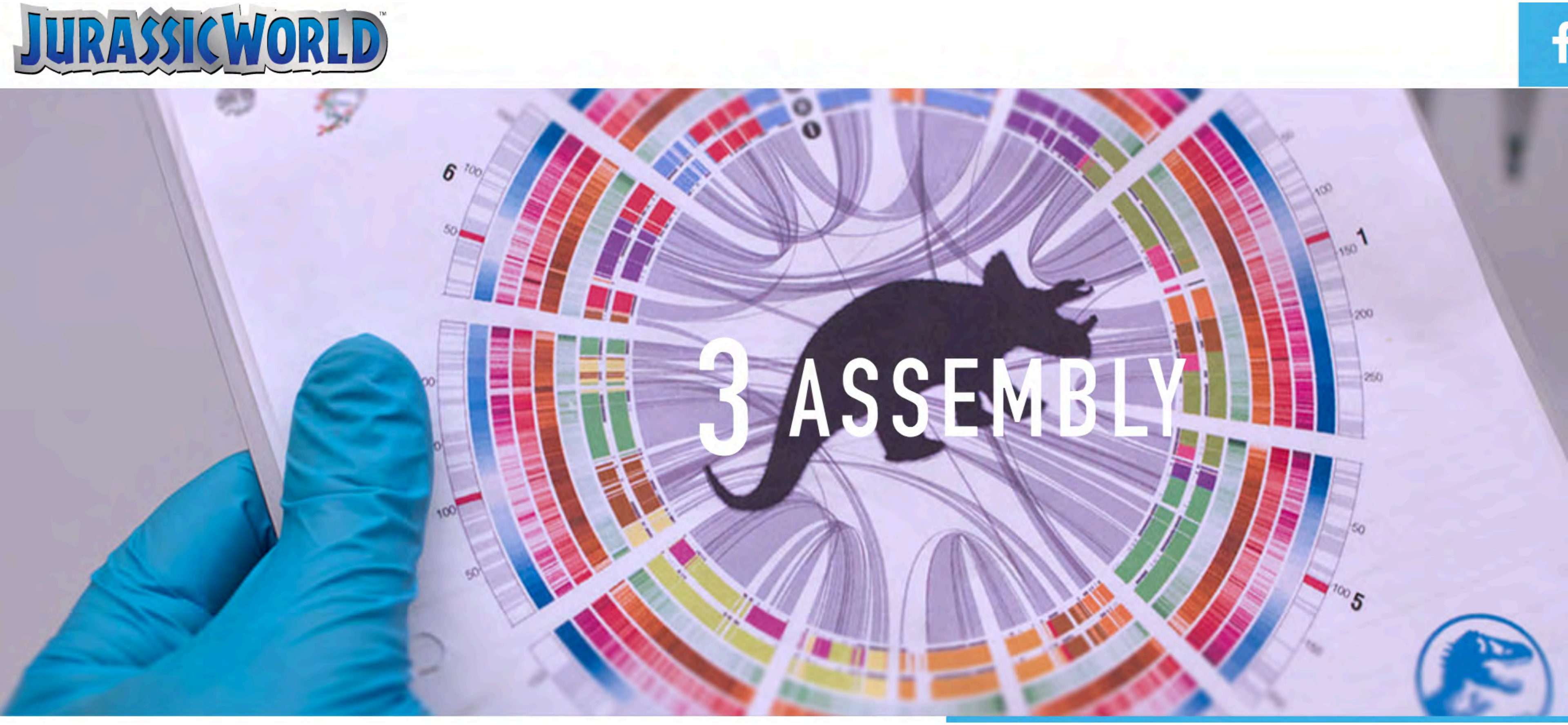
Sequence genomes provides us the parts list and allows us to see what is the same or different between organisms

Genes in corn, rice, and sorghum are in similar places in the genome



Schnable, Ware et al. Science (2009)

Triceratops Genome



http://www.jurassicworld.com/creation-lab/

Genomes sequences allow us to see all the variations (mutations) that exist in nature

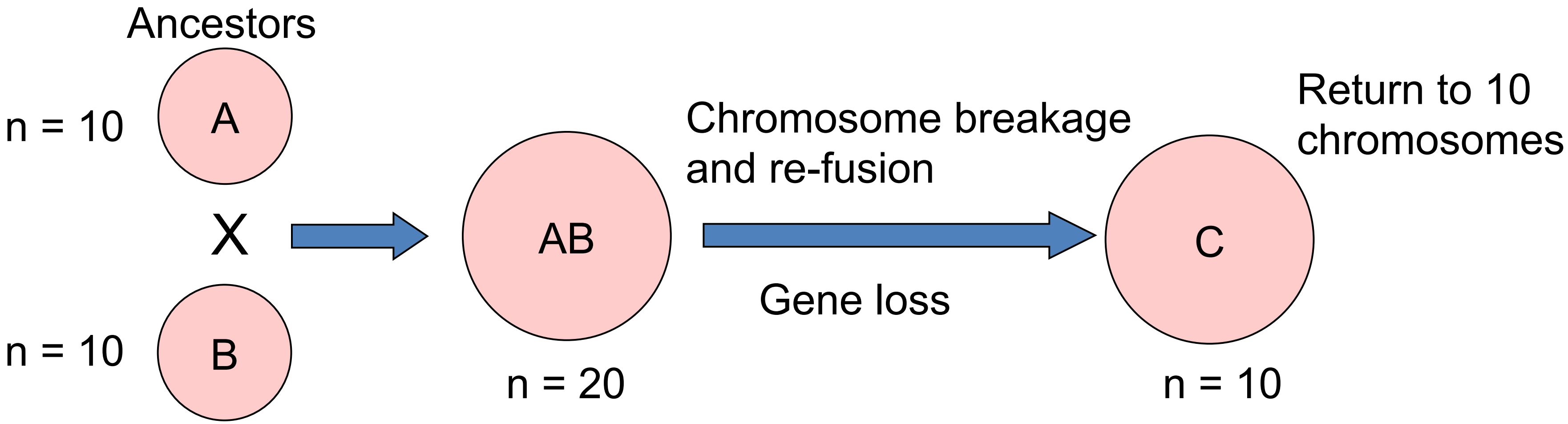
Letters or Single nucleotides polymorphisms (SNPs)

Gene content or parts list, known as copy number variations (CNVs)

Jumping genes, Transposable Elements (TEs) associated with the regulatory sequence between the Genes

Barbara McClintock
1983 Nobel Prize in Physiology & Medicine

Maize is a "Tale of Two Genomes"



Ancestors

n = 10 A

X

n = 10 B

AB

n = 20

Chromosome breakage and re-fusion

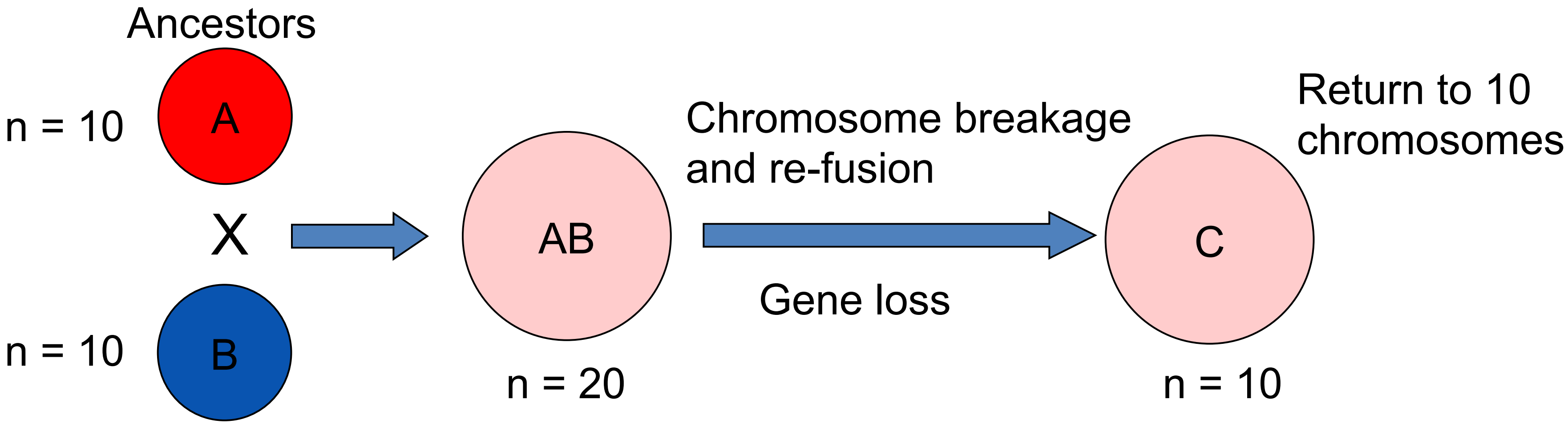
Gene loss

C

n = 10

Return to 10 chromosomes

Evolutionary History of the Maize Genome



Ancestors

n = 10 A

X

n = 10 B

AB

n = 20

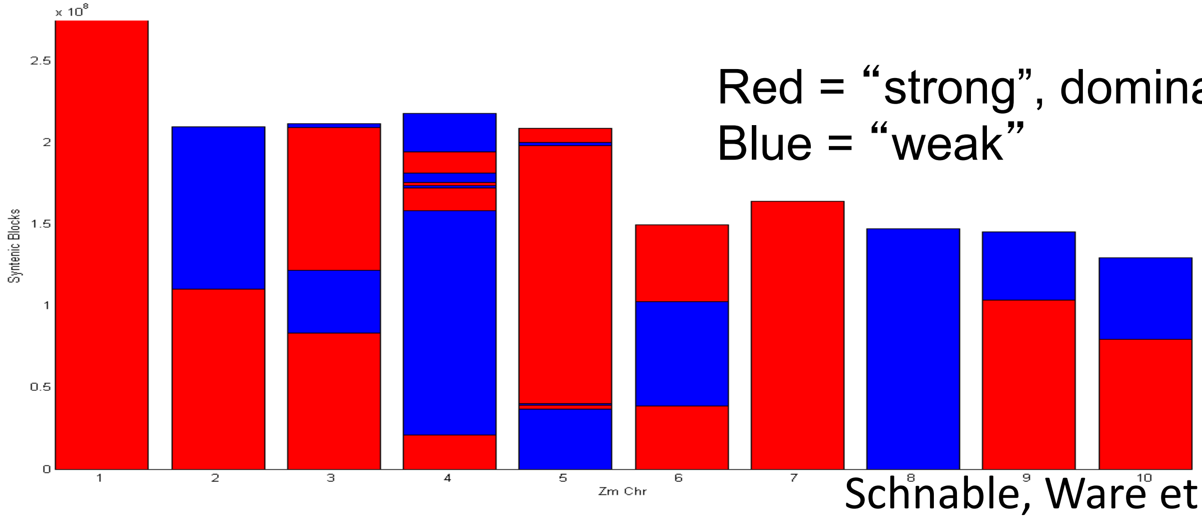
Chromosome breakage and re-fusion

Gene loss

Return to 10 chromosomes

C

n = 10



Red = “strong”, dominant genome
Blue = “weak”

Schnable, Ware et.al, Science 2009

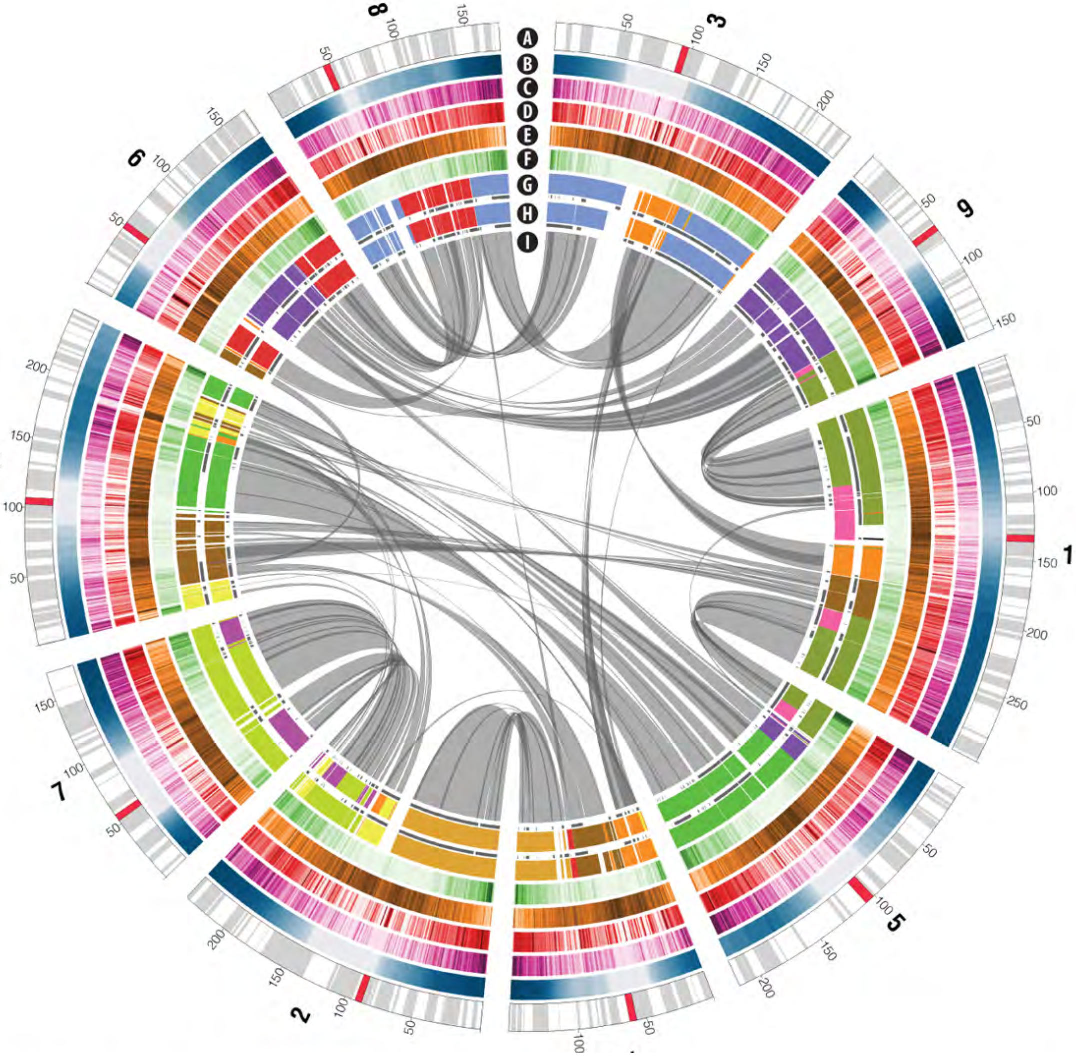
Maize is a “Tale of Two Genomes”

Maize, also known as corn experienced a whole genome duplications and then lost many of the genes

The genes that were kept by corn can tell us about how corn is adapting

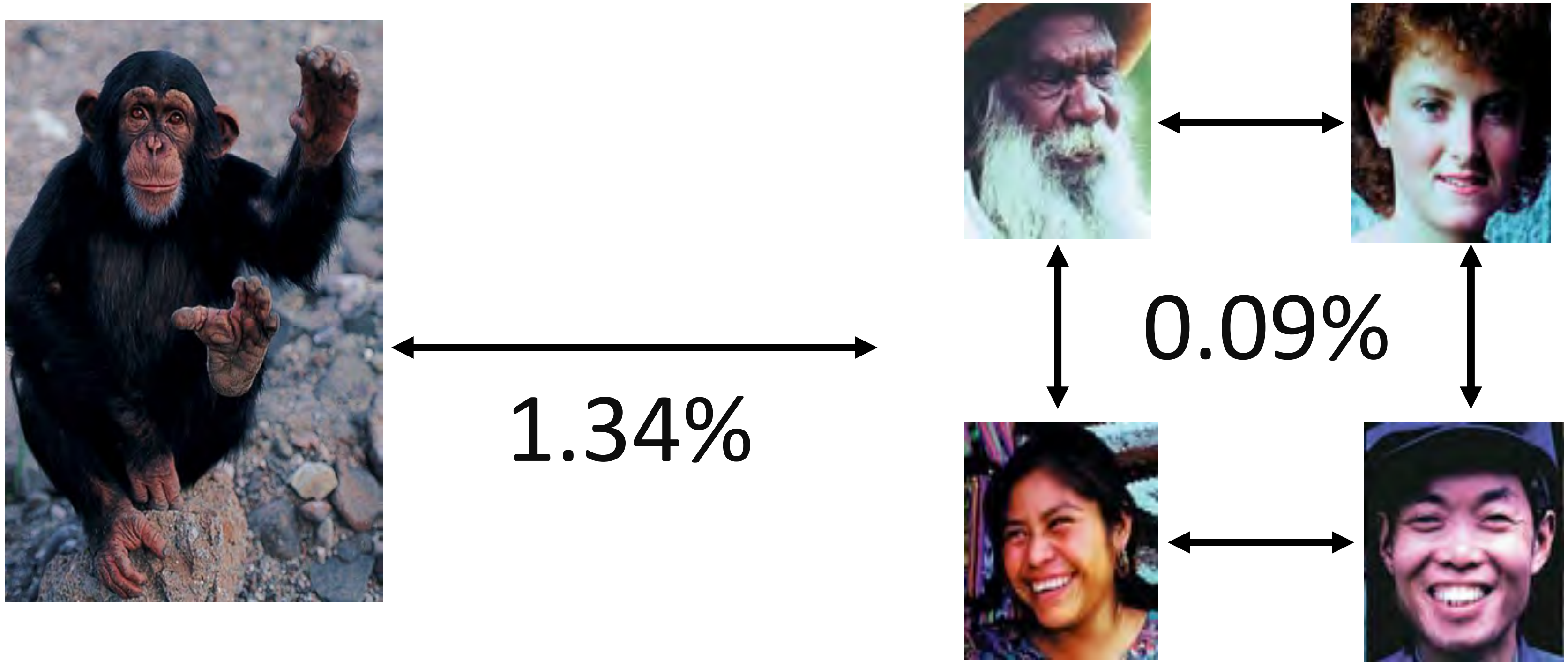
Transcription factors, kinases, chromatin modifiers

Not all genomes have the same potential



Schnable, Ware et.al, Science 2009

Humans Have Limited Molecular Diversity

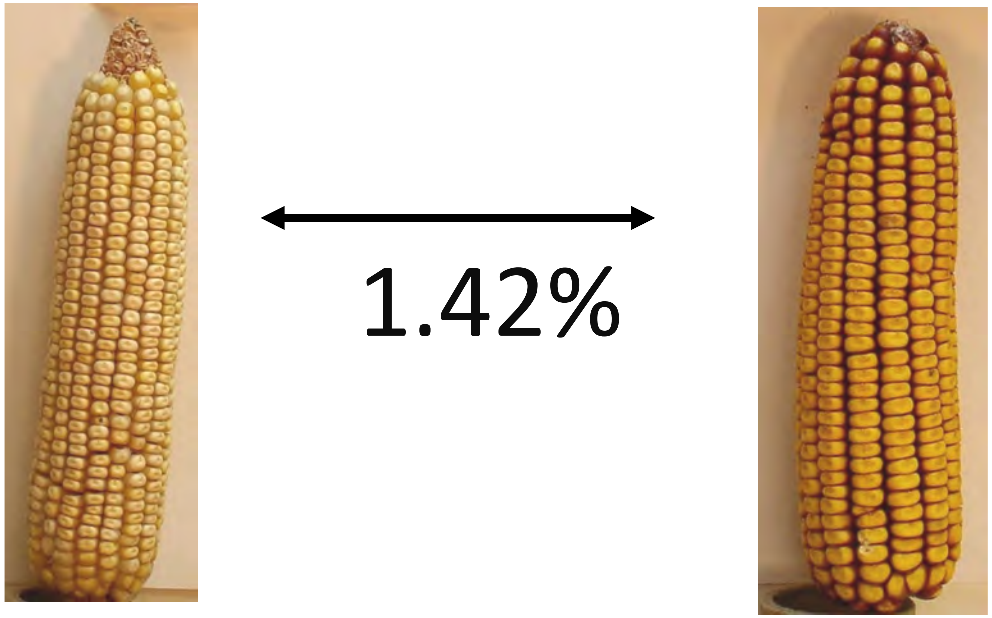


1.34%

0.09%

Silent Diversity; Zhao, et al. (2001) *PNAS*

Maize diversity is greater than the difference between human and chimps



1.42%

Silent Diversity; Tenallion, et al. (2001) *PNAS*

Individual Maize lines are very different from each other

The SNPs and gene differences affect how corn plants grow

Access to these sequences can accelerate the time it takes to make new lines of corn

Differences come from locally duplicated genes

Images Bo Wang

An additional copy of gene confers tolerance to acidic soil

Maron *et. al*, *PNAS*, 2013

Maize genomes are highly variable


nature genetics

VOLUME 44 NUMBER 7 JULY 2012
www.nature.com/naturegenetics

Putting genome variation to work in maize

Red beet pigmentation

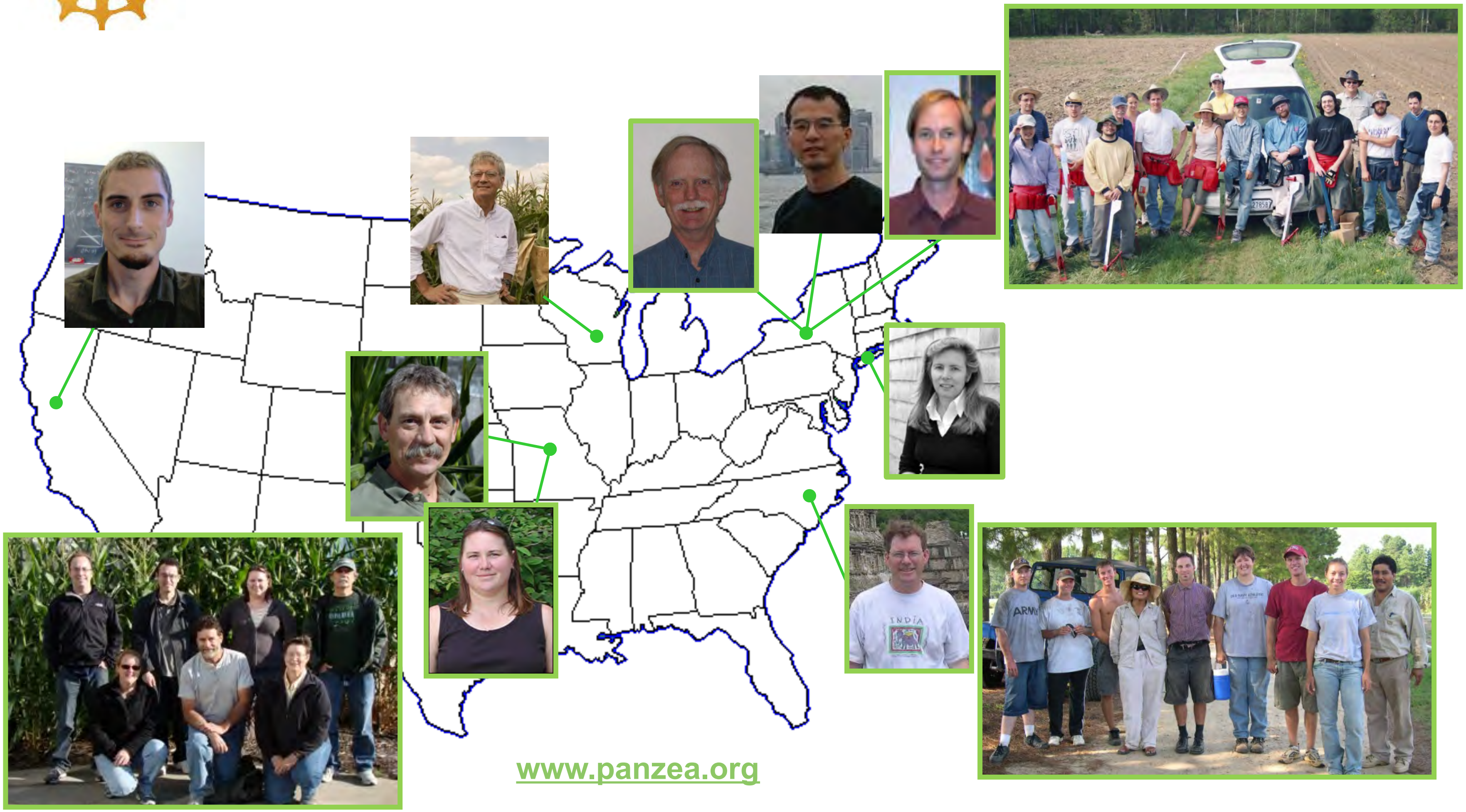
Classifying renal cancer

- High rate of SNP and **structure variation** in the population
- Structure variations are highly **associated with phenotypic variation**
- **Structural variation** in non coding region was enriched for phenotypic variation
- **One genome is not enough** to represent the diversity of the population

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC *et al*: **Maize HapMap2 identifies extant variation from a genome in flux**. *Nat Genet* 2012, **44**(7):803-807.

MAIZE DIVERSITY PROJECT

Meet the Family (2002 - present)

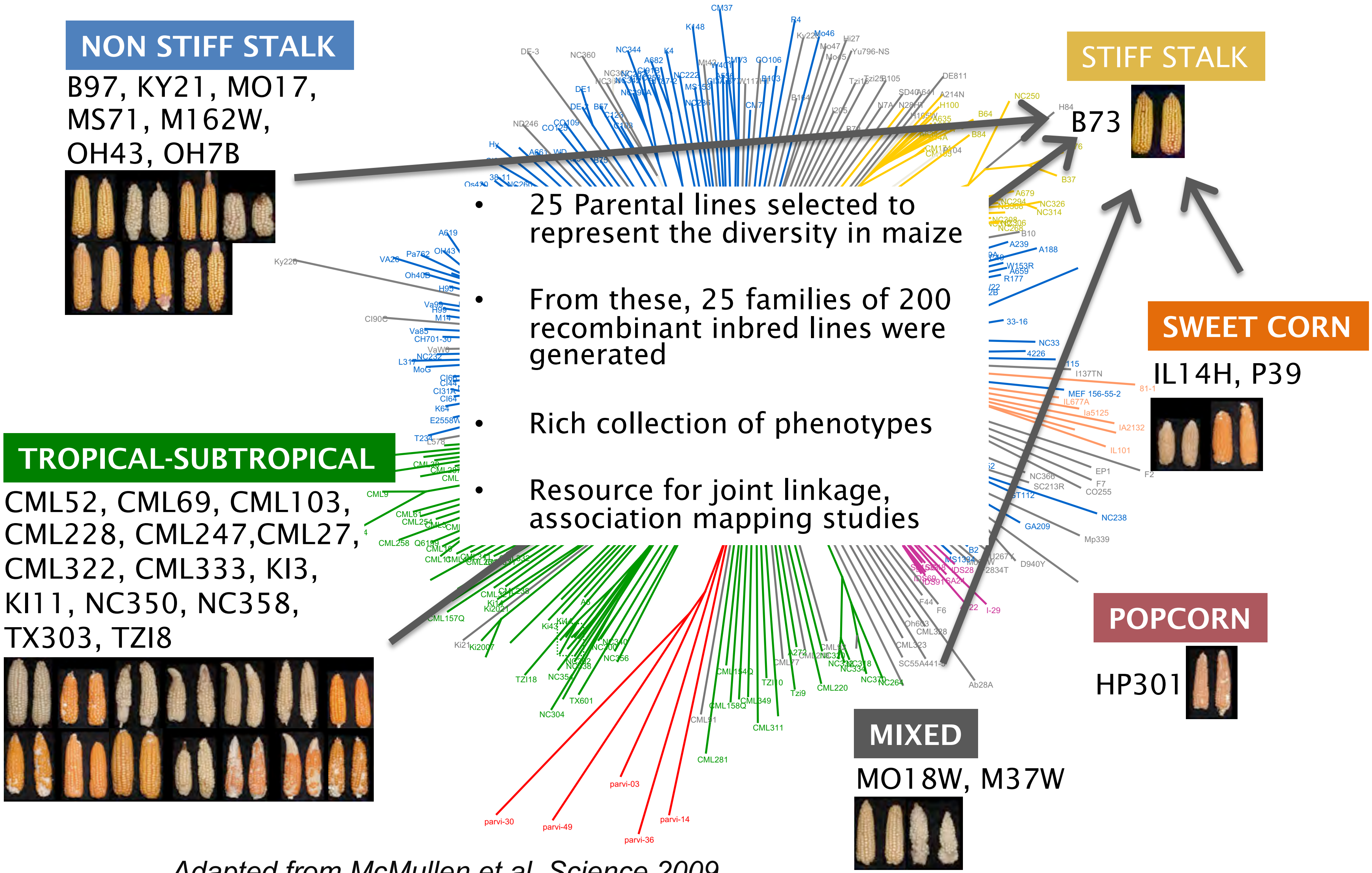


www.panzea.org

USDA
United States Department Of Agriculture
Agricultural Research Service

COURTESY SHERRY FLINT-GARCIA

26 Reference Assemblies for the Maize Nested Association Mapping (NAM) Population



NON STIFF STALK

B97, KY21, MO17, MS71, M162W, OH43, OH7B

STIFF STALK

B73

SWEET CORN

IL14H, P39

TROPICAL-SUBTROPICAL

CML52, CML69, CML103, CML228, CML247,CML27, CML322, CML333, KI3, KI11, NC350, NC358, TX303, TZI8

POPCORN

HP301

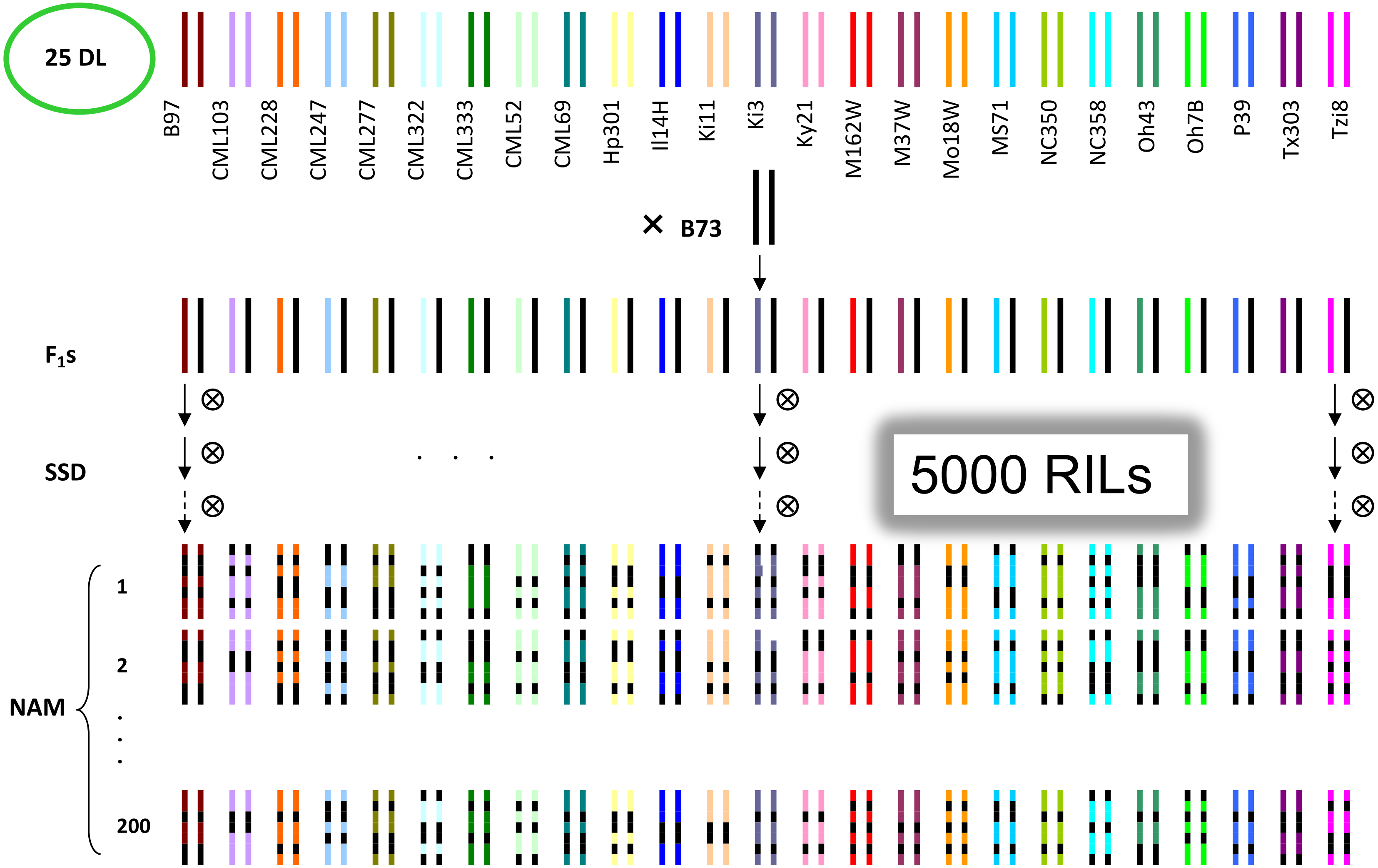
MIXED

MO18W, M37W

- 25 Parental lines selected to represent the diversity in maize
- From these, 25 families of 200 recombinant inbred lines were generated
- Rich collection of phenotypes
- Resource for joint linkage, association mapping studies

Adapted from McMullen et al. Science 2009

Maize Nested Association Mapping (NAM) Population

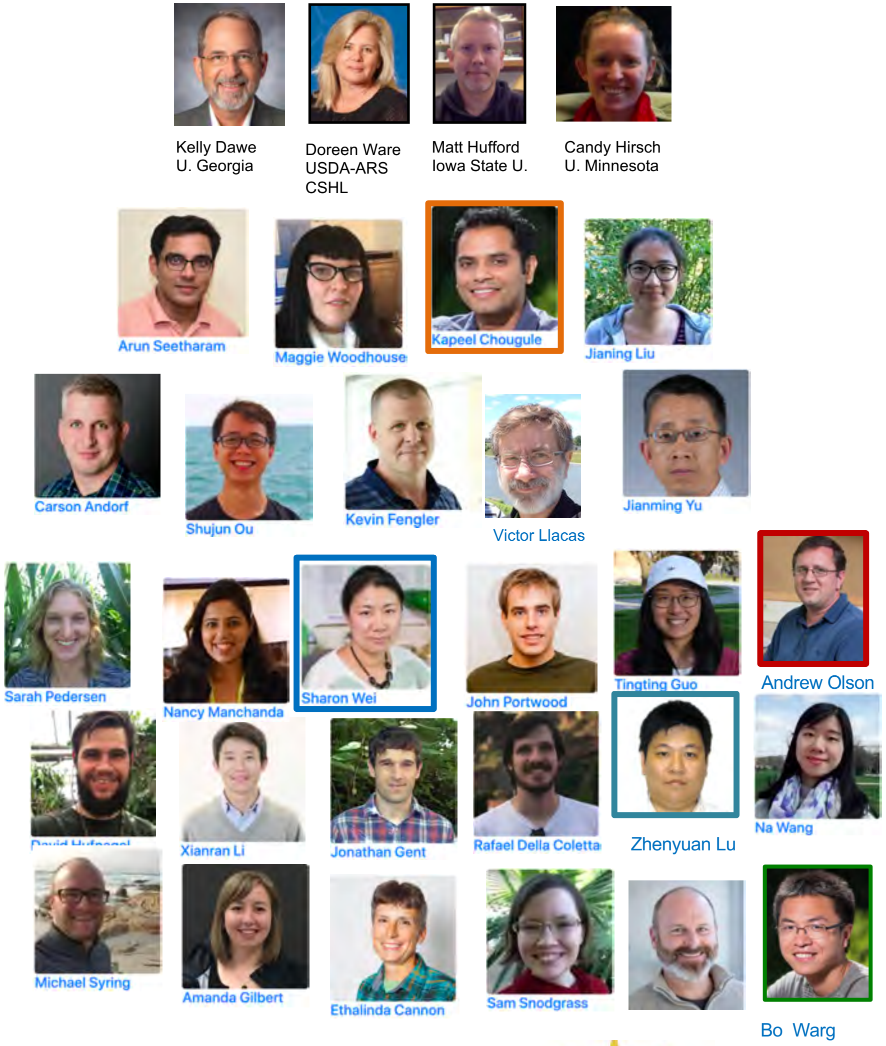


5000 RILs

Yu, et al. (2008) *Genetics*

26 Maize NAM founders reference assemblies (2019- 2021)

Arun Seetharam -ISU
Margaret Woodhouse – MaizeGDB
Kapeel Chougule -CSHL
Shujun Ou -ISU
Jianing Liu -UGA
Xuehong Wei –CSHL
Zhenyuan Lu –CSHL
Andrew Olson –CSHL
Bo Wang -CSHL
Sharon Wei -CSHL
TingTing Guo -ISU
Rafael Della Coletta -UM
Xianran Li -ISU
John Portwood –MaizeGDB
Kevin Fengler -Corteva
Victor Llaca -Corteva
Amanda Gilbert -UM
Nancy Manchanda -ISU
Samantha Snodgrass -ISU
David Hufnagel -ISU
Sarah Pedersen -ISU
Michael Syring –ISU
Ethy Cannon - MaizeGDB
Carson Andorf -MaizeGDB
Jonathan Gent –UGA
Todd Michael - JCVI
Jianming Yu –ISU
Candice Hirsch -UM
Doreen Ware –CSHL
Matthew B. Hufford -ISU
R. Kelly Dawe -UGA



USDA United States Department Of Agriculture
Agricultural Research Service

NSF

CORTEVA™
agriscience

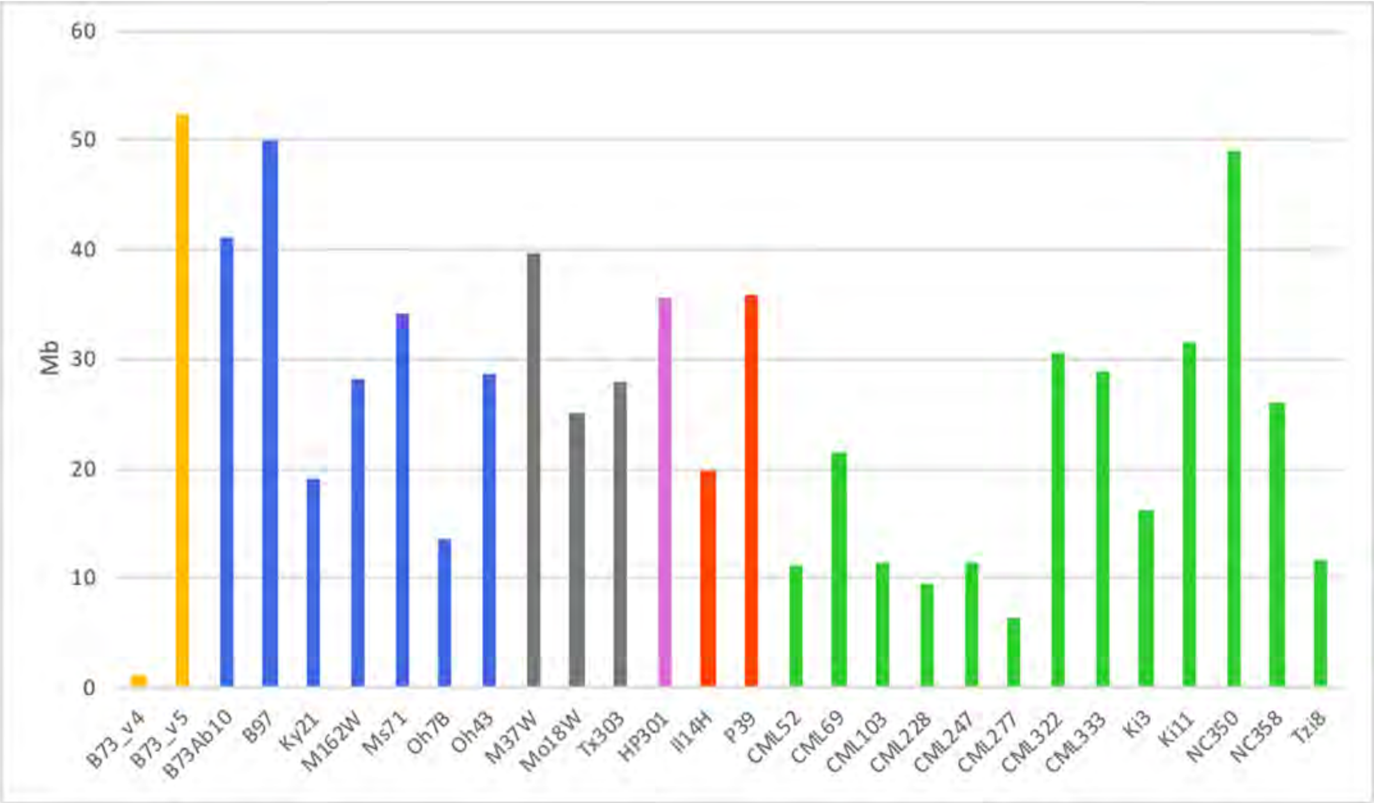
New assemblies have a vast improvement in the contiguity of the sequence


Kevin
Fengler


Victor Llaca

CORTEVA™
agriscience

Matrices	Assembled Contigs
Total Bases in Assembly	**2,180,413,054**
Contig Contiguity (NG50)	**52,409,415**
Number of Contigs	**811**
Longest Contig	**161,290,055**

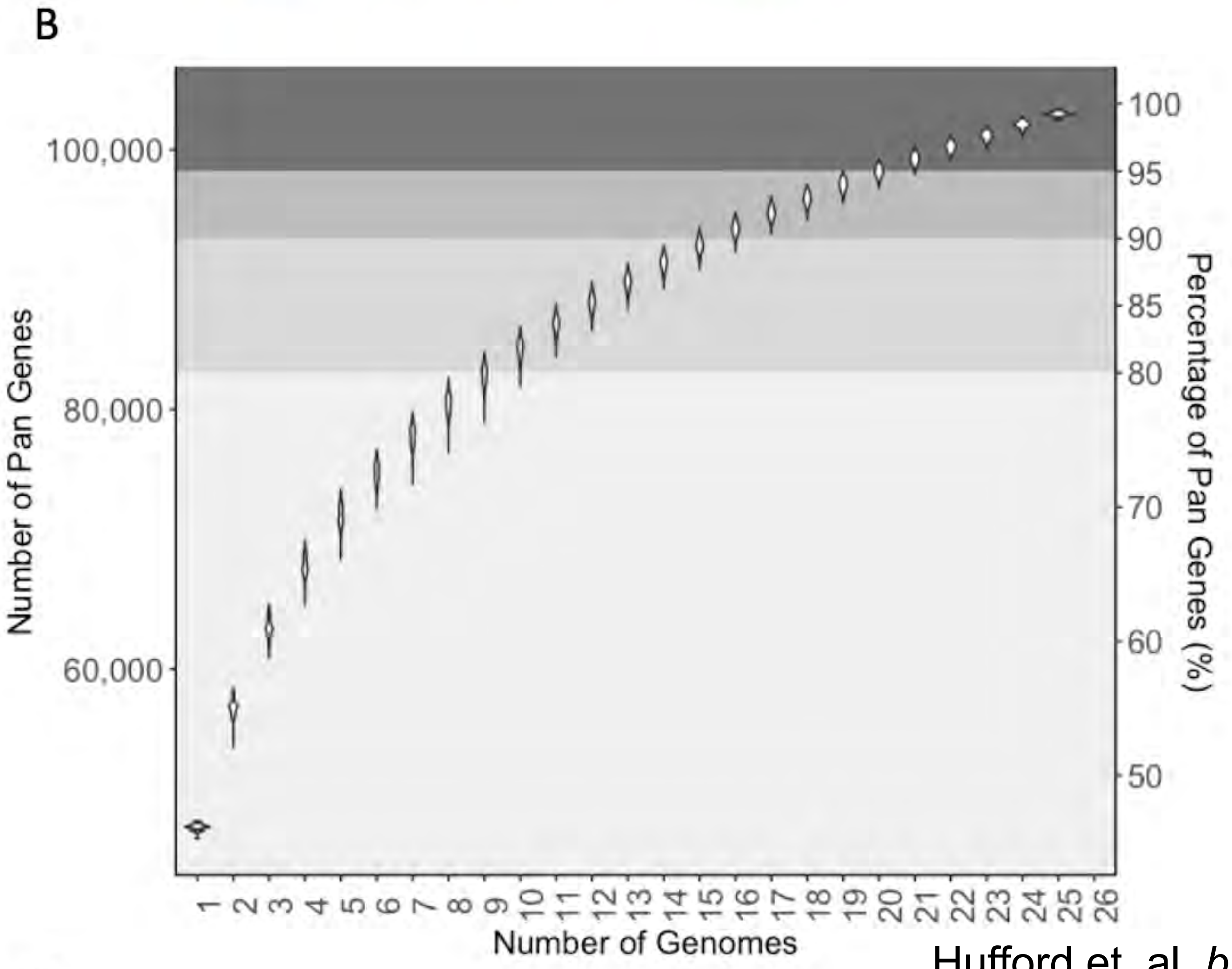


🟨 StiffStalk
🟦 NonStiffStalk
⬛ Mixed
🟪 PopCorn
🟥 SweetCorn
🟩 Tropical

N50 is the shortest contig length needed to cover 50% of the **genome**. -> Half of the **genome sequence** is covered by contigs larger than or equal the **N50** contig size.

Hufford et, al, *bioRxiv*, 2021

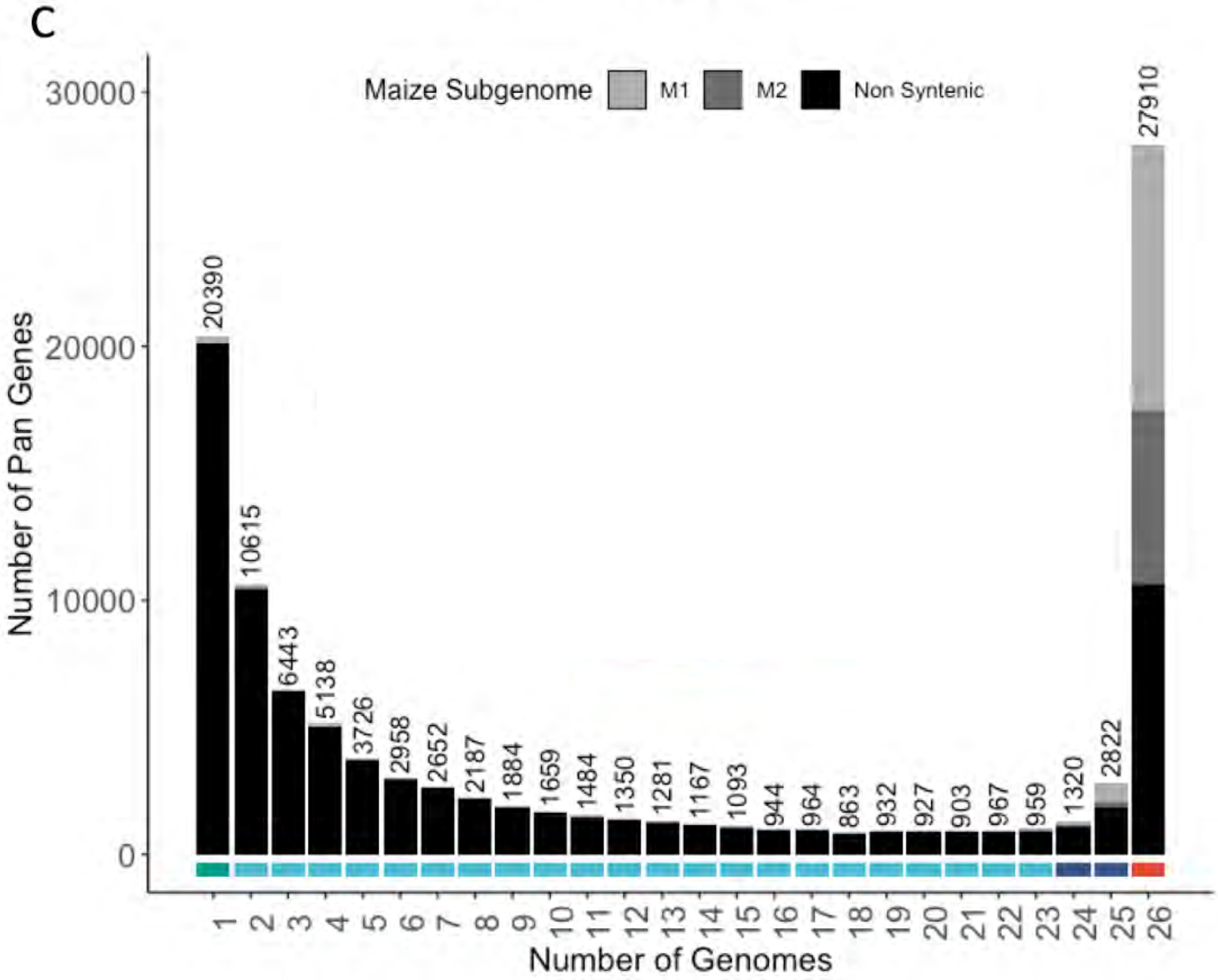
More than 100, 000 genes found in the 26 maize accession



Hufford et, al, *bioRxiv*, 2021

Maize Pan Gene Set

Candy Hirsch
U. Minnesota

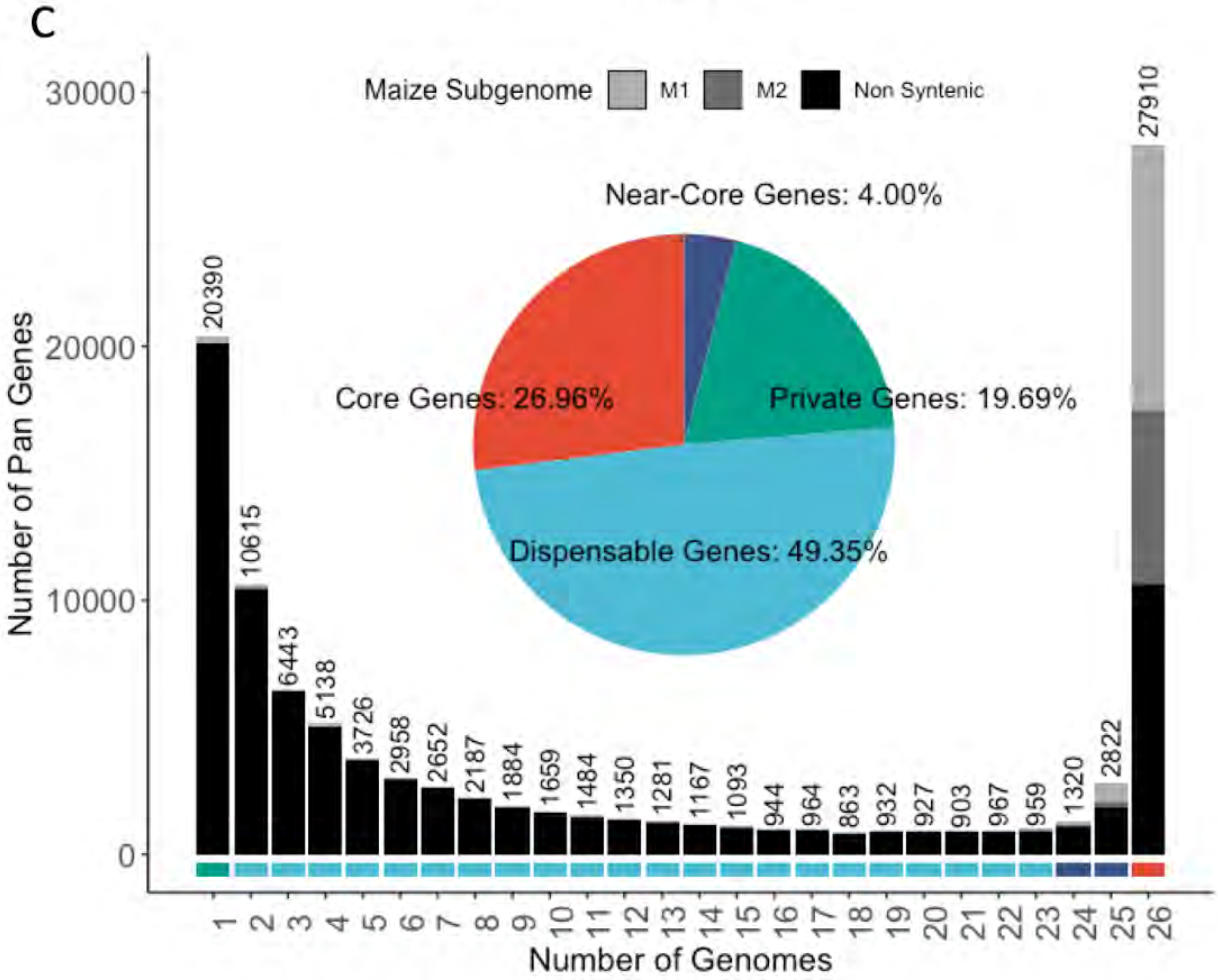


Hufford et, al, *bioRxiv*, 2021

Maize Pan Gene Set

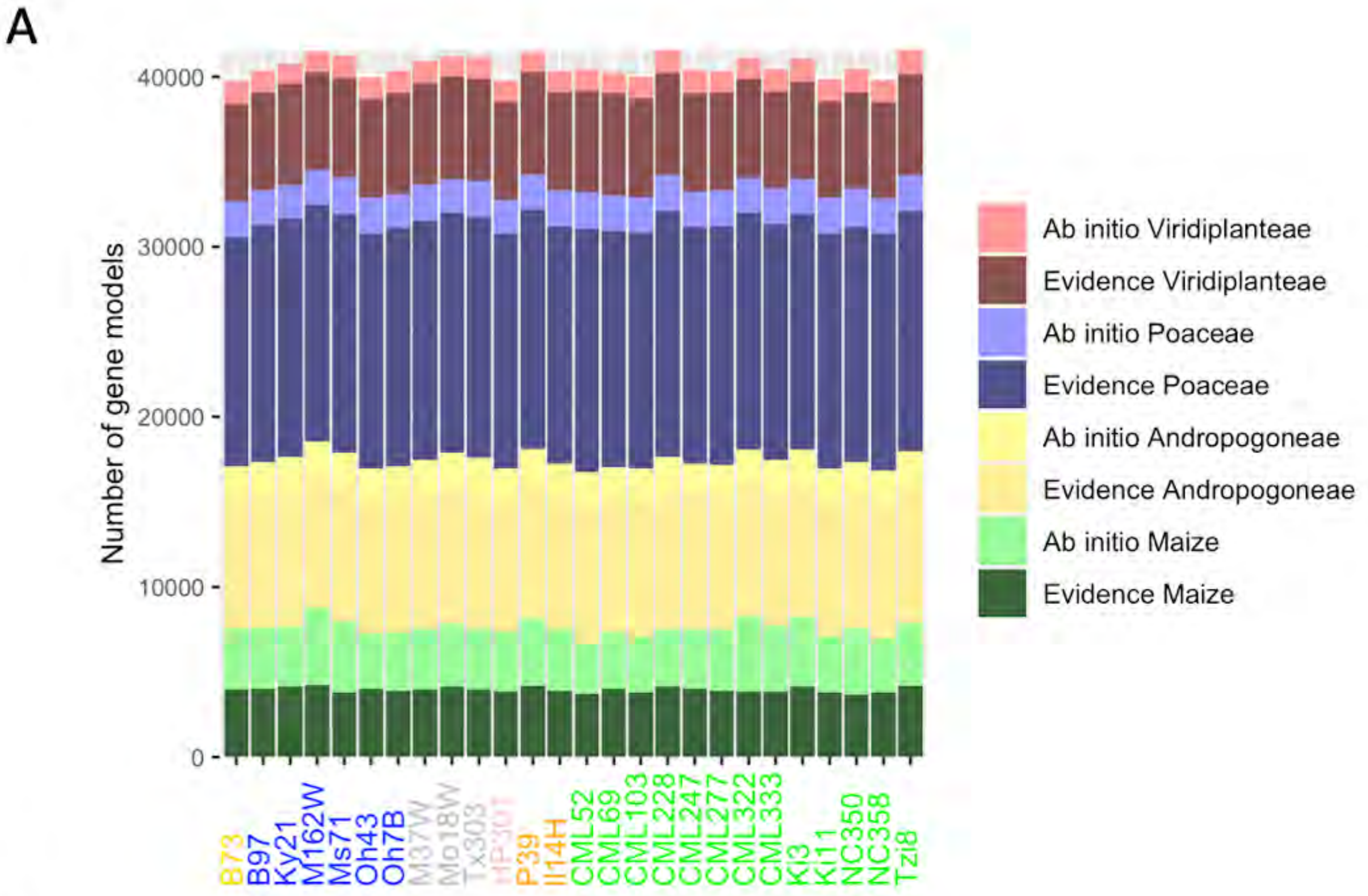
Core genes are more likely to be Syntenic

Candy Hirsch
U. Minnesota



Hufford et, al, *bioRxiv*, 2021

The bulk of the genes in maize are found in other species



Hufford et, al, *bioRxiv*, 2021

Genome contains life history of the species

Teosinte
Zea mays L. ssp. *parviglumis*

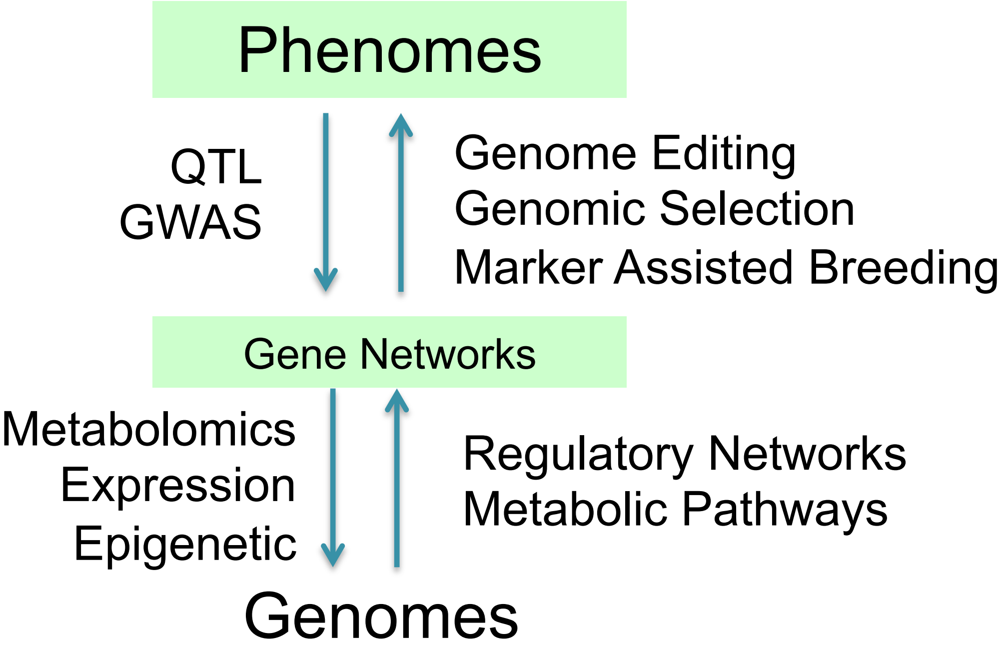
Landraces

Maize, Modern Inbreds
Zea mays L. ssp. *mays*

Biology Enabled Agriculture

G X (environment +management) = P

Complex Traits: Yield & Quality



Phenomes

QTL
GWAS

Genome Editing
Genomic Selection
Marker Assisted Breeding

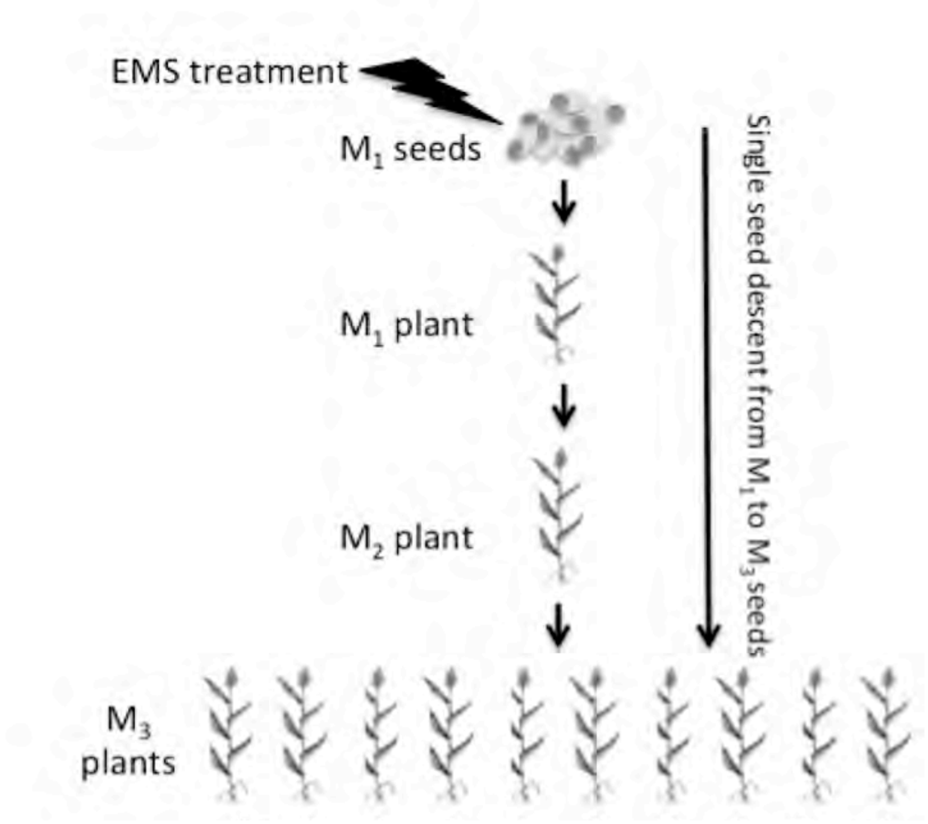
Gene Networks

Metabolomics
Expression
Epigenetic

Regulatory Networks
Metabolic Pathways

Genomes

Profiling new sorghum genetic & phenotypic variation



EMS Mutagenesis
 -Random
- Single nucleotide change
- >99% GC→AT

Zhanguo Xin

Cropping Systems Research Lab,
USDA-ARS, Lubbock TX

- Parental line: BTX 623
- >10,000 individual M2 seed pools
- >6,400 M3 seeds obtained and
- Phenotyping is on-going and need to be expanded
- High quality DNA prepared for all lines

Jiao *et al*. The Plant Cell, 2016

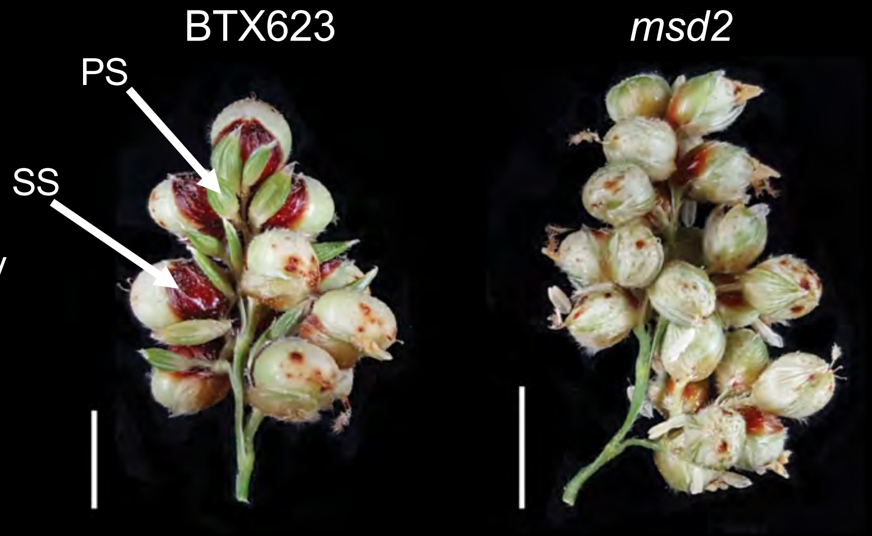
Mutation Detection by whole genome sequencing of 256 mutants for forward genetics

Yinping Jiao

- **Sequencing summary**
  - 20 M3 plants pooled together for sequence to averagely 16X
  - Average whole genome coverage – 86%; gene space coverage -95%
- **Quality control of the population:** 2 contamination lines + 2 sibling lines
- **Mutation detection:**
  - 1,862,560 EMS-induced mutations
  - Sanger sequencing validation rate >98%
  - 7,660 mutation/mutant = 1,798 homozygous + 5,862 heterozygous
- **Mutation Effect:**
  - 22% of mutations are located in genes, covering 95% of Sorghum genes
  - 57% (18,684) of the genes harbor >35,000 disruptive mutations, ~2 disruptive mutations per gene.

Jiao *et al*. The Plant Cell, 2016

Multiseeded (msd1, msd2, msd3)



~50% Fertility

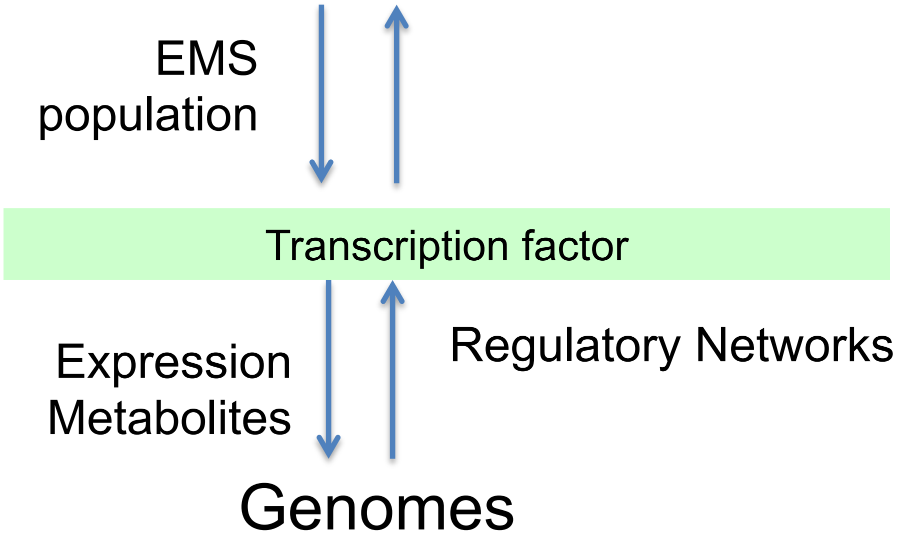
100% Fertility

Collaborator Zhanguo Xin
USDA ARS, Lubbock

Biology Enabled Agriculture

G X (environment +management) = P

Yield > Flower development > grain number > fertility/branching



EMS
population

Transcription factor

Expression
Metabolites

Regulatory Networks

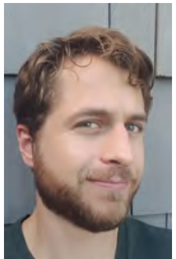
Genomes

MSD1

TCP transcription factor

Yinping Jiao

Nick Gladman



oung Koung Lee

Zhanguo Xin

Jiao et al., Nature Comm. (2018)

Biology Enabled Agriculture

G X (environment +management) = P

Yield > Flower development > grain number > fertility/branching

EMS
population

Transcription factor

Expression
Metabolites

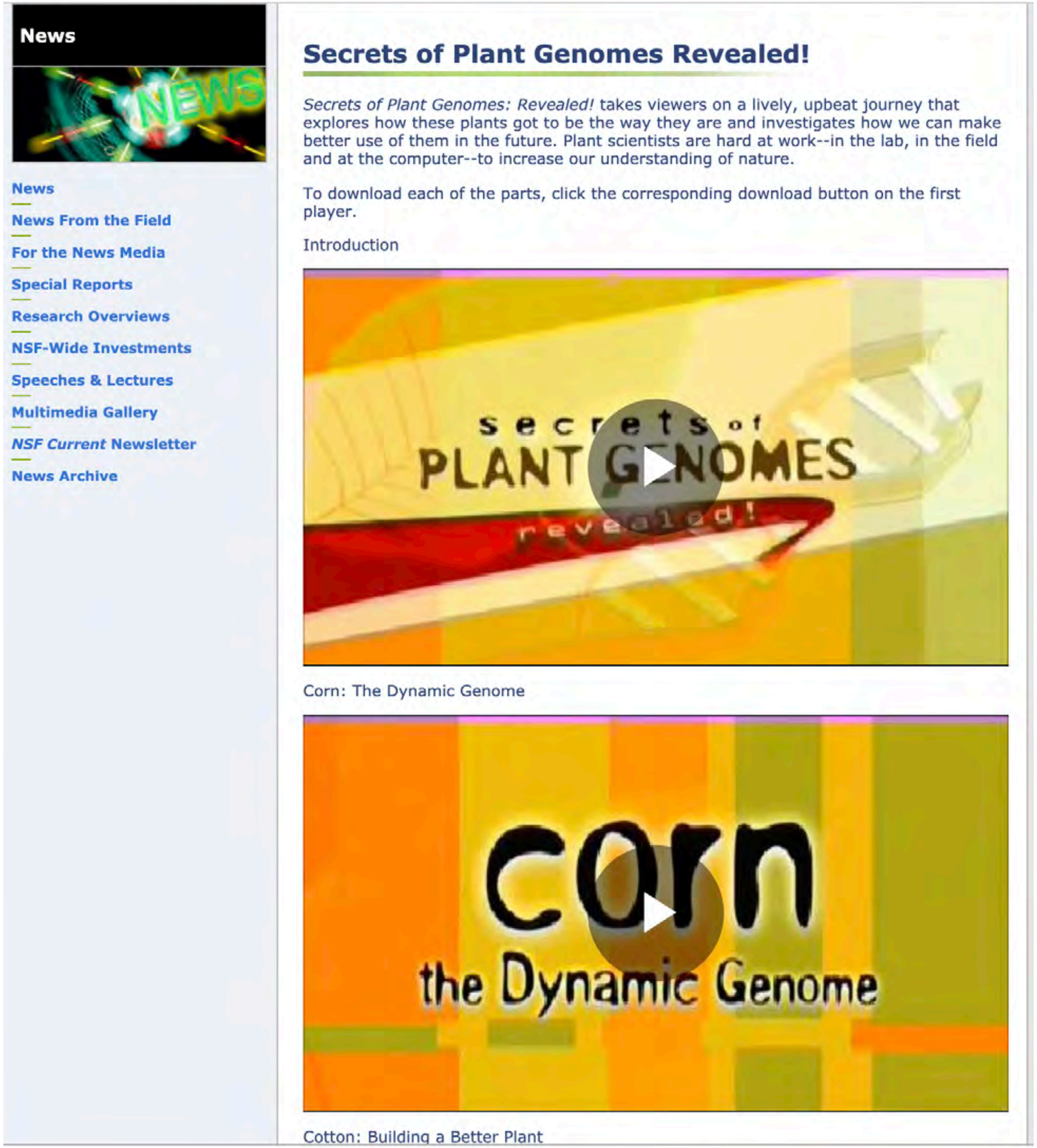
Regulatory Networks

Genomes

TCP transcription factor

Teosinte Branched

Doebley et al., Geneitics (1995)



Secrets of Plant Genomes Revealed!

Secrets of Plant Genomes: Revealed! takes viewers on a lively, upbeat journey that explores how these plants got to be the way they are and investigates how we can make better use of them in the future. Plant scientists are hard at work--in the lab, in the field and at the computer--to increase our understanding of nature.

To download each of the parts, click the corresponding download button on the first player.

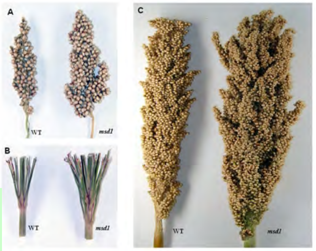
Introduction

Corn: The Dynamic Genome

Cotton: Building a Better Plant

Biology Enabled Agriculture

G X (environment +management) = P



Yield > Flower development > grain number > fertility/branching

EMS population

Genome Editing
Marker Assisted Breeding

Transcription factor, Plant hormones

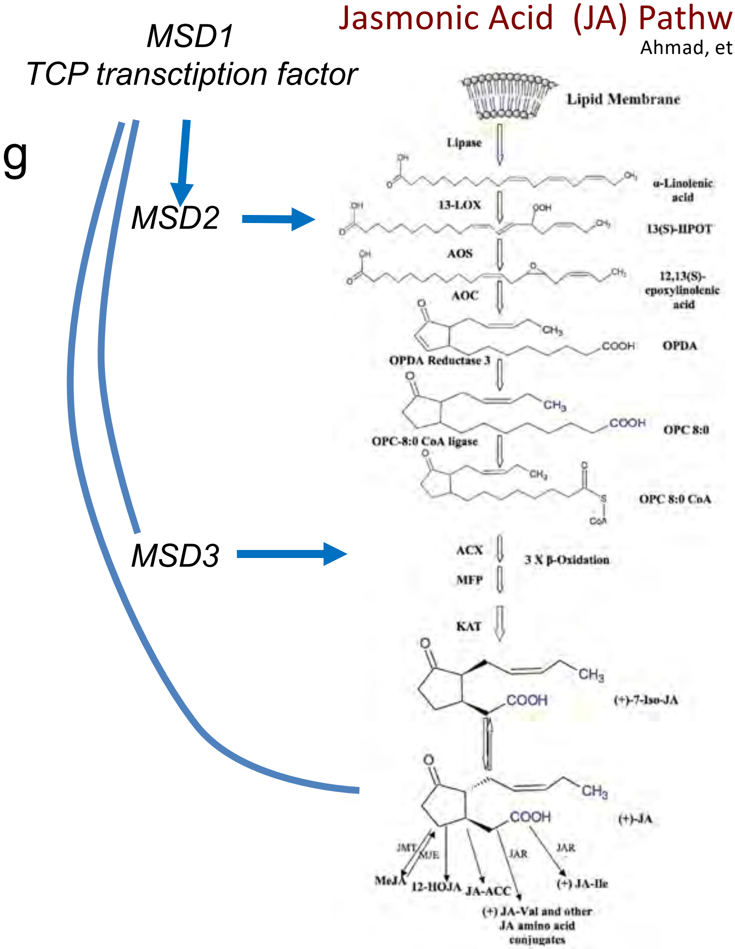
Expression
Metabolites

Regulatory Networks
Metabolic Pathways

Genomes

Jasmonic Acid (JA) Pathway

Ahmad, et al. 2016. *Frontiers*



Jiao et al., Nature Comm. (2018)
Gladman et al., Int. J. Mol. Science (2019)
Dampanaboina, L., *Int. J. Mol. Science (2019)*

Nitrogen, soil, and agricultural sustainability

Insufficient Nitrogen fertilizer

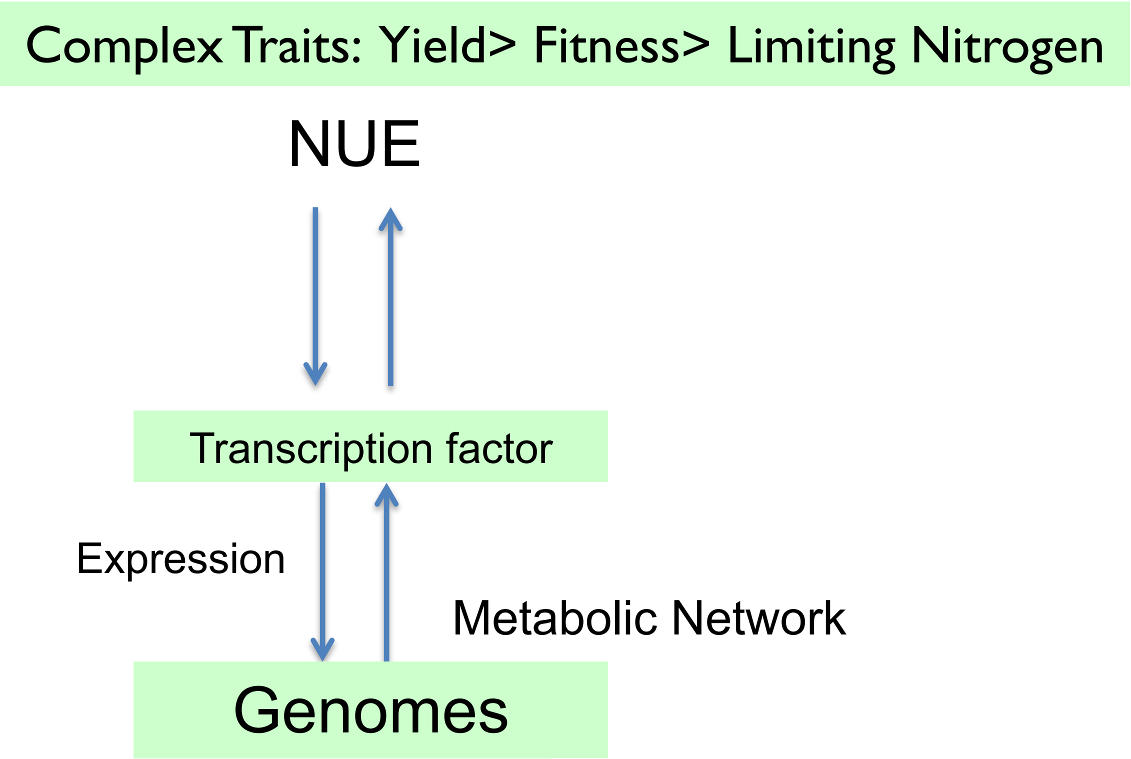
Nitrogen deficiencies, LSU (courtesy, Dr. Brenda Tubana)

Excess Nitrogen fertilizer

October 2011. Credit, USGS, NASA.

Biology Enabled Agriculture

G X (environment +management) = P



Complex Traits: Yield> Fitness> Limiting Nitrogen

NUE

Transcription factor

Expression

Metabolic Network

Genomes

Lifang Zhang

Andrew Olson

Christophe
Liseron-Monfils

Allie Gaudinier
UC DAVIS

Siobhan Brady
UC DAVIS

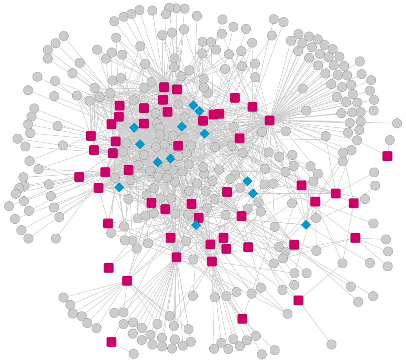
Guardinier *et al*, *Nature 2017*

Biology Enabled Agriculture

G X (environment +management) = P

Complex Traits: Yield> Fitness> Limiting Nitrogen

NUE



Transcription factor

Expression
Gene regulation

Regulatory Networks
Metabolic Network

Genomes


Lifang Zhang


Andrew Olson


Christophe
Liseron-Monfils


Allie Gaudinier
UC DAVIS


Siobhan Brady
UC DAVIS

UCDAVIS
UNIVERSITY OF CALIFORNIA

DUPONT
PIONEER.

Guardinier *et al*, *Nature 2017*

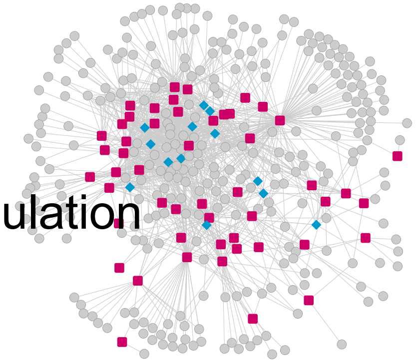
Biology Enabled Agriculture

G X (environment +management) = P

Complex Traits: Yield> Fitness> Limiting Nitrogen

NUE

tDNA population



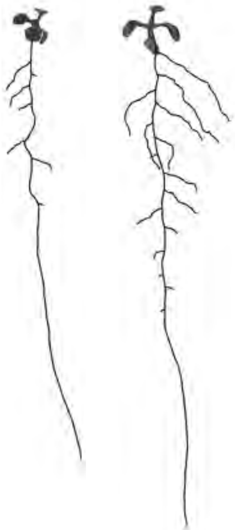
Transcription factor

Expression
Gene regulation

Regulatory Networks
Metabolic Network

Genomes

Sufficient N
10mM NO_3



Limiting N
1mM NO_3

Lifang Zhang

Andrew Olson

Christophe
Liseron-Monfils

Allie Gaudinier
UC DAVIS

Siobhan Brady
UC DAVIS

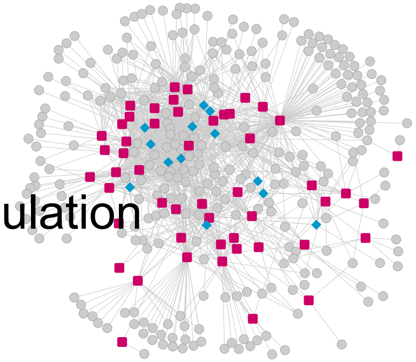
Guardinier *et al*, *Nature 2017*

Biology Enabled Agriculture

G X (environment +management) = P

Complex Traits: Yield> Fitness> Limiting Nitrogen

NUE



tDNA population

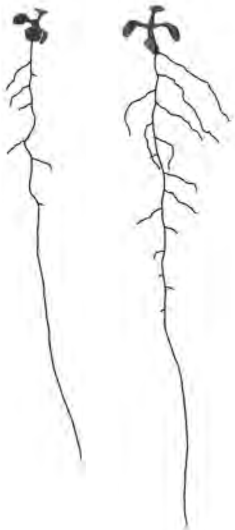
Transcription factor

Expression
Gene regulation

Regulatory Networks
Metabolic Network

Genomes

Sufficient N
10mM NO_3



Limiting N
1mM NO_3

Lifang Zhang

Andrew Olson

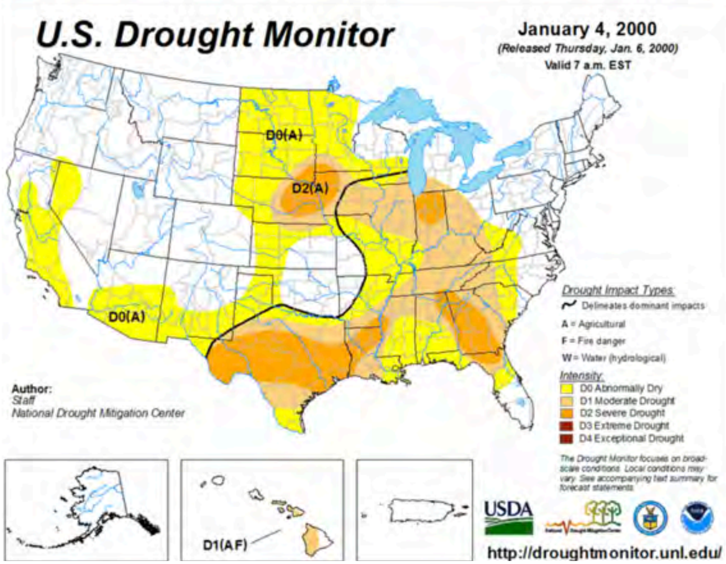
Christophe
Liseron-Monfils

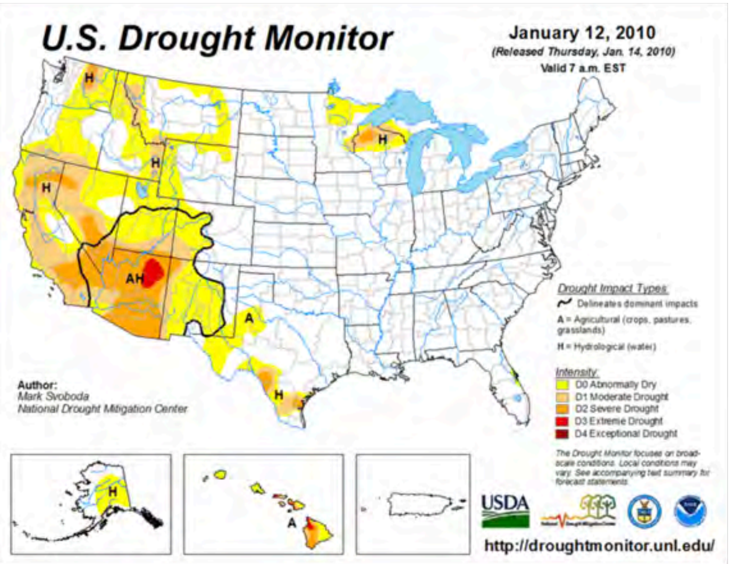
Allie Gaudinier
UC DAVIS

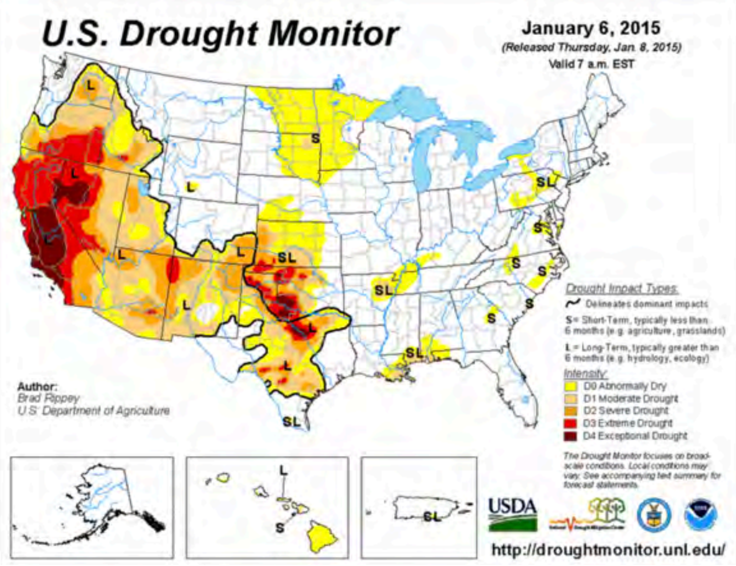
Siobhan Brady
UC DAVIS

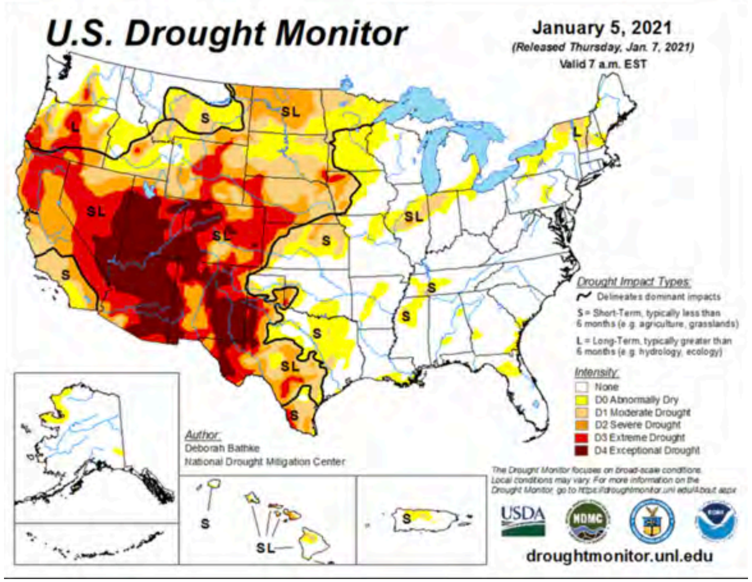
Guardinier *et al*, *Nature 2017*

Changing climate increasing temperature and drought





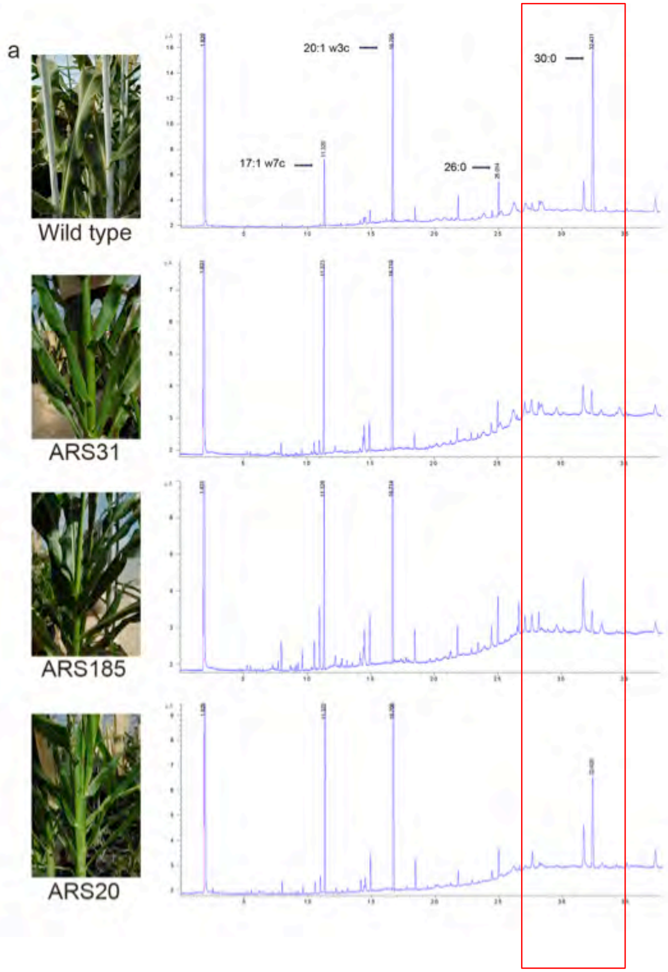




Reverse Genetics: From Gene to Phenotype

Epicuticular wax (bloom) of sorghum plays important roles in tolerance of environmental stress.

Gene in Arabidopsis	Sorghum gene	Amino acid change	Mutant Id
CER6	Sobic.001G453200	E159K	ARS20
KCS12	Sobic.004G341300	R189C	ARS73
CER5	Sobic.009G083300	P581L	ARS73
CER5	Sobic.009G083300	L244F	ARS20
CER6	Sobic.006G020600	A133T	ARS205
KCS7	Sobic.002G268300	P449S	ARS31
KCS4	Sobic.002G268500	A49V	ARS31
KCS20	Sobic.005G168700	R303Q	ARS185
CER1	Sobic.001G222700	L100F	ARS185

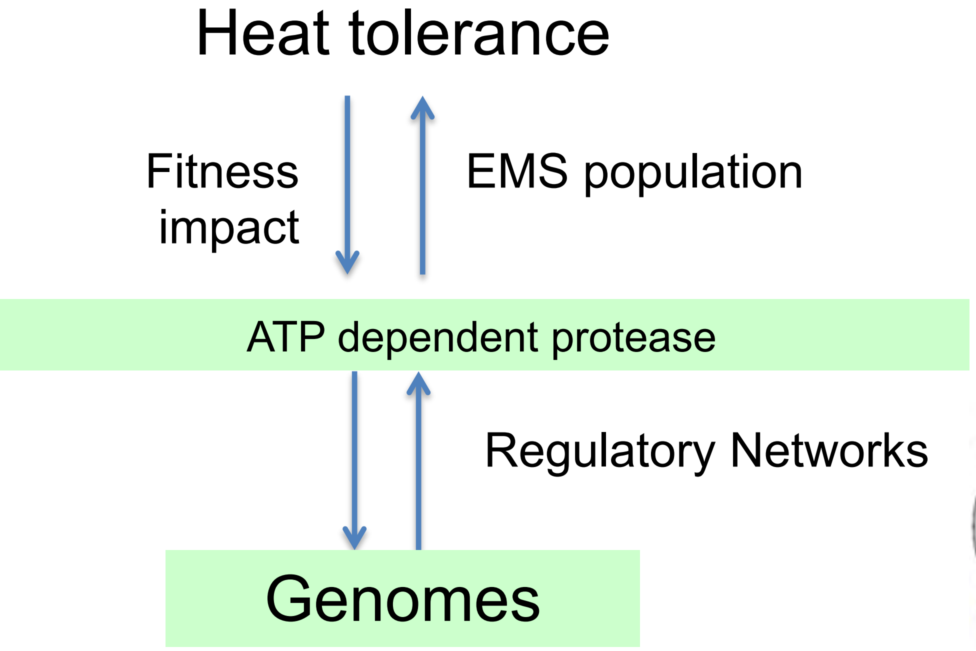


Jiao *et al*. The Plant Cell, 2016

Biology Enabled Agriculture

G X (environment +management) = P

Complex Traits: Yeild > Stress > Heat tolerance



Heat tolerance

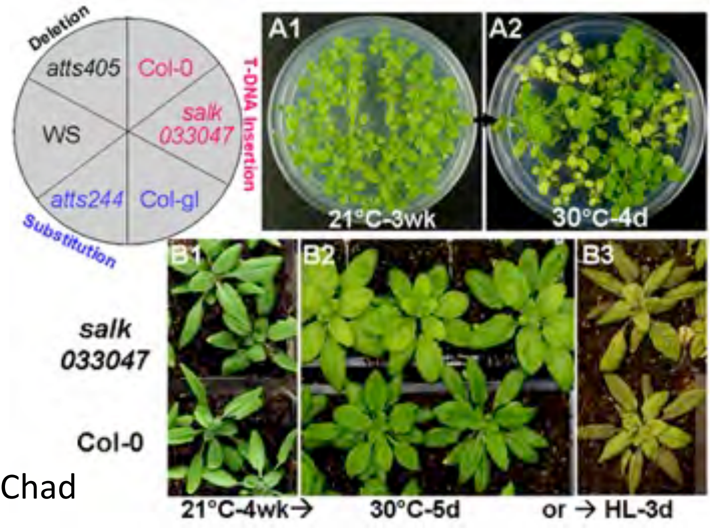
Fitness impact

EMS population

ATP dependent protease

Regulatory Networks

Genomes



Ftsh11 identified in a model plant

Yinping Jiao

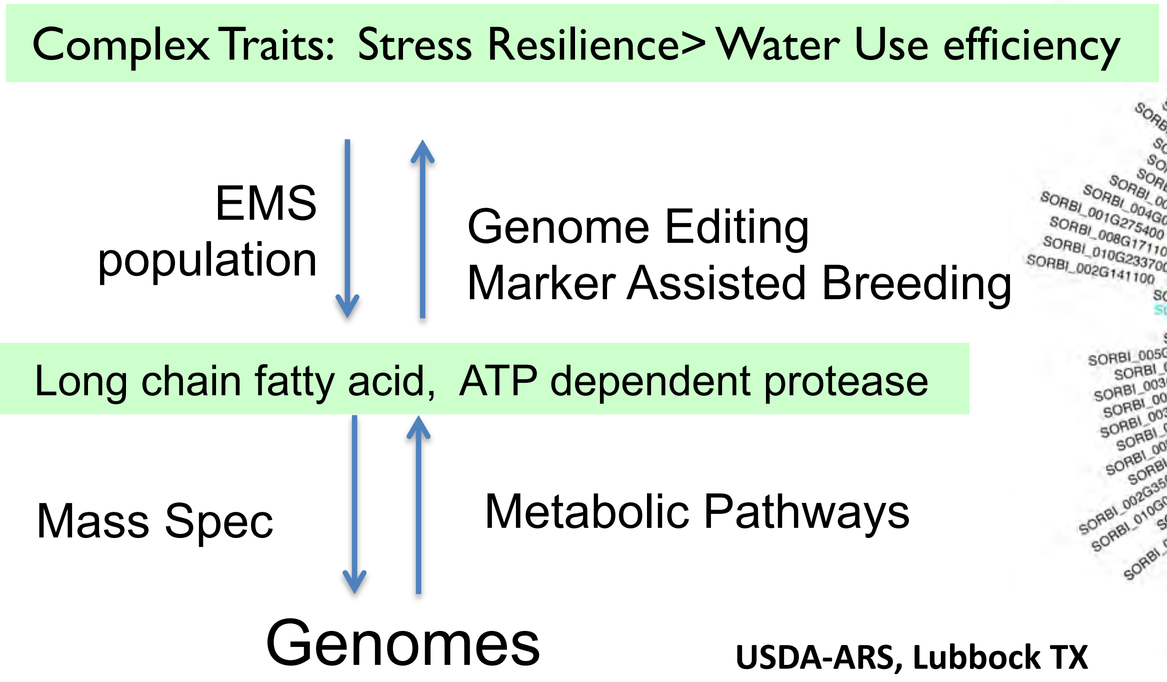
Zhanguo Xin

USDA-ARS, Lubbock TX
Zhanguo Xin, Gloria Burow, Ratan Chopra, John Burke, Chad Hayes

Jiao et al. Plant Cell 2016

Biology Enabled Agriculture

G X (environment +management) = P



Complex Traits: Stress Resilience> Water Use efficiency

EMS population

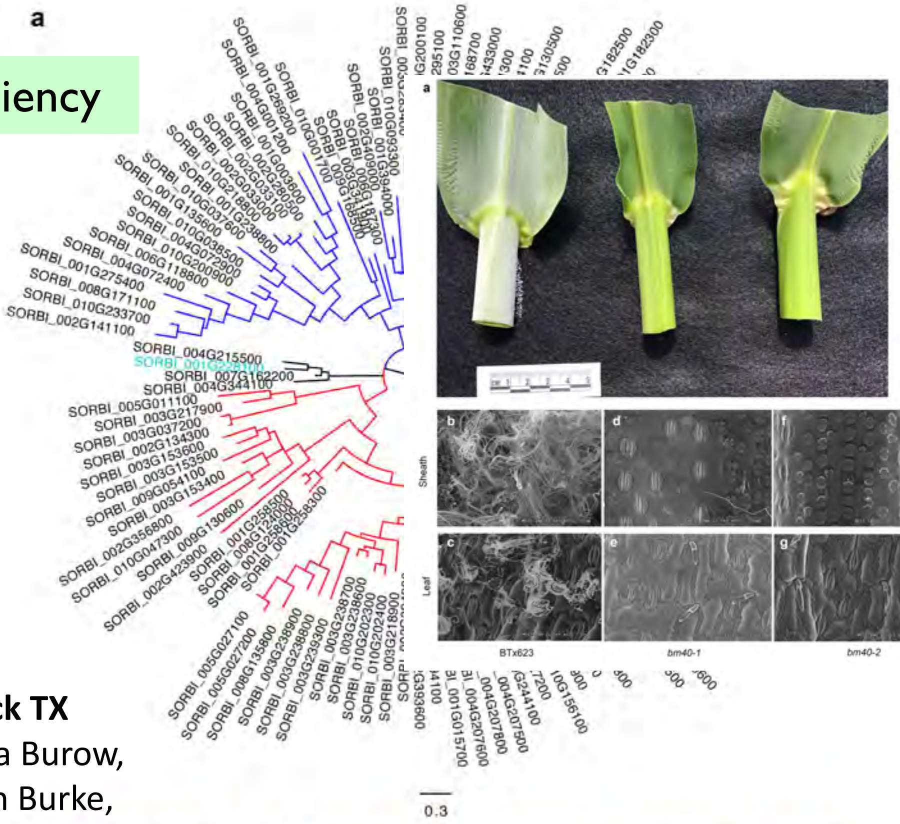
Genome Editing
Marker Assisted Breeding

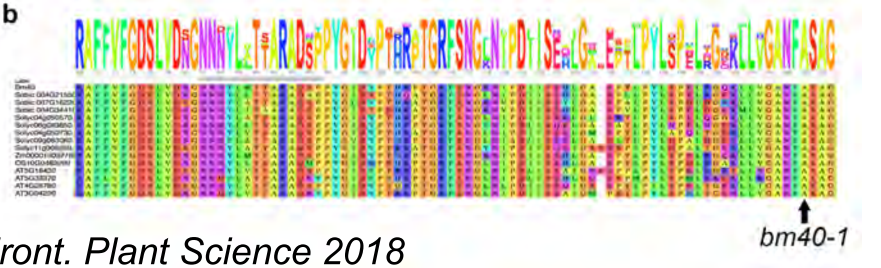
Long chain fatty acid, ATP dependent protease

Mass Spec

Metabolic Pathways

Genomes



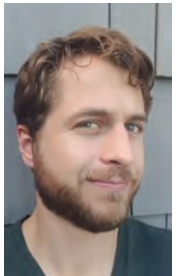


USDA-ARS, Lubbock TX
Zhanguo Xin, Gloria Burow,
Ratan Chopra, John Burke,
Chad Hayes

Jiao et al. Front. Plant Science 2018

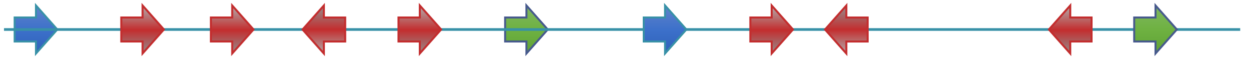
Yinping Jiao

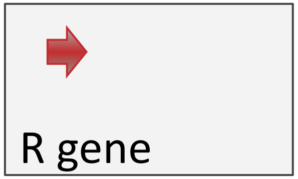
Nick Gladman



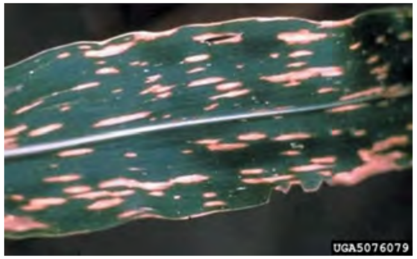
Zhanguo Xin

Climate change impacts disease pressures





Sorghum Leaf spot
http://texassorghum.org/wp-content/uploads/2015/10/Fig.-2.jpg



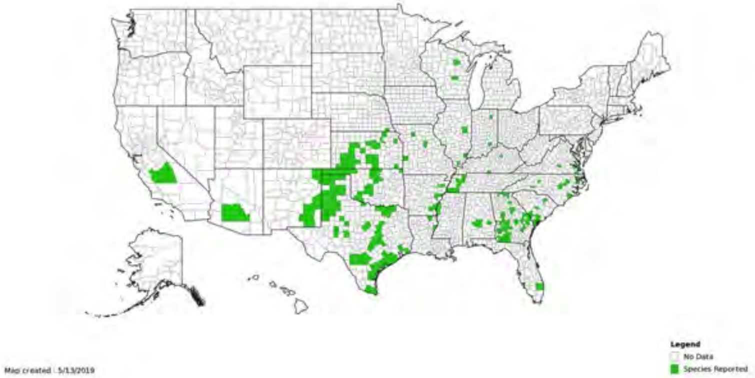
Resistance to Southern Leaf Blight
Kump et al. Nat Genet 2011

- Yield loss: Pesticide spraying increases direct cost and impacts to the environment, and human health
- Disease resistance “R”genes (NLR) are rapidly evolving and often seen in cluster. Good candidate for structural variation
- Pan Genomes: High quality reference assemblies to support characterization of core and dispensible (adaptive) genes

Sugarcane aphids

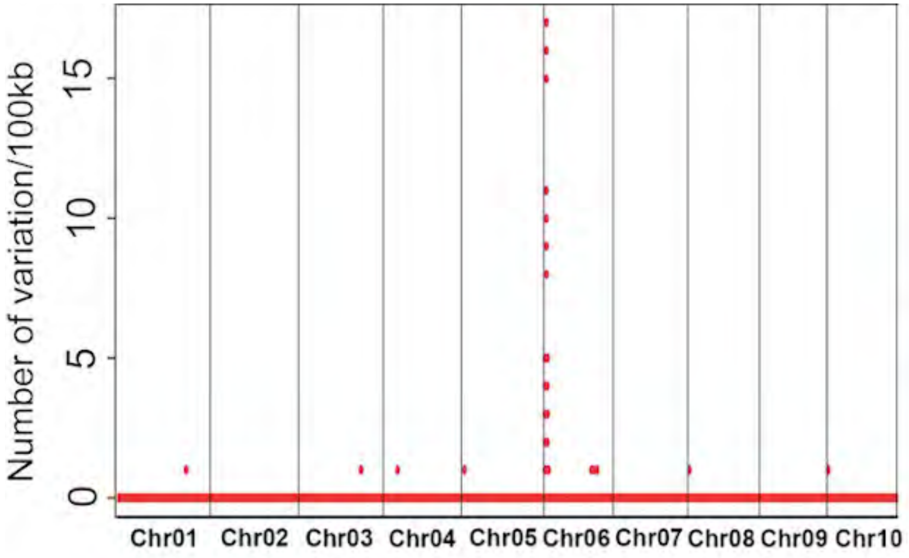
Collaborators:Yinghua Huang (USDA-ARS), Zhanguo Xin (USDA-ARS) Chad Hayes (USDA-ARS)

- Since 2013 in the US sugarcane aphids have been causing enormous damage to sorghum crop
- Tx2783 has high resistance to SCA



https://www.myfields.info/pests/sugarcane-aphid

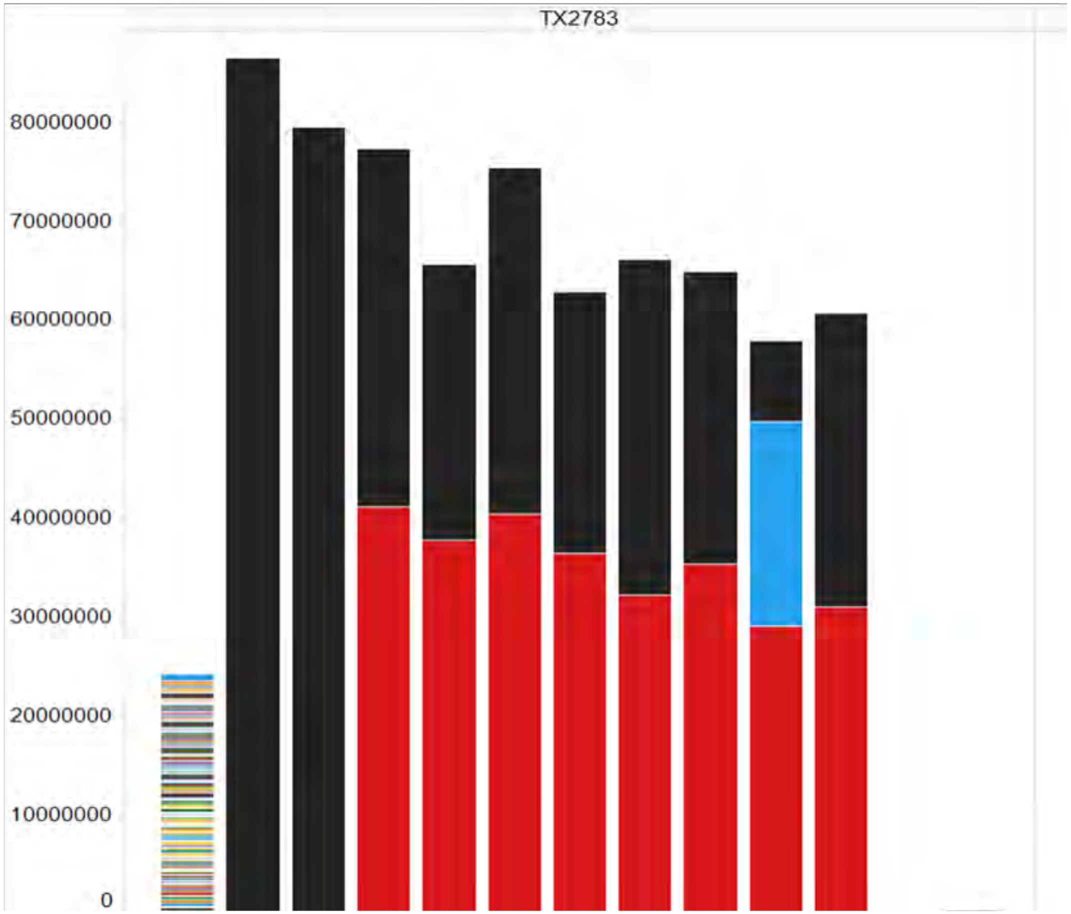
SCA locus mapped to the top of chromosojme 6



Wang et al., *bioRxiv*, 2021

Sorghum sugarcane aphid tolerant TX2783 reference assembly

447 PacBio contigs (25.6 Mb contig N50)



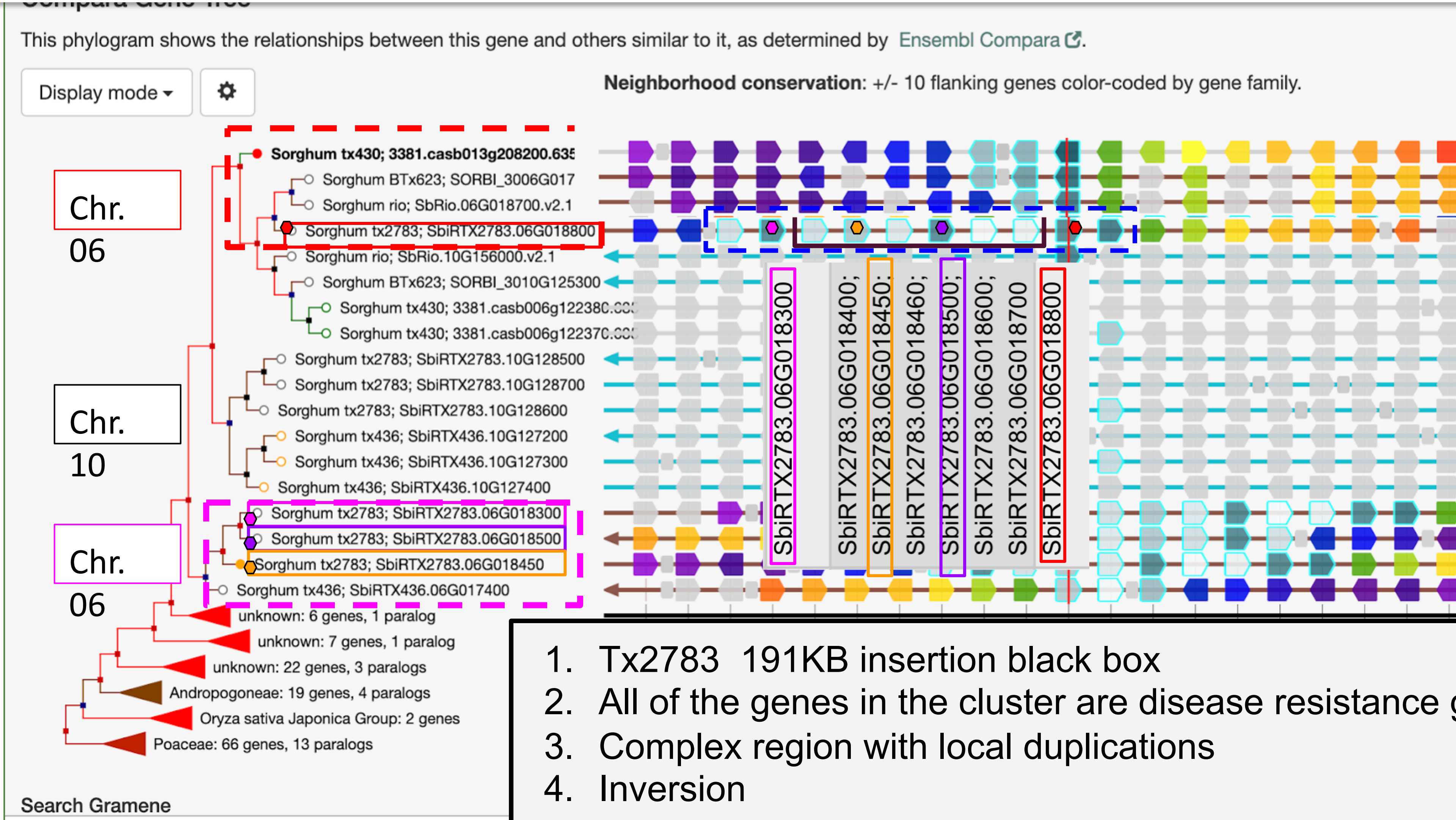
Wang et al., *bioRxiv*, 2021

Sorghum Pan Genomes site:
TX2783, BTX623, TX436, TX430, RIO
Assembly Gene Neighborhood conservation view


Andrew Olson


Sharon Wei


Kapeel Chou



Compara Gene Tree

This phylogram shows the relationships between this gene and others similar to it, as determined by Ensembl Compara.

Neighborhood conservation: +/- 10 flanking genes color-coded by gene family.

1. Tx2783 191KB insertion black box
2. All of the genes in the cluster are disease resistance
3. Complex region with local duplications
4. Inversion

Biology Enabled Agriculture

G X (environment +management) = P

Yield> Biotic Stress > Disease Resistance

Genome Editing
Marker Assisted Breeding

NLR R Genes

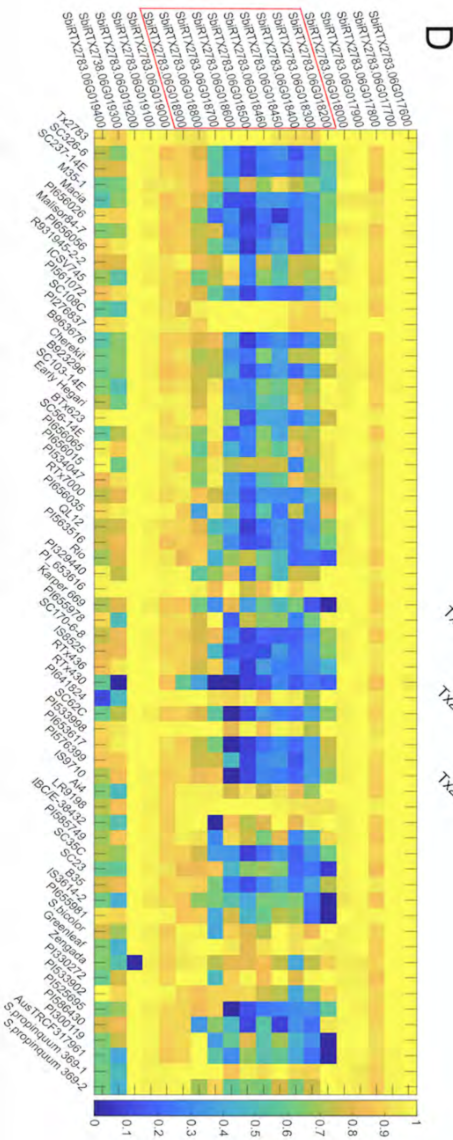
Expression

Signaling pathway

Genomes

TX2783 191 kb structural variant (SV) containing a cluster of R Genes

Resequencing data from 62 accessions identified the SV is segregating at a low frequency in these lines.



Yinping Jiao

Zhanguo Xin

Bo Wang

Liya Wang

Wang et al., *bioRxiv*, 2021

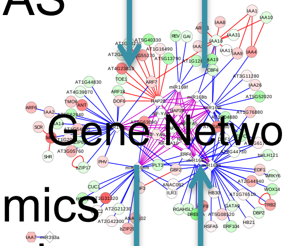
Exploring Genomes with an eye toward breeding

Biology Enabled Agriculture

Phenomes

QTL
GWAS

Genome Editing
Genomic Selection
Marker Assisted Breeding



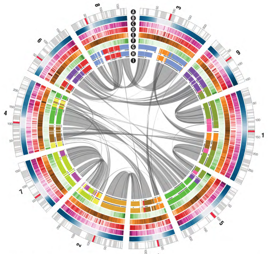
Gene Networks

Metabolic Pathways
Protein Complexes
Regulatory Networks

Metabolomics
Expression
Epigenetic

Genomes

Metabolites
Proteins
Transcripts
Chromatin



Association Analyses
Genes with major affect

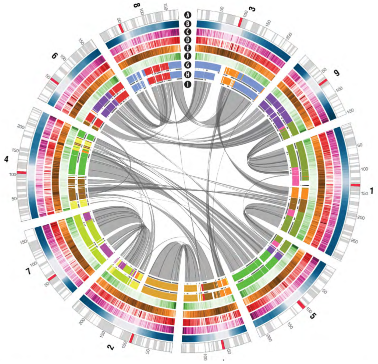
Regulatory and Inference Networks
Many genes with minor contributions associated with a complex trait

Comparative analysis
Transfer information across species

Improved understanding of biological mechanisms & systems will improve breeding models and support for genetic engineering

Genomes X (Environment +Management) = Phenotypes

Decreasing cost of sequencing leads to increasing computes and data management



$50 million (2009) Sequencing Centers

BAC library, Sanger sequencing library, finishing libraries,computes

$250- 180 thousand (2016) Sequencing Centers

PacBio long single molecule, Optical map, illumina short read

High quality DNA, Library prep, access sequencer & ***compute

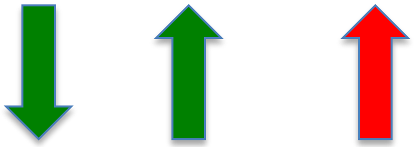
$90- 50 thousand (2017) Sequencing Centers

PacBio long single molecule, Optical map, 10X illumina short read

High quality DNA, Library prep, access sequencer & ***compute

$45- 25 thousand (2018) Local/ Sequencing Centers

$2-6 thousand (2021) Long single molecule



Decrease Sequence Cost

Improve Assembly Quality

Increase Compute Cost

Schnable, Ware et al, Science 2009
Jiao *et al*., Nature, 2017
Ou et. al, Genome Biology 2020
Liu et al, Nature Comm 2020
Wang et. al, submitted 2021
Hufford et. al, submitted 2021

BioNano Genomics

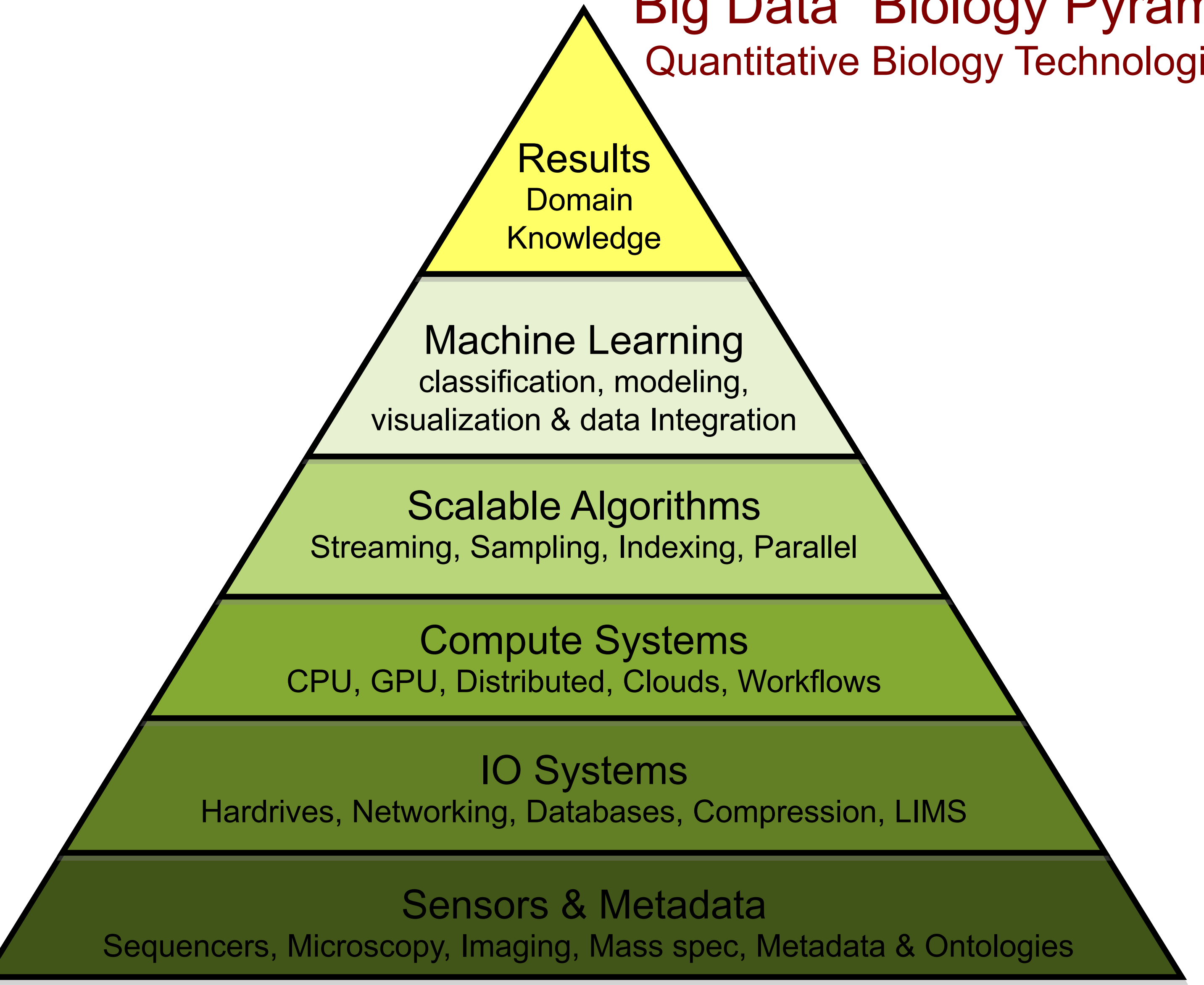
Pacific Biosciences

10x Genomics

Biology has transitioned to an Information Science

Big Data” Biology Pyramid

Quantitative Biology Technologies



Results
Domain
Knowledge

Machine Learning
classification, modeling,
visualization & data Integration

Scalable Algorithms
Streaming, Sampling, Indexing, Parallel

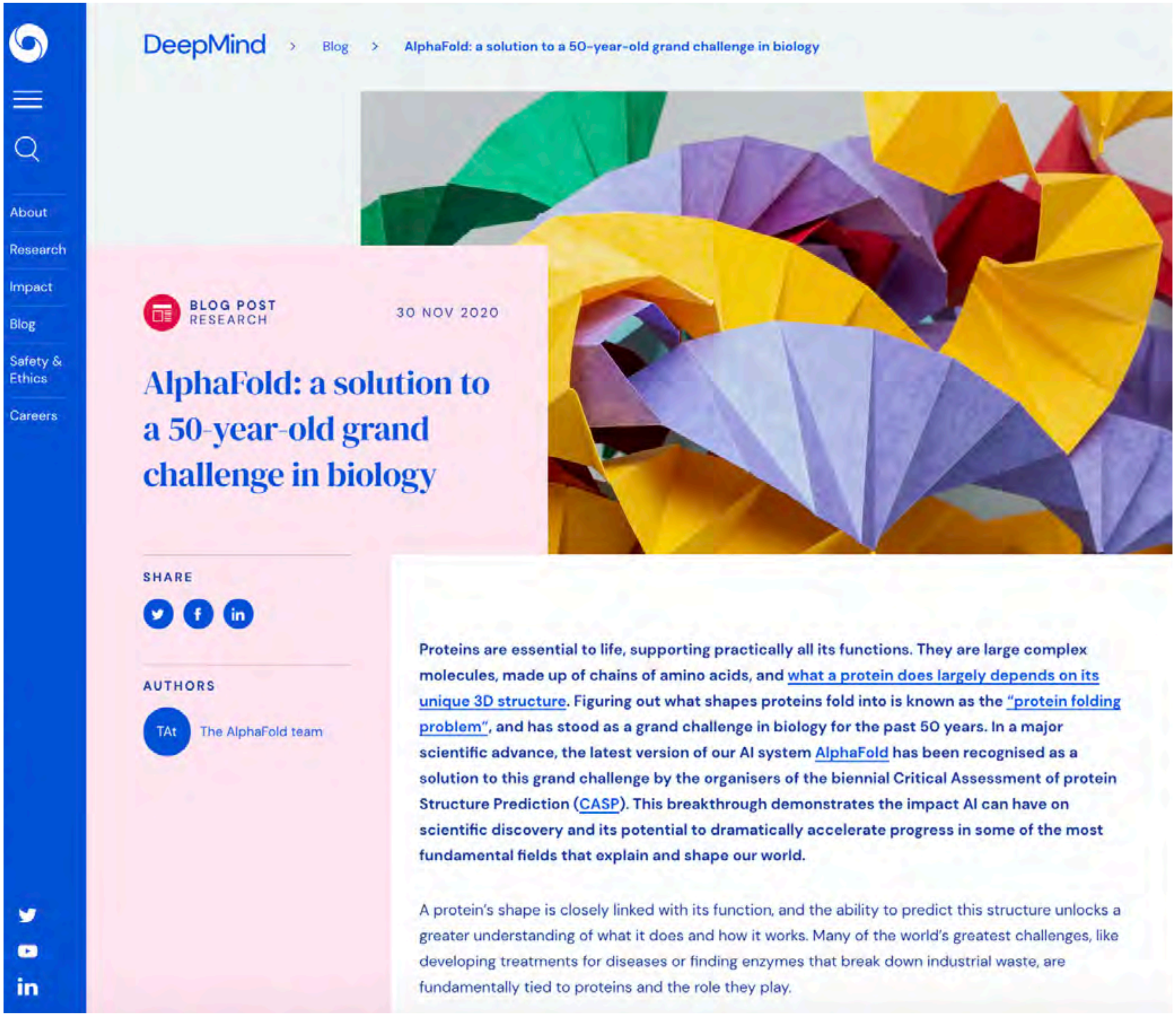
Compute Systems
CPU, GPU, Distributed, Clouds, Workflows

IO Systems
Hardrives, Networking, Databases, Compression, LIMS

Sensors & Metadata
Sequencers, Microscopy, Imaging, Mass spec, Metadata & Ontologies

Keystone Big Data in Biology 2014   Stein, Schatz, Ware

Artificial Inteligence (AI) is allowing us to reimagine how we approach science

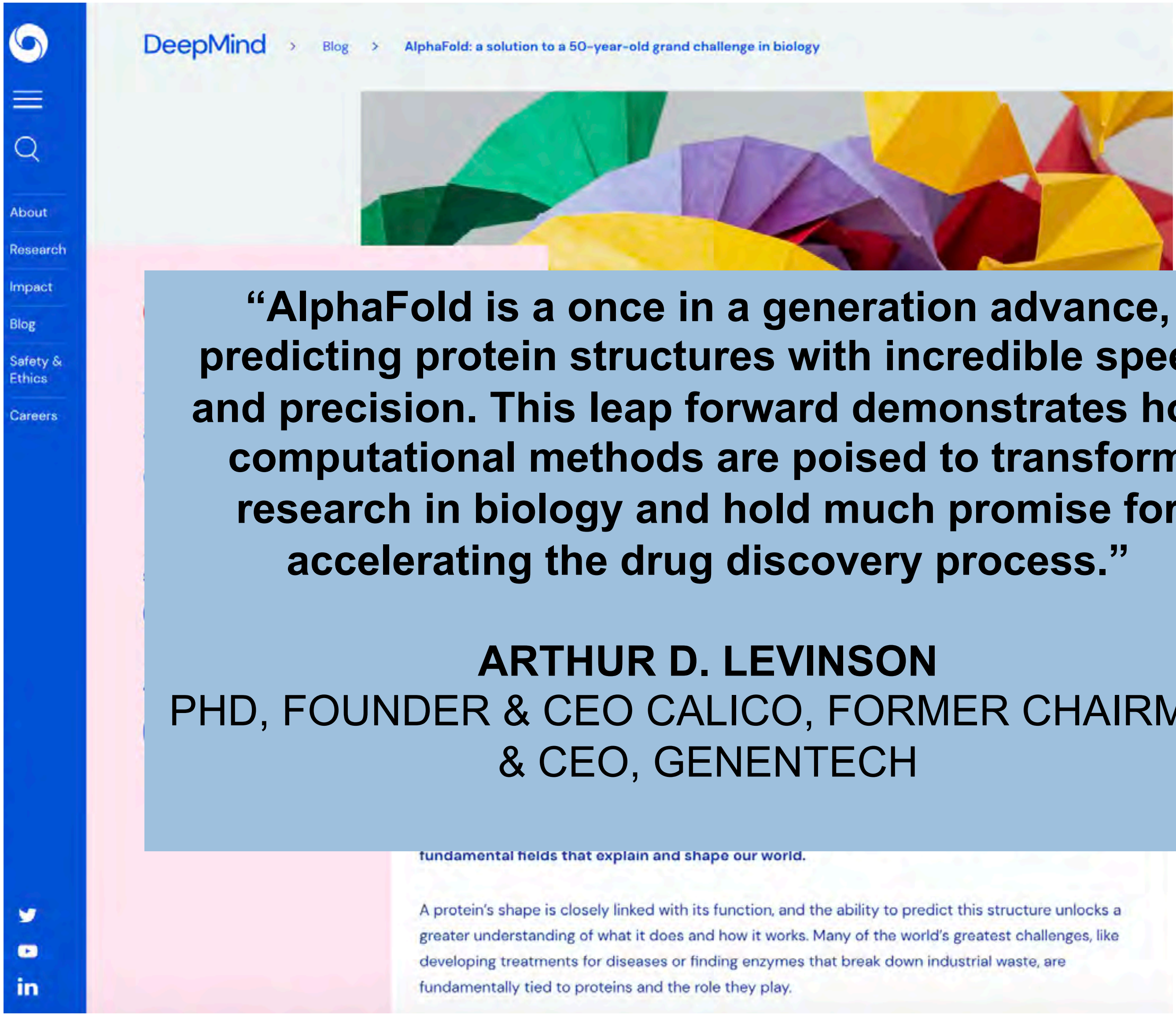


AlphaFold: a solution to a 50-year-old grand challenge in biology

Proteins are essential to life, supporting practically all its functions. They are large complex molecules, made up of chains of amino acids, and what a protein does largely depends on its unique 3D structure. Figuring out what shapes proteins fold into is known as the "protein folding problem", and has stood as a grand challenge in biology for the past 50 years. In a major scientific advance, the latest version of our AI system AlphaFold has been recognised as a solution to this grand challenge by the organisers of the biennial Critical Assessment of protein Structure Prediction (CASP). This breakthrough demonstrates the impact AI can have on scientific discovery and its potential to dramatically accelerate progress in some of the most fundamental fields that explain and shape our world.

A protein's shape is closely linked with its function, and the ability to predict this structure unlocks a greater understanding of what it does and how it works. Many of the world's greatest challenges, like developing treatments for diseases or finding enzymes that break down industrial waste, are fundamentally tied to proteins and the role they play.

Artificial Inteligence (AI) is allowing us to reimagine how we approach science



DeepMind > Blog > AlphaFold: a solution to a 50-year-old grand challenge in biology

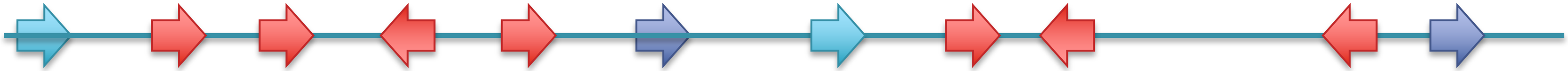
“AlphaFold is a once in a generation advance, predicting protein structures with incredible speed and precision. This leap forward demonstrates how computational methods are poised to transform research in biology and hold much promise for accelerating the drug discovery process.”

ARTHUR D. LEVINSON
PHD, FOUNDER & CEO CALICO, FORMER CHAIRMAN & CEO, GENENTECH

fundamental fields that explain and shape our world.

A protein’s shape is closely linked with its function, and the ability to predict this structure unlocks a greater understanding of what it does and how it works. Many of the world’s greatest challenges, like developing treatments for diseases or finding enzymes that break down industrial waste, are fundamentally tied to proteins and the role they play.

Wild Grapes have natural resistance to disease



Wild Grapes come with baggage they have shed with domestication

Unpleasant taste

Wild:
Male and Female

Domesticated:
Hermaphrodite

This, (2006); Feechan, et.al (2015)

Plant NLR R-genes

CC

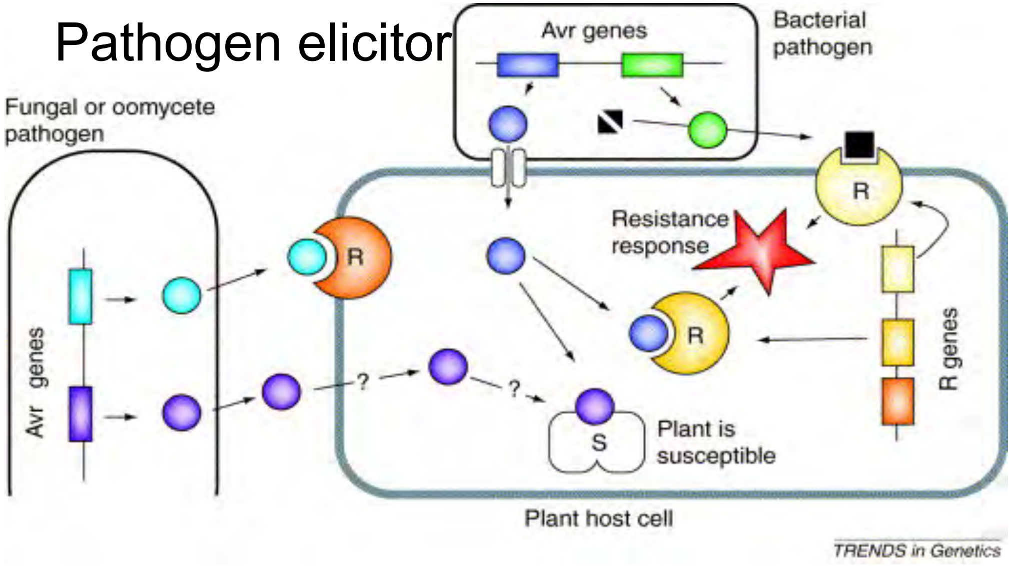
coil-coil

NB-ARC

ATP-binding

LRR

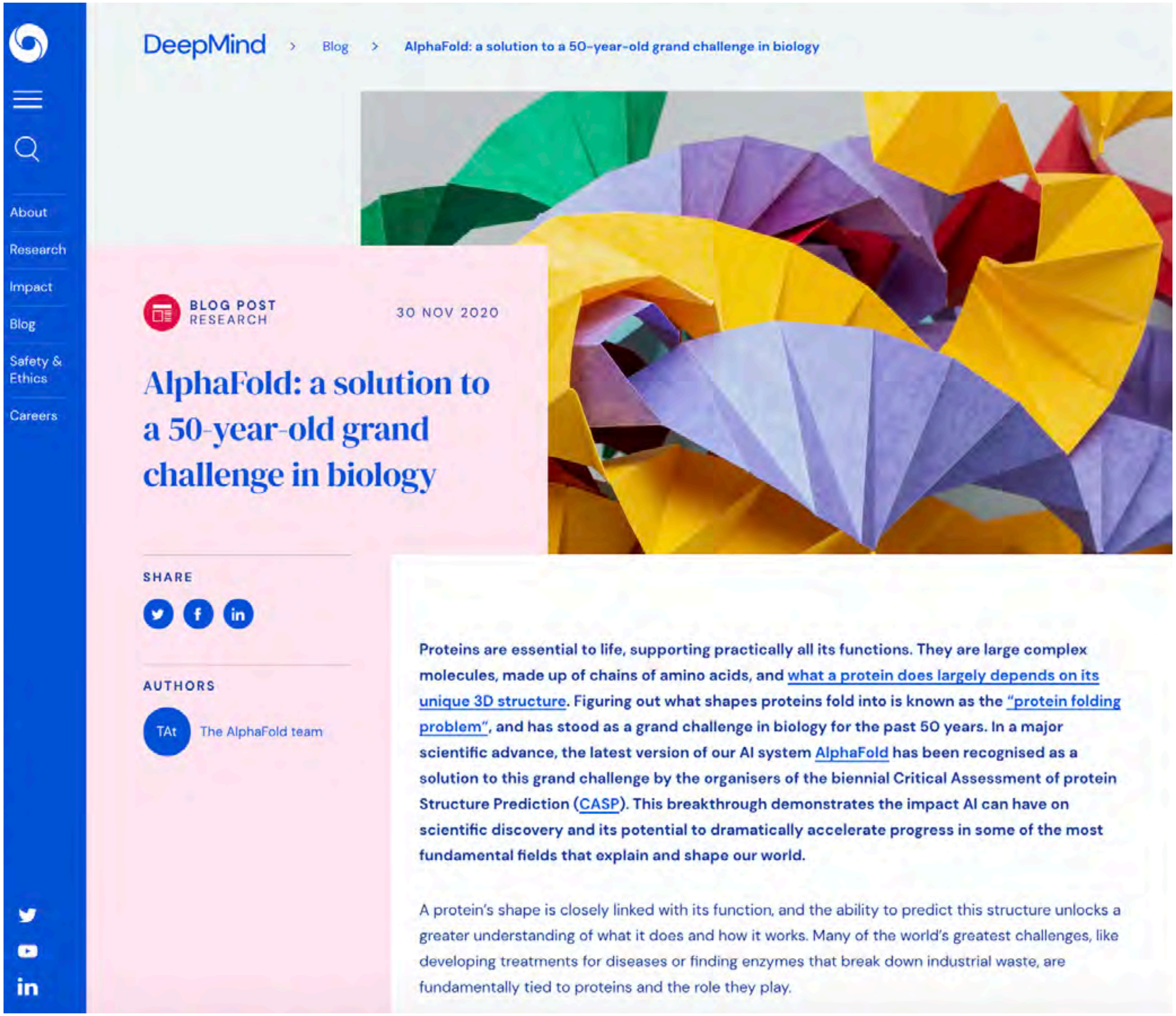
Pathogen elicitor



Gene-for-gene hypothesis

Tyler B. (2001). Trends in Genetics. 17(11):611-614

Imagine a world where you can begin to model protein ligand models for disease resitance genes!



BLOG POST
RESEARCH

30 NOV 2020

AlphaFold: a solution to a 50-year-old grand challenge in biology

SHARE

AUTHORS

The AlphaFold team

Proteins are essential to life, supporting practically all its functions. They are large complex molecules, made up of chains of amino acids, and what a protein does largely depends on its unique 3D structure. Figuring out what shapes proteins fold into is known as the "protein folding problem", and has stood as a grand challenge in biology for the past 50 years. In a major scientific advance, the latest version of our AI system AlphaFold has been recognised as a solution to this grand challenge by the organisers of the biennial Critical Assessment of protein Structure Prediction (CASP). This breakthrough demonstrates the impact AI can have on scientific discovery and its potential to dramatically accelerate progress in some of the most fundamental fields that explain and shape our world.

A protein's shape is closely linked with its function, and the ability to predict this structure unlocks a greater understanding of what it does and how it works. Many of the world's greatest challenges, like developing treatments for diseases or finding enzymes that break down industrial waste, are fundamentally tied to proteins and the role they play.

The Next Green Revolution will be "Data" and "Design" driven

Agriculture has transition to a Data Science. Massive data generation—genotypes, phenotypes, soil & environment.

Data now exceeds human capacity to formulate and test hypotheses about gene function, regulatory networks, and predictive agriculture.

Develop new approaches and systems that can supply new hypotheses for researchers for crop breeding, fermentation systems, solar energy capture, pest and disease management.

Adapt what has evolved in nature and design the space nature missed

Norman Borlaug
Nobel Peace Prize 1970

While CRISPR has the power to enhance breeding by rapidly customizing and optimizing crop productivity

Genome Biology provides the insights to drive the translation

Unraveling Complex Traits using Genomic, Genetic, Systems & Computational Approaches

Ware Lab

Kapeel Chougule
Nick Gladman
YInping Jiao
Vivek Kumar
Sunita Kumari
Forrest Li
Augusto Lima Dinz
Zhenyuan Lu
Andrew Olson
Michael Regulski
Bo Wang
Liya Wang
Xiaofei Wang
Marcela Tello Ruiz
Sharon Wei
Peter Van Buren
Lifang Zhang
Carol Hu

CSHL

Dick McCombie
Rob Martienssen
Dave Jackson
Tom Gingeras
Dave Micklos

Uplands Farm

Tim Mulligan
Kyle Schlecht

HPCC resources

Todd Heywood

USDA

NSF

Cold Spring Harbor Laboratory

Advancing Agriculture Through Collaborative Research on Crop & Model Species