Exploring and exploiting the cotton pan-genome

John Lovell, Avinash Sreedasyam & Jeremy Schmutz

HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
Department of Energy Joint Genome Institute, Berkeley, CA, USA

16-July 2024

Team & Funding

HUDSONALPHA
GENOME SEQUENCING
CENTER

Don C. Jones
Cotton Inc.

Peng W. Chee
U of Georgia, Tifton

Christopher A. Saski
Clemson University

Shengqiang Shu

Tomáš Brůna

JGI ANNOTATION / PHYTOZOME

Cotton Incorporated

USDA

NIFA

NERSC

National Science Foundation
Directorate for Biological Sciences (BIO)

JGI JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

HudsonAlpha Genome Sequencing Center

Z. Jeffrey Chen
UT Austin

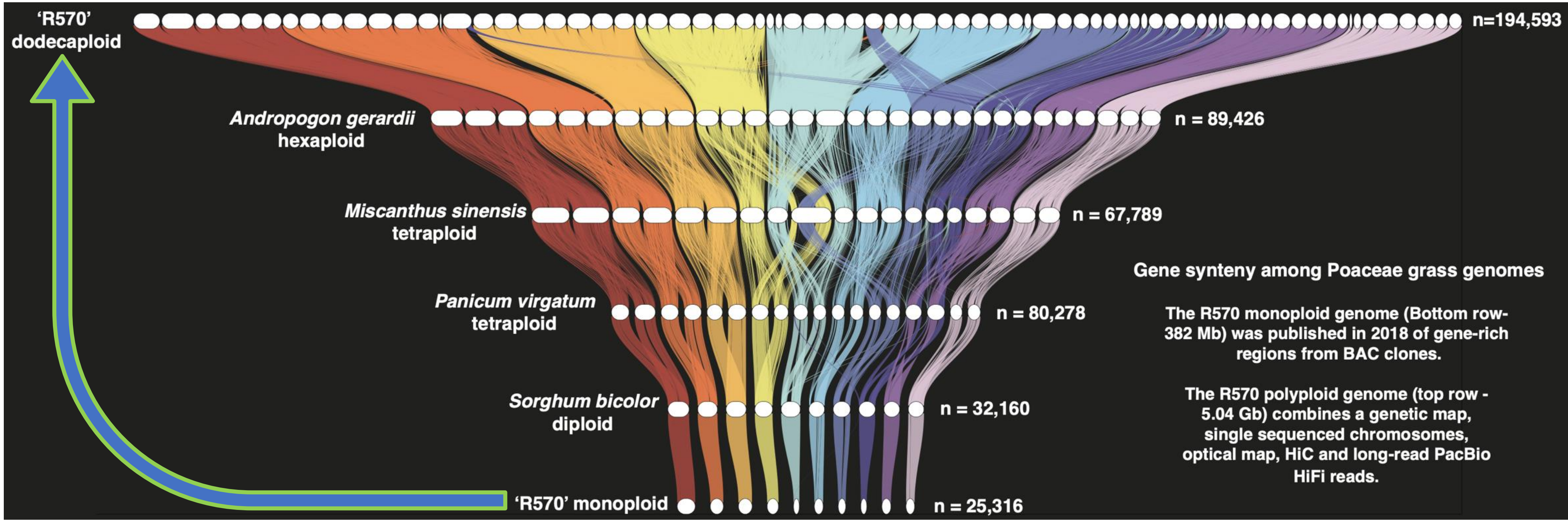
Fred Bourland
U of Arkansas

Warwick N. Stiller
CSIRO, Australia

HAGSC Genome Analysis

HUDSONALPHA
GENOME SEQUENCING
CENTER

We build complex reference genome assemblies



Gene synteny among Poaceae grass genomes

The R570 monoploid genome (Bottom row- 382 Mb) was published in 2018 of gene-rich regions from BAC clones.

The R570 polyploid genome (top row - 5.04 Gb) combines a genetic map, single sequenced chromosomes, optical map, HiC and long-read PacBio HiFi reads.

Healey et al. (2024) Nature

Cursi et al. (2021)

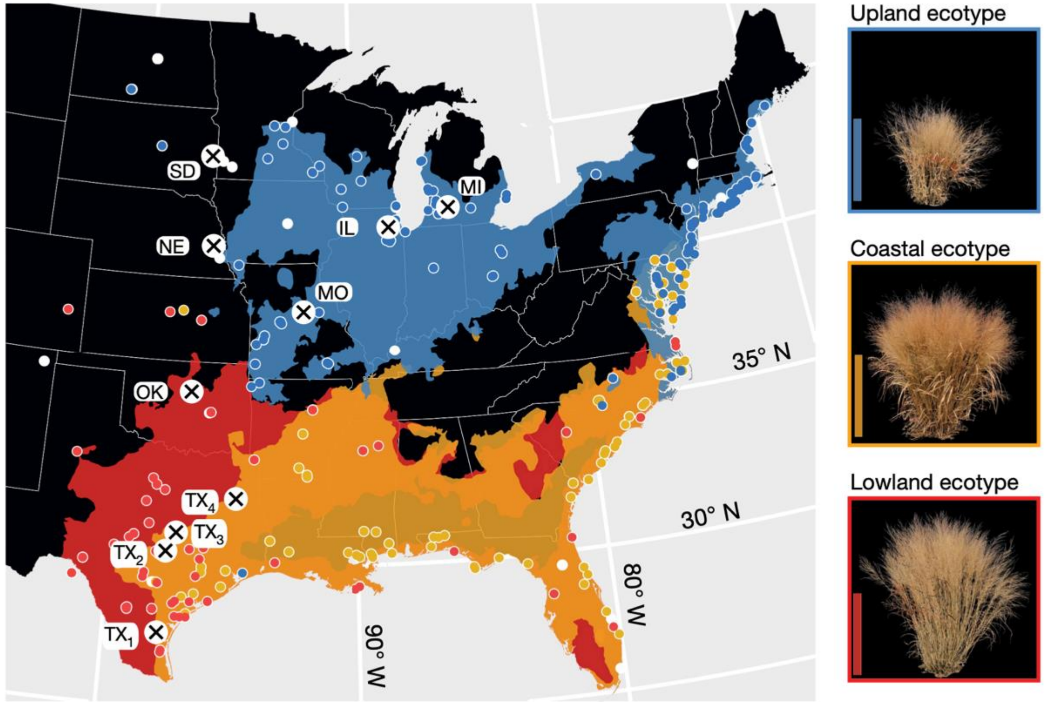

eLife

TOOLS AND RESOURCES

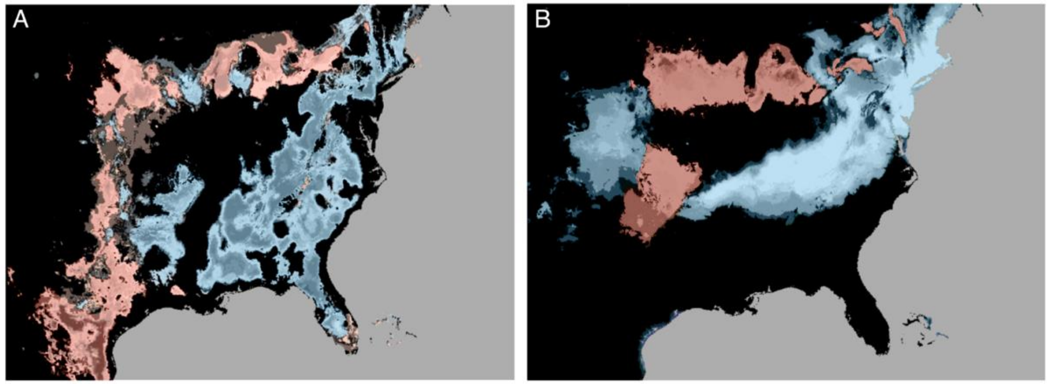
GENESPACE tracks regions of interest and gene copy number variation across multiple genomes

John T Lovell[1,2]*, Avinash Sreedasyam[1], M Eric Schranz[3], Melissa Wilson[4], Joseph W Carlson[2], Alex Harkess[1,5], David Emms[6], David M Goodstein[2], Jeremy Schmutz[1,2]

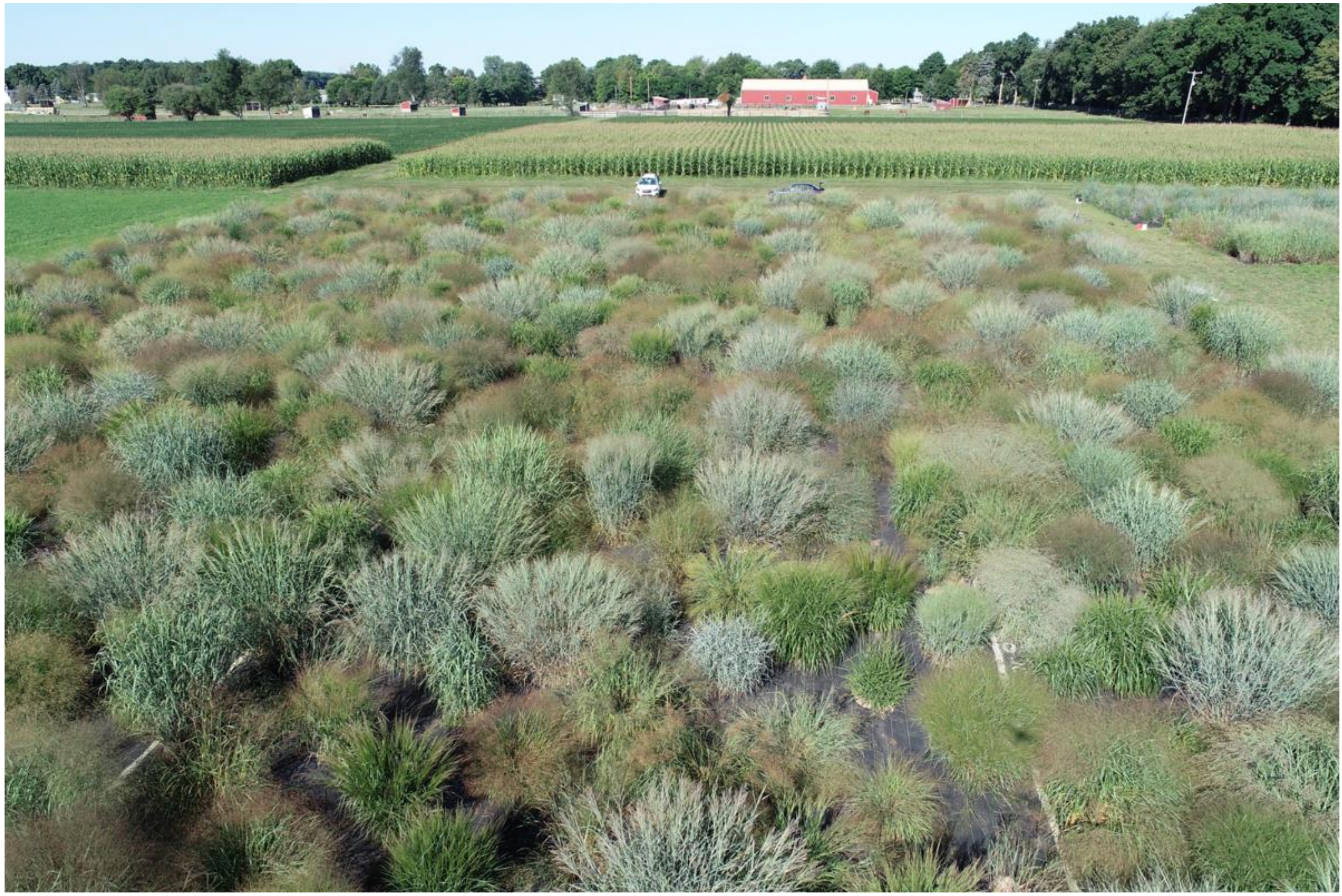
Custom evolutionary analysis to optimize complex breeding outcomes



Lovell et al. (2021) Nature



Napier et al. (2022) PNAS



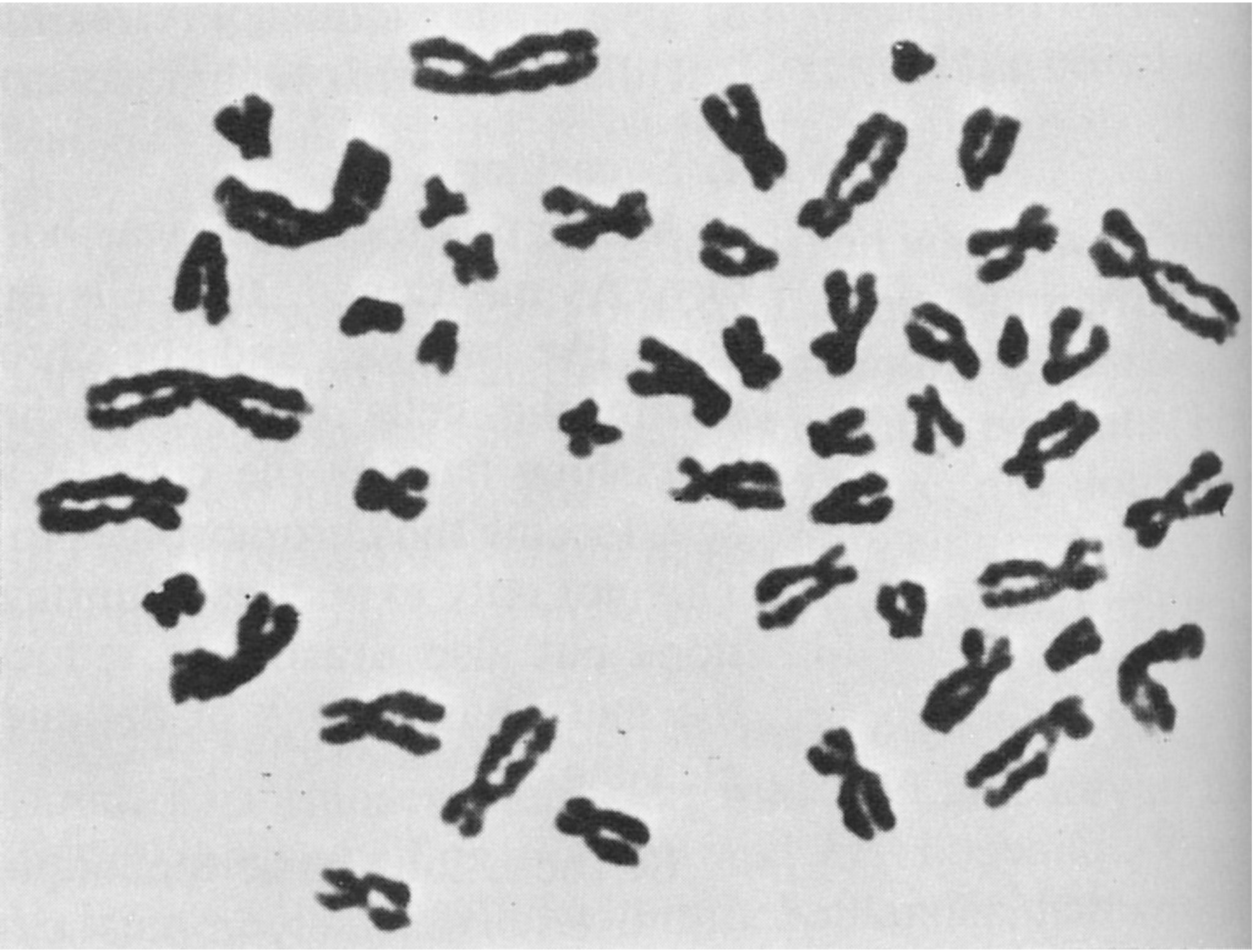
Outline

HUDSONALPHA
GENOME SEQUENCING
CENTER

- **Genomics**: why it's important and what progress has been made
- **Cotton genomes**: what has been done, and what is left to do
- **Conceptual issues with pan-genomes**: why we need integrated resources
- Current **state of the cotton pan-genome**: genes, variants and tools
- **Using the pan-genome**: tracking QTL and Pima introgressions
- **Plans for the pan-genome**: what do we need to do to aid breeders and biotechnology (open for questions/comments)

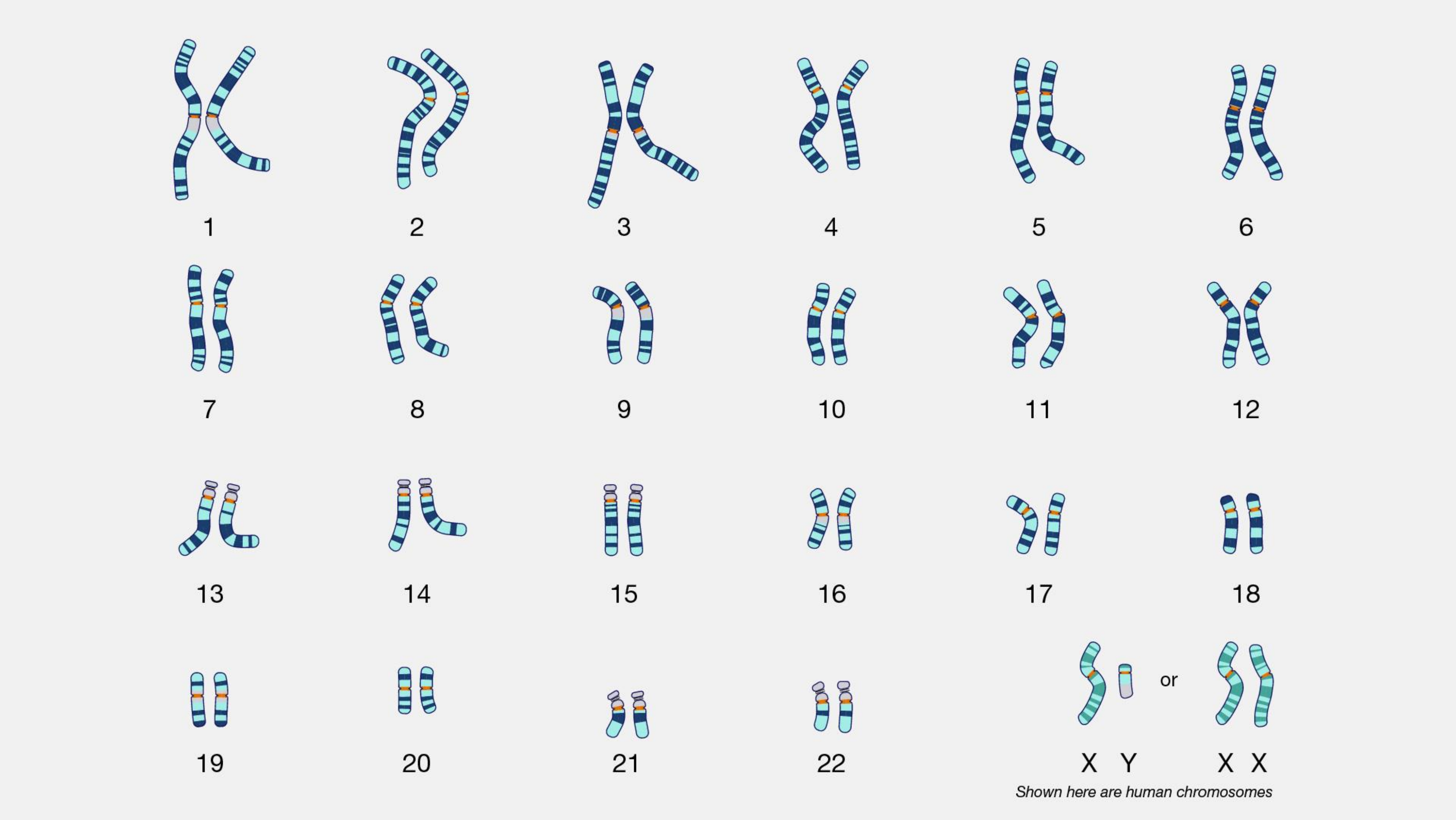
What is a genome?

HUDSONALPHA GENOME SEQUENCING CENTER

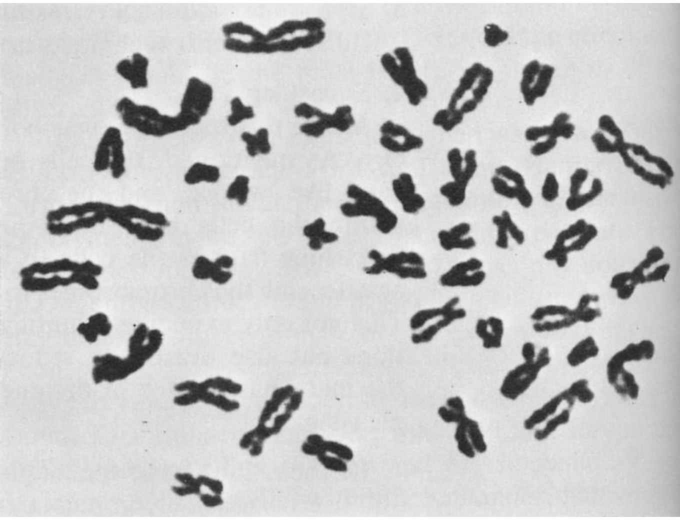


What is a genome?

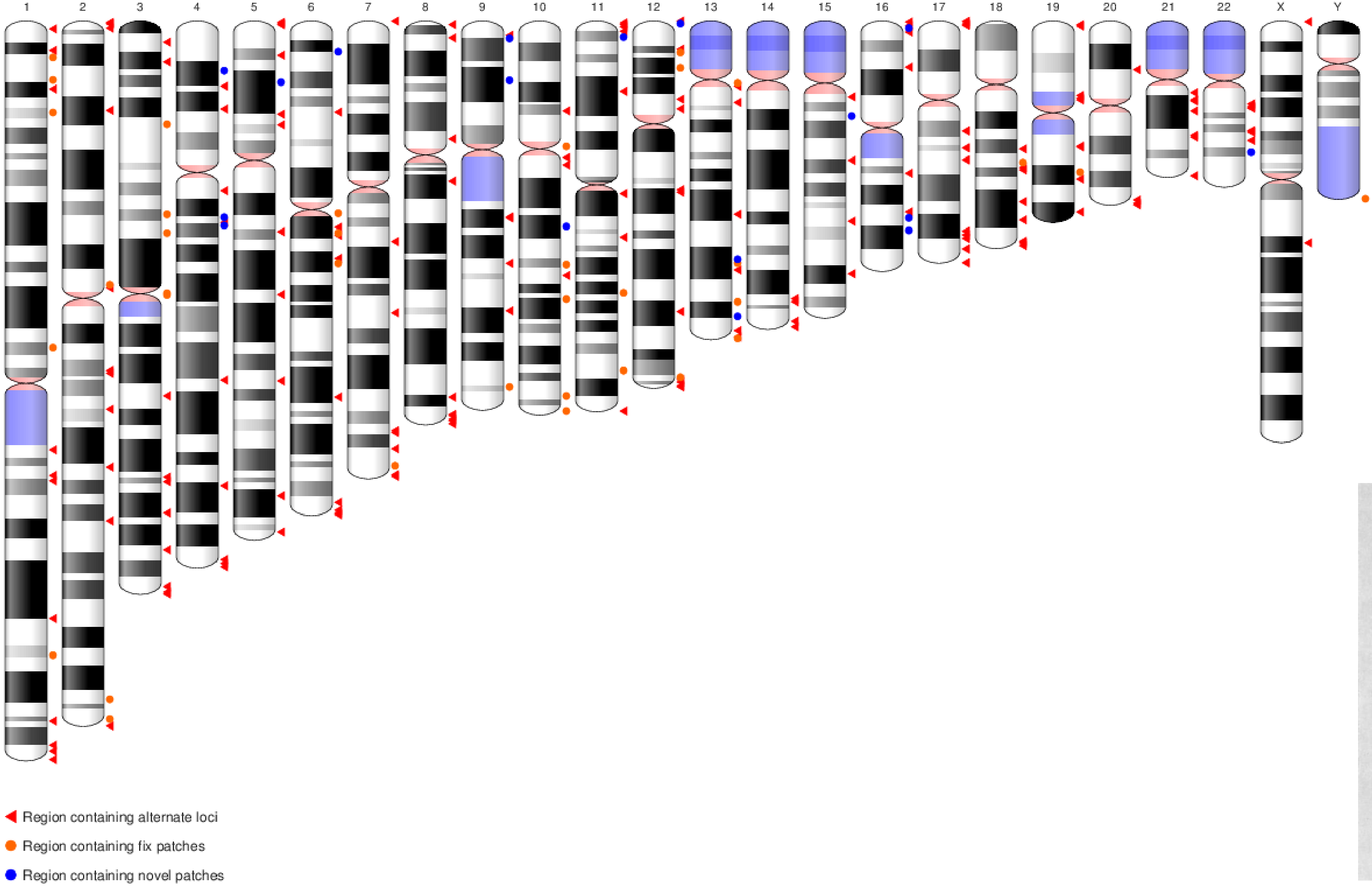
HUDSONALPHA GENOME SEQUENCING CENTER

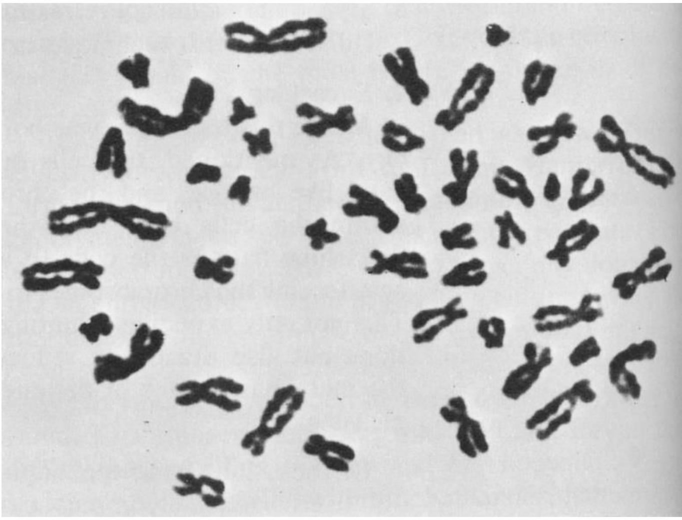


Shown here are human chromosomes



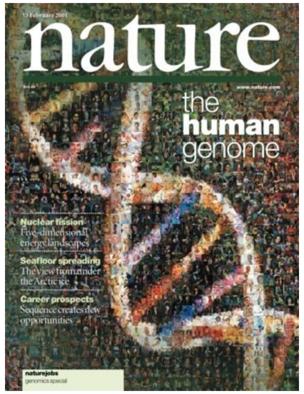
What is a genome?





State of the art in genomics

HUDSONALPHA
GENOME SEQUENCING
CENTER



15 February 2001

nature

www.nature.com

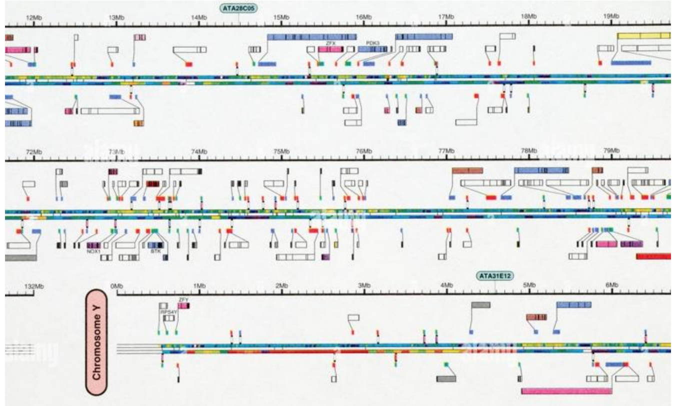
the
human
genome

Nuclear fission
Five-dimensional energy landscapes

Seafloor spreading
The view from under the Arctic ice

Career prospects
Sequence creates new opportunities

naturejobs
genomics special



Earth's heart of iron begins to yield its secrets p. 18

Microglia in chronic pain recovery and relapse pp. 33 & 86

Particle acceleration in a nova explosion p. 77

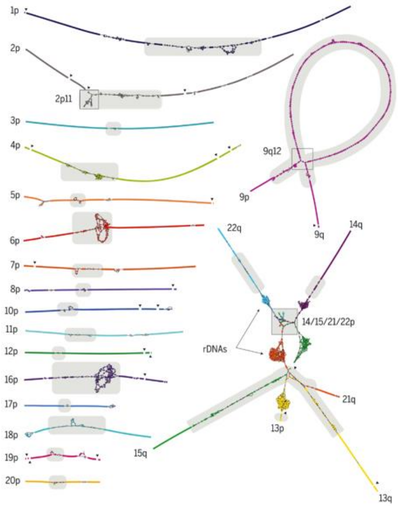
Science

$15
1 APRIL 2022
SPECIAL ISSUE
science.org

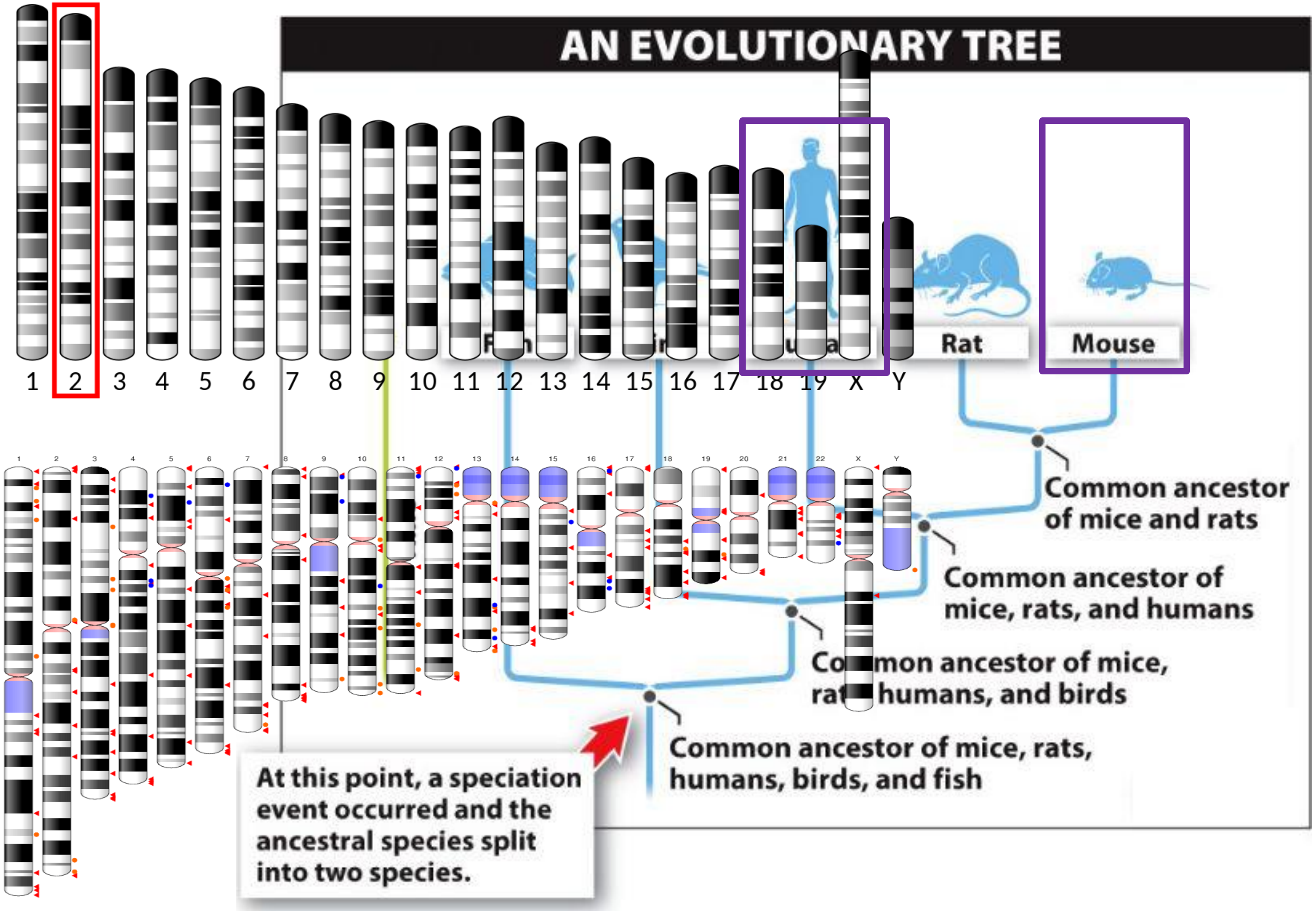
AAAS

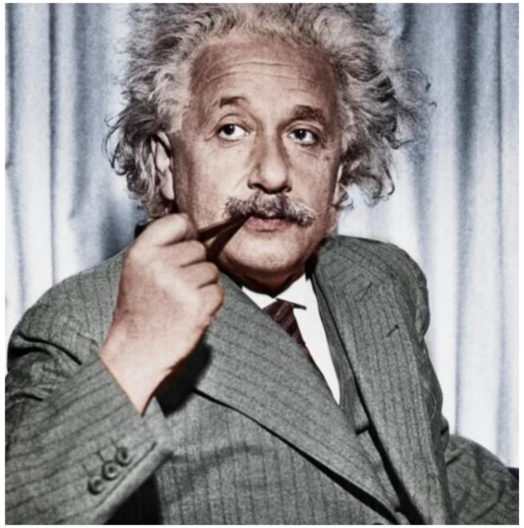
FILLING THE GAPS

Closing in on a complete human genome p. 42

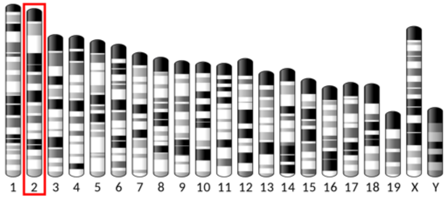


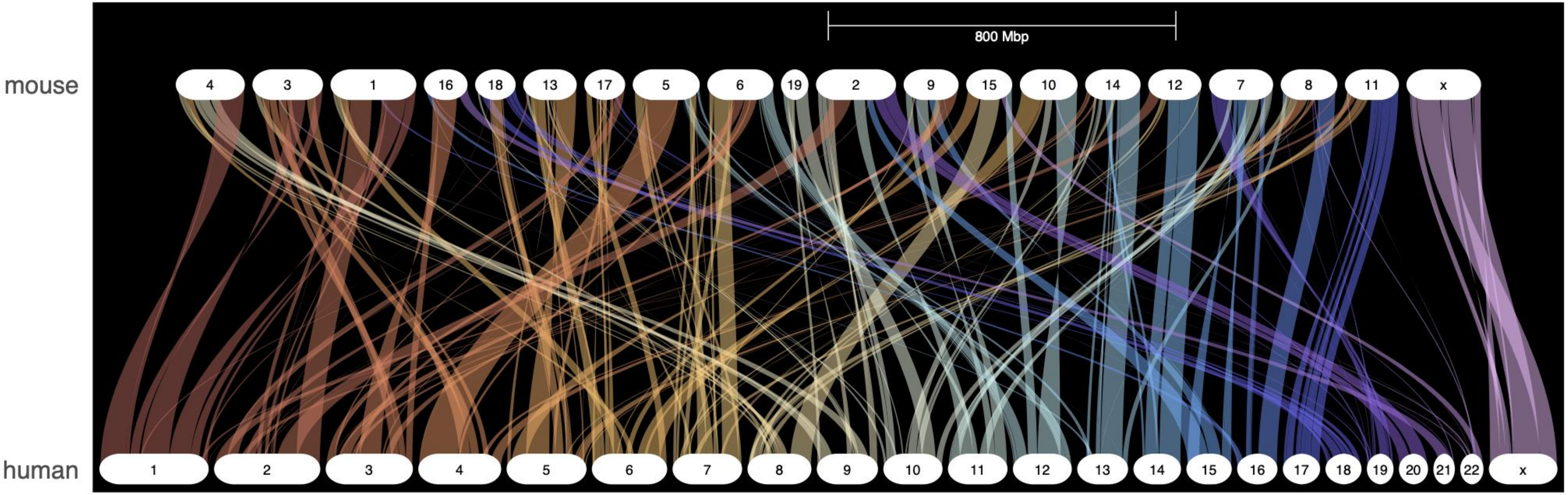
Comparing genomes



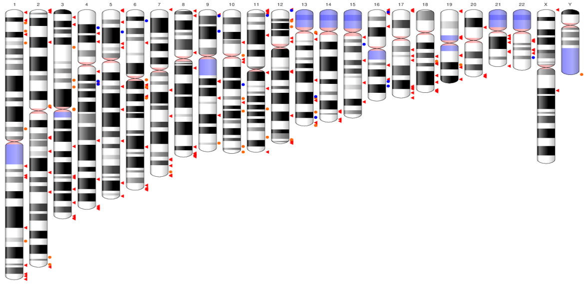


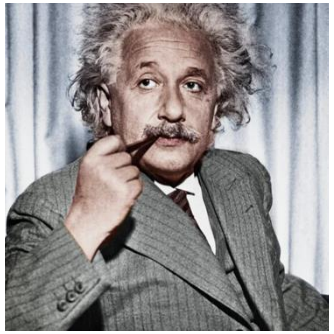
Comparing genomes



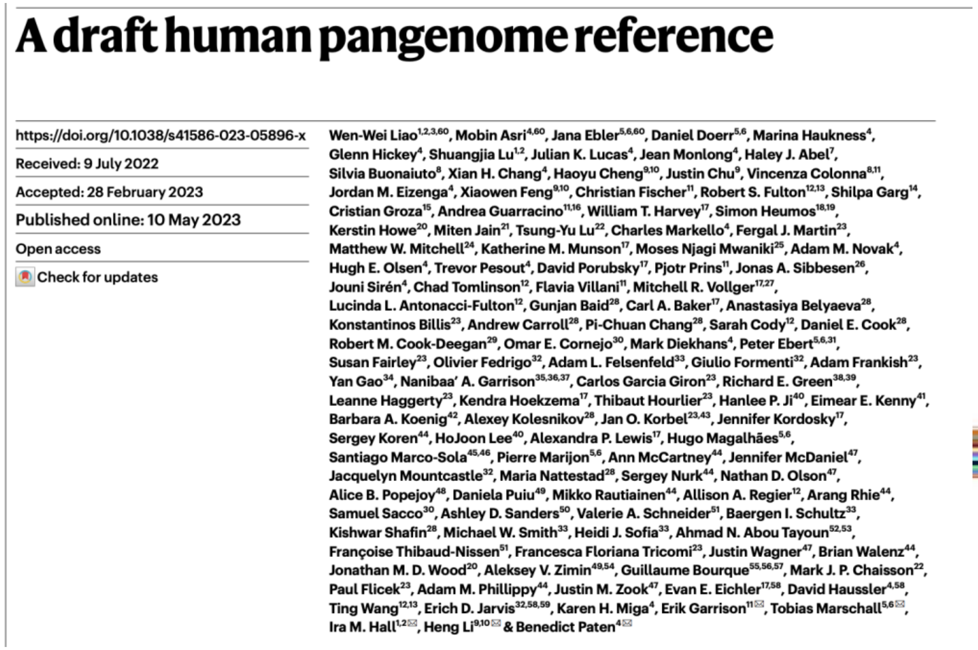


Chromosomes scaled by physical position





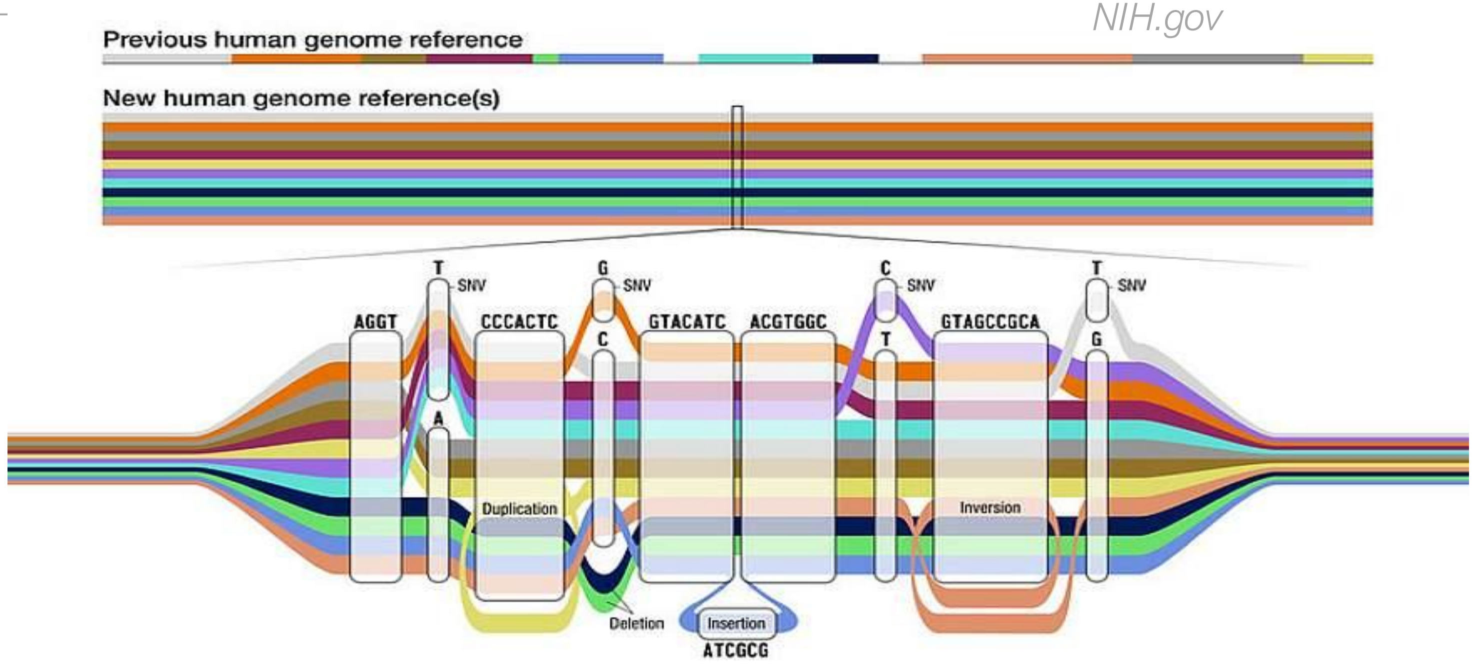
The concept of a pan-genome graph



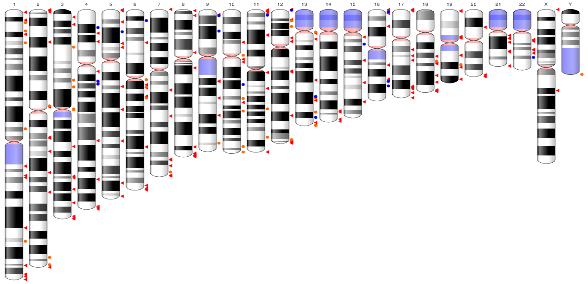
A draft human pangenome reference

https://doi.org/10.1038/s41586-023-05896-x
Received: 9 July 2022
Accepted: 28 February 2023

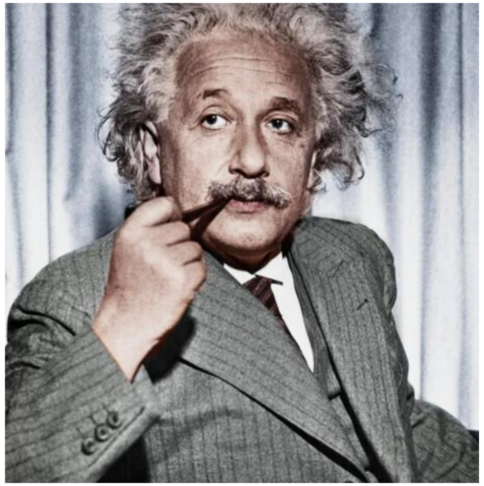
Wen-Wei Liao[1,2,3,60], Mobin Asri[4,60], Jana Ebler[5,6,60], Daniel Doerr[5,6], Marina Haukness[4], Glenn Hickey[4], Shuangjia Lu[1,2], Julian K. Lucas[4], Jean Monlong[4], Haley J. Abel[7], Silvia Buonaiuto[8], Xian H. Chang[4], Haoyu Cheng[9,10], Justin Chu[9], Vincenza Colonna[8,11], Jordan M. Eizenga[4], Xiaowen Feng[9,10], Christian Fischer[11], Robert S. Fulton[12,13], Shilpa Garg[14], Cristian Groza[15], Andrea Guarracino[11,16], William T. Harvey[17], Simon Heumos[18,19], Kerstin Howe[20], Miten Jain[21], Tsung-Yu Lu[22], Charles Markello[4], Fergal J. Martin[23], Matthew W. Mitchell[24], Katherine M. Munson[17], Moses Njagi Mwaniki[25], Adam M. Novak[4], Hugh E. Olsen[4], Trevor Pesout[4], David Porubsky[17], Pjotr Prins[11], Jonas A. Sibbesen[26], Jouni Sirén[4], Chad Tomlinson[12], Flavia Villani[11], Mitchell R. Vollger[17,27], Lucinda L. Antonacci-Fulton[12], Gunjan Baid[28], Carl A. Baker[17], Anastasiya Belyaeva[28], Konstantinos Billis[23], Andrew Carroll[28], Pi-Chuan Chang[28], Sarah Cody[12], Daniel E. Cook[28], Robert M. Cook-Deegan[29], Omar E. Cornejo[30], Mark Diekhans[4], Peter Ebert[5,6,31], Susan Fairley[23], Olivier Fedrigo[32], Adam L. Felsenfeld[33], Giulio Formenti[32], Adam Frankish[23], Yan Gao[34], Nanibaa’ A. Garrison[35,36,37], Carlos Garcia Giron[23], Richard E. Green[38,39], Leanne Haggerty[23], Kendra Hoekzema[17], Thibaut Hourlier[23], Hanlee P. Ji[40], Eimear E. Kenny[41], Barbara A. Koenig[42], Alexey Kolesnikov[28], Jan O. Korbel[23,43], Jennifer Kordosky[17], Sergey Koren[44], HoJoon Lee[40], Alexandra P. Lewis[17], Hugo Magalhães[5,6], Santiago Marco-Sola[45,46], Pierre Marijon[5,6], Ann McCartney[44], Jennifer McDaniel[47], Jacquelyn Mountcastle[32], Maria Nattestad[28], Sergey Nurk[44], Nathan D. Olson[47], Alice B. Popejoy[48], Daniela Puiu[49], Mikko Rautiainen[44], Allison A. Regier[12], Arang Rhie[44], Samuel Sacco[30], Ashley D. Sanders[50], Valerie A. Schneider[51], Baergen I. Schultz[33], Kishwar Shafin[28], Michael W. Smith[33], Heidi J. Sofia[33], Ahmad N. Abou Tayoun[52,53], Françoise Thibaud-Nissen[51], Francesca Floriana Tricomi[23], Justin Wagner[47], Brian Walenz[44], Jonathan M. D. Wood[20], Aleksey V. Zimin[49,54], Guillaume Bourque[55,56,57], Mark J. P. Chaisson[22], Paul Flicek[23], Adam M. Phillippy[44], Justin M. Zook[47], Evan E. Eichler[17,58], David Haussler[4,58], Ting Wang[12,13], Erich D. Jarvis[32,58,59], Karen H. Miga[4], Erik Garrison[11] ✉, Tobias Marschall[5,6] ✉, Ira M. Hall[1,2] ✉, Heng Li[9,10] ✉ & Benedict Paten[4] ✉

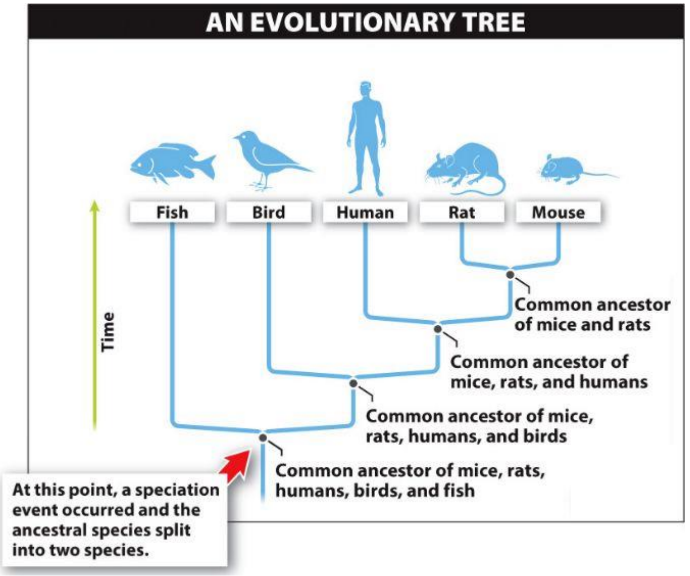


NIH.gov

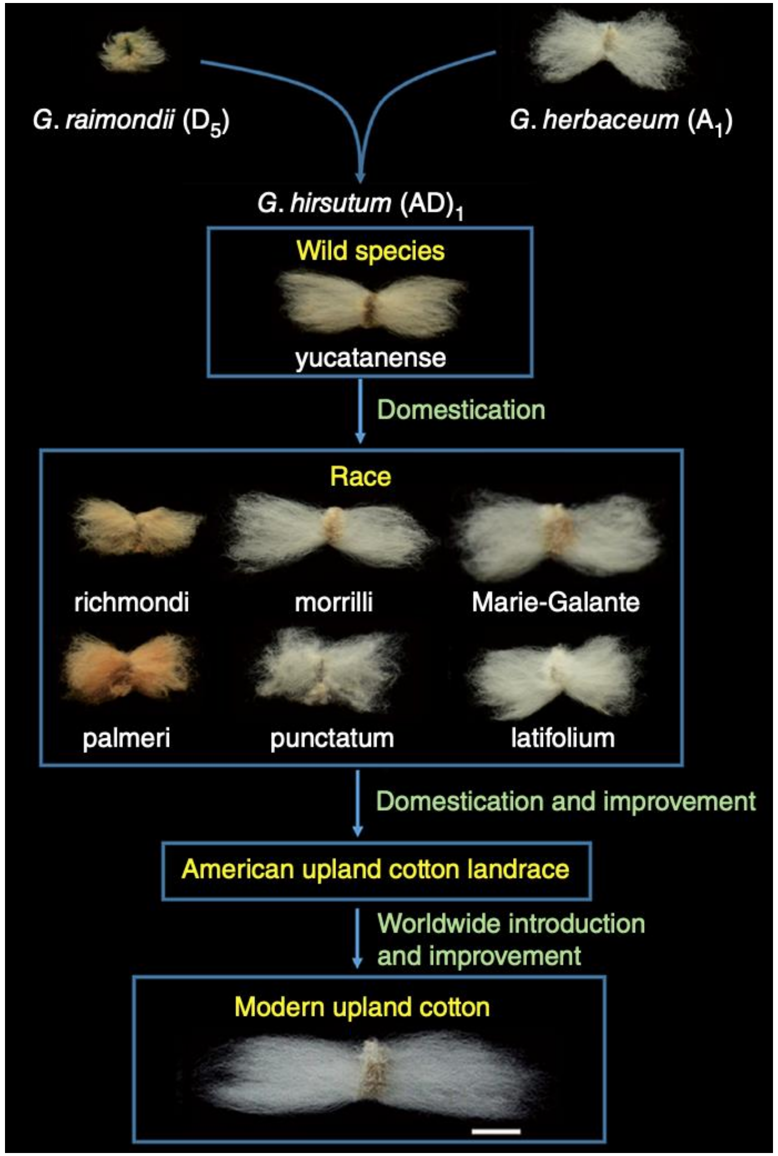


Structure of the tetraploid cotton genomes

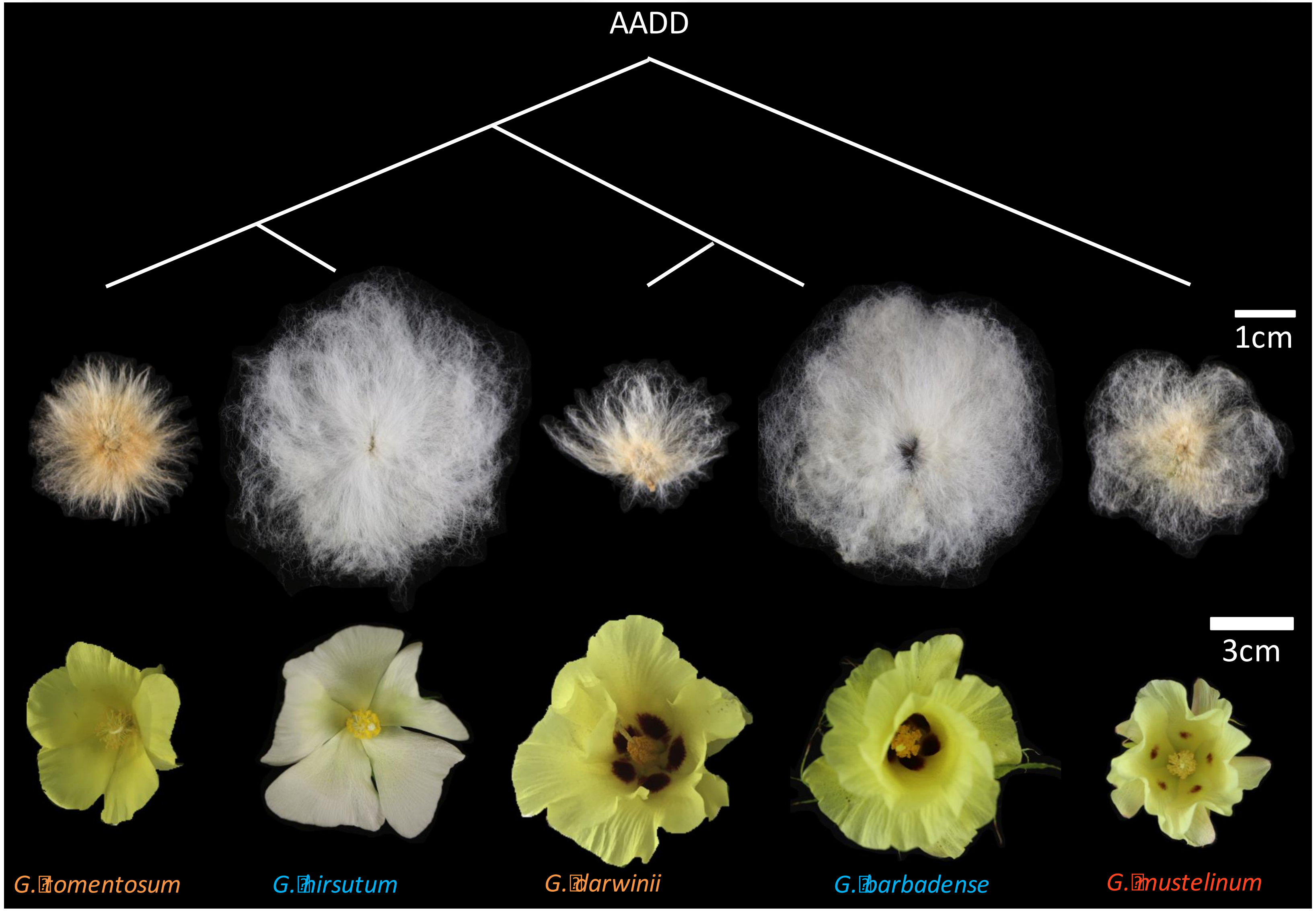


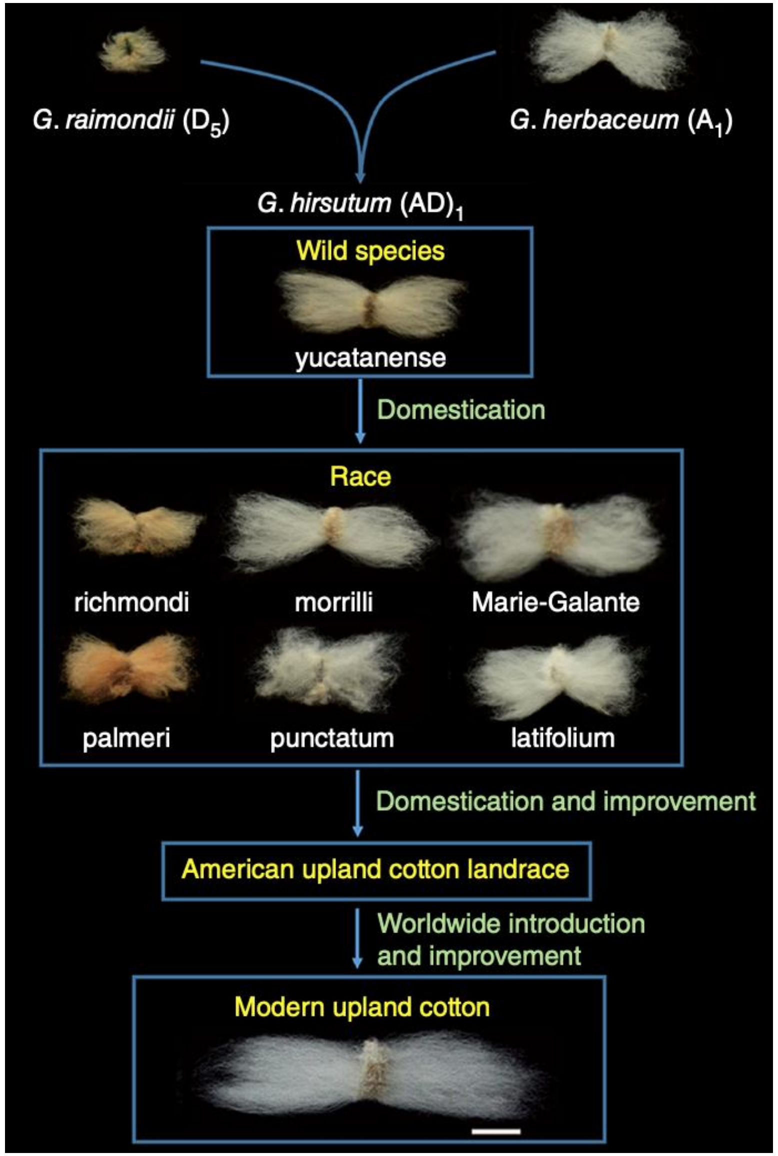


$\neq$



Evolutionary history of cotton




G. raimondii (D_5)

G. herbaceum (A_1)

G. hirsutum $(AD)_1$

Wild species

yucatanense

Domestication

Race

richmondi

morrilli

Marie-Galante

palmeri

punctatum

latifolium

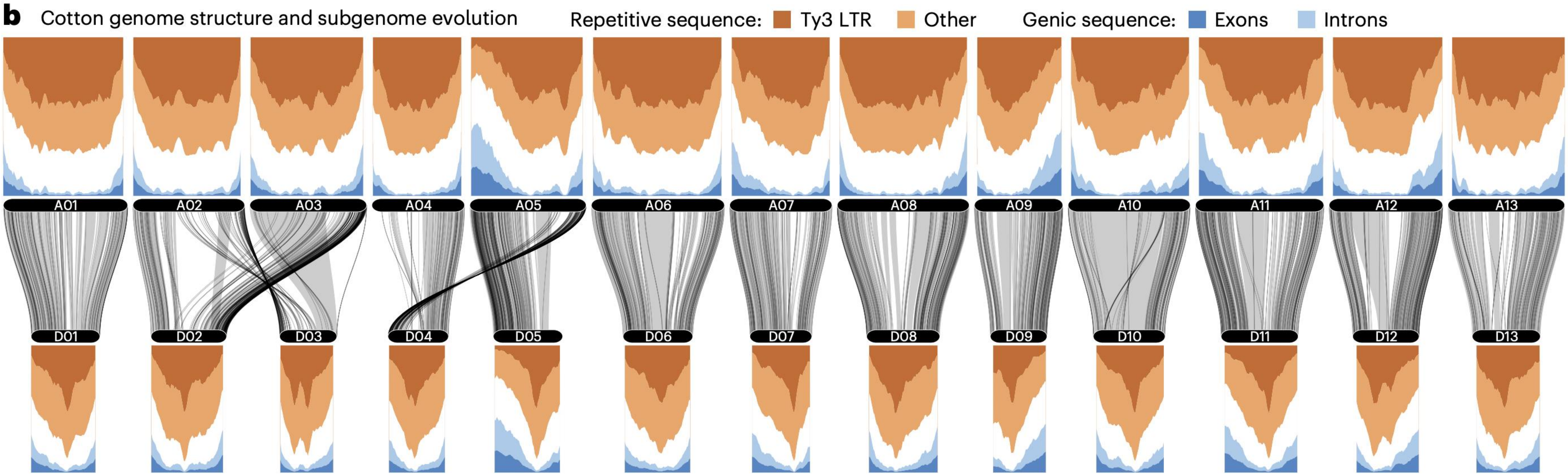
Domestication and improvement

American upland cotton landrace

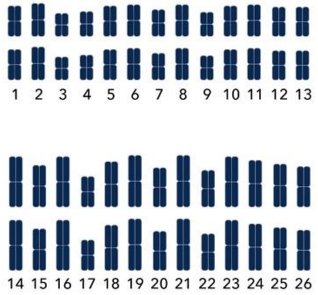
Worldwide introduction and improvement

Modern upland cotton

Structure of the tetraploid cotton genomes

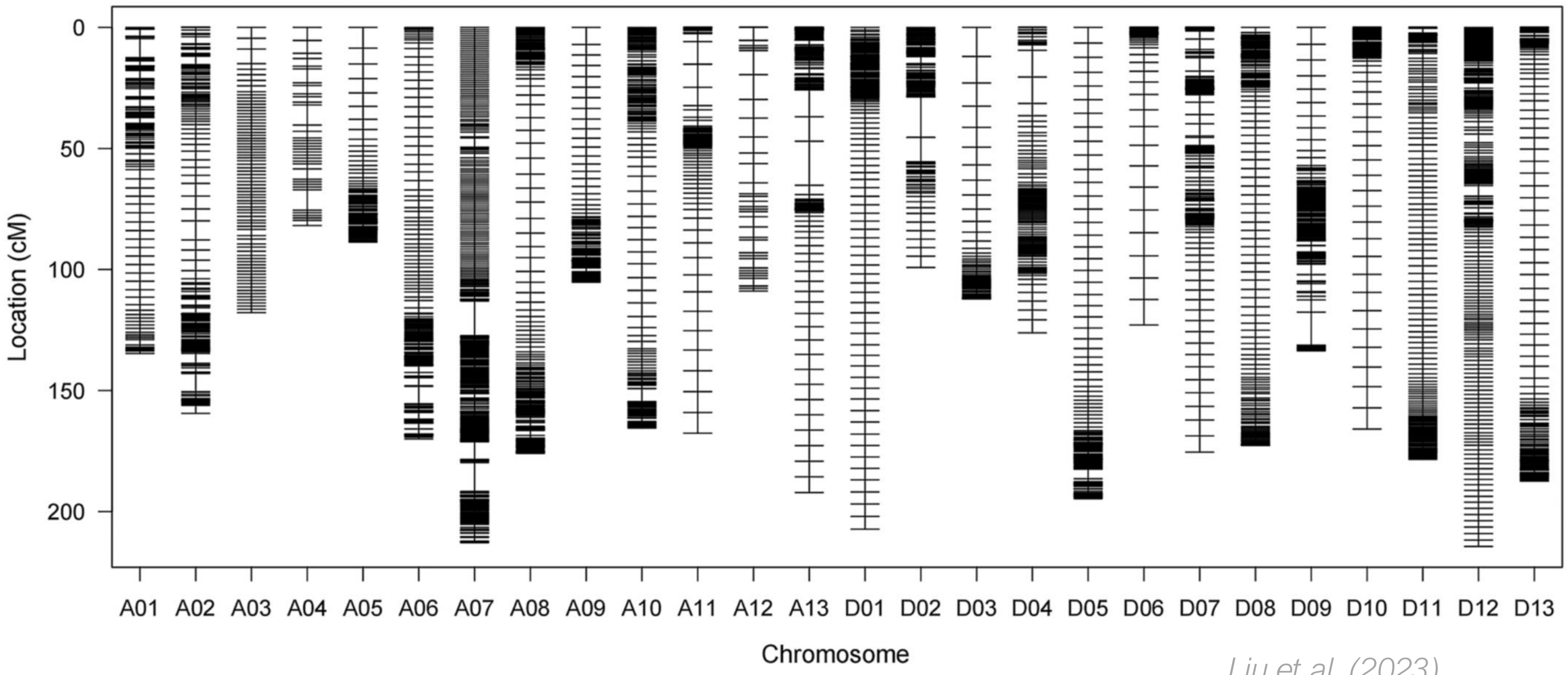


Sreedasyam et al. (2024)



Cotton genomes: evolution and quality

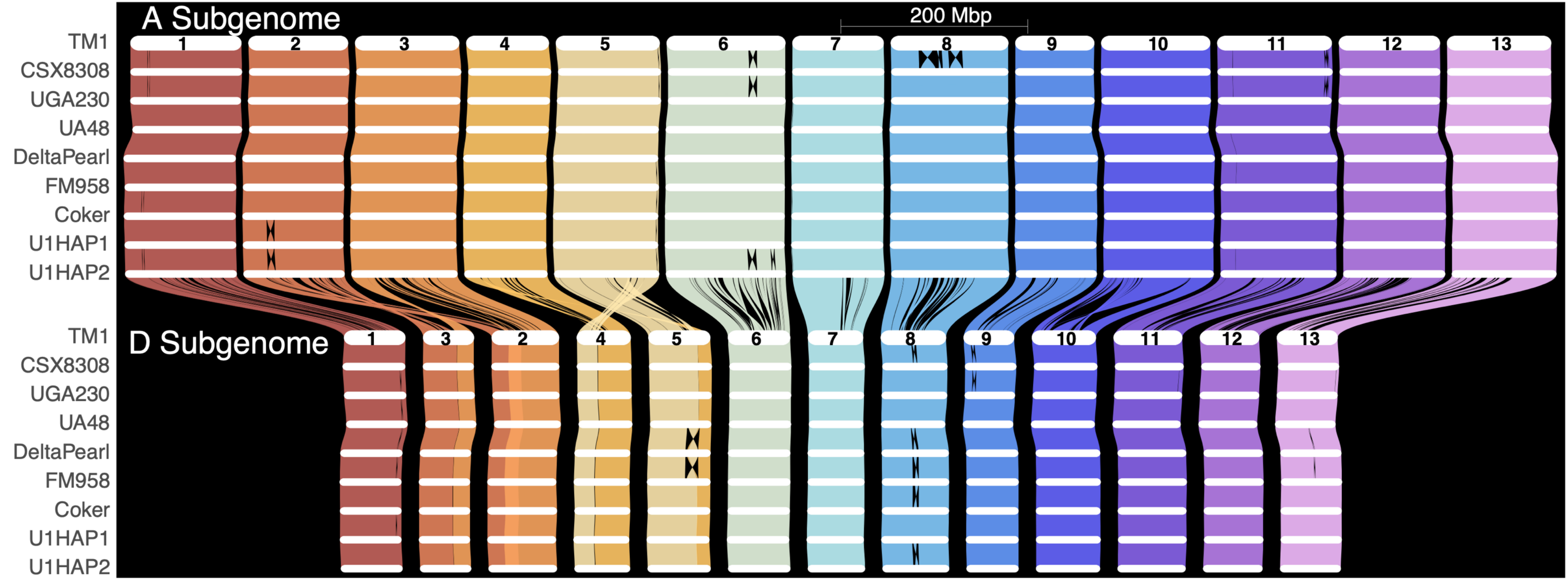
HUDSONALPHA GENOME SEQUENCING CENTER



Liu et al. (2023)

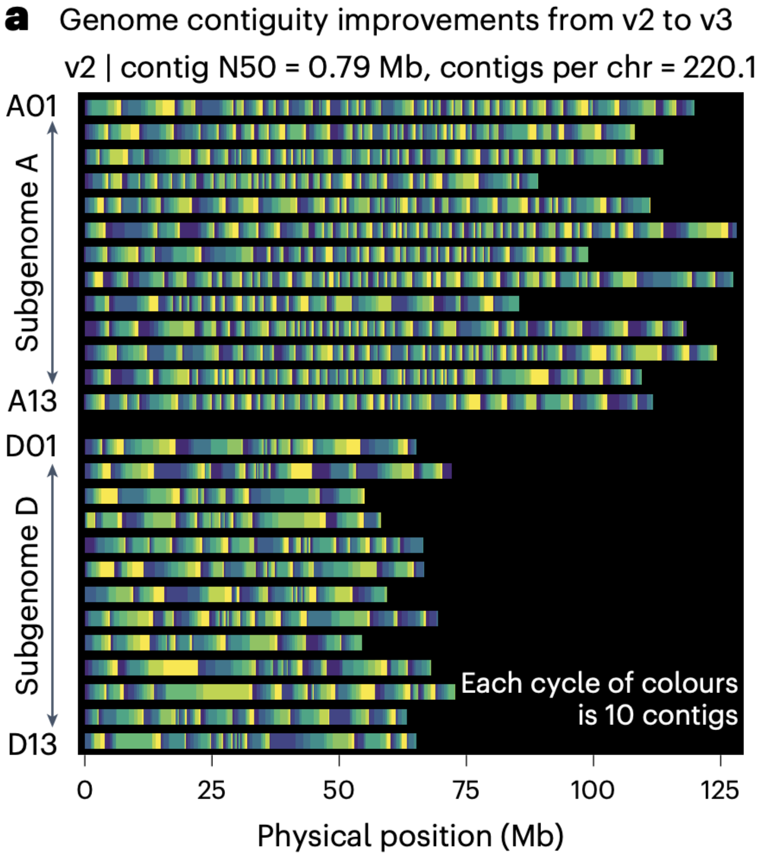
The emerging cotton “pan-genome”

HUDSONALPHA
GENOME SEQUENCING
CENTER



Chromosomes scaled by physical position

Cotton genomes: evolution and quality



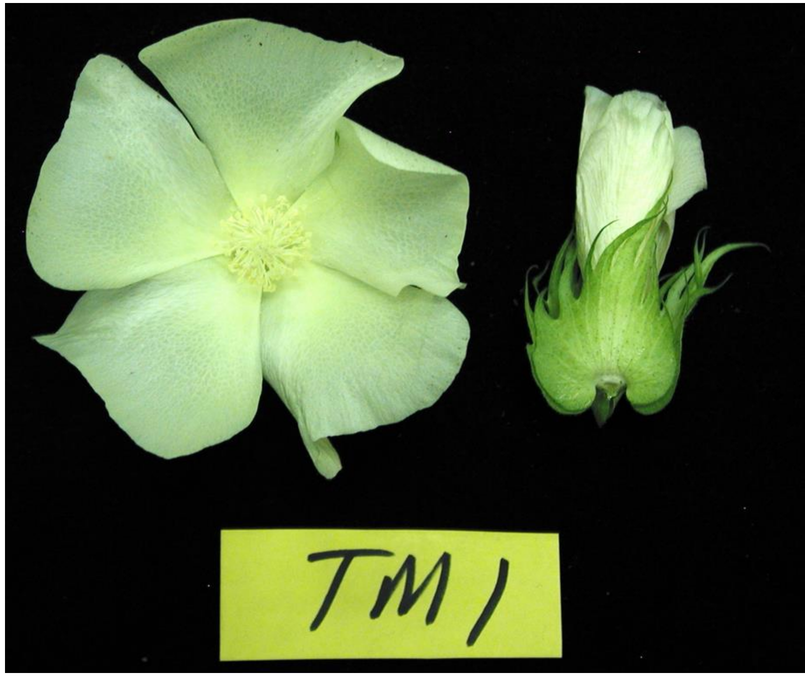
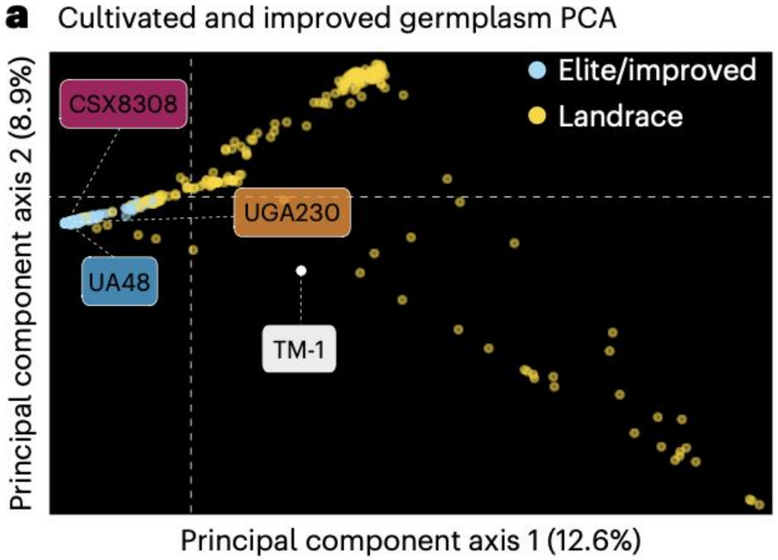
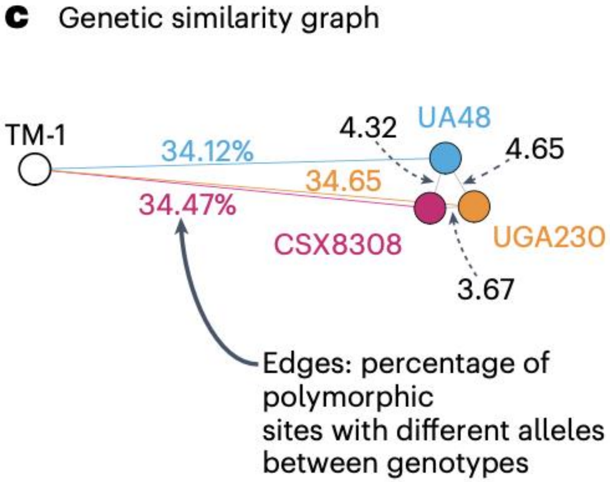


Photo courtesy of Z.J. Chen

Cotton genomes: evolution and quality

HUDSONALPHA GENOME SEQUENCING CENTER

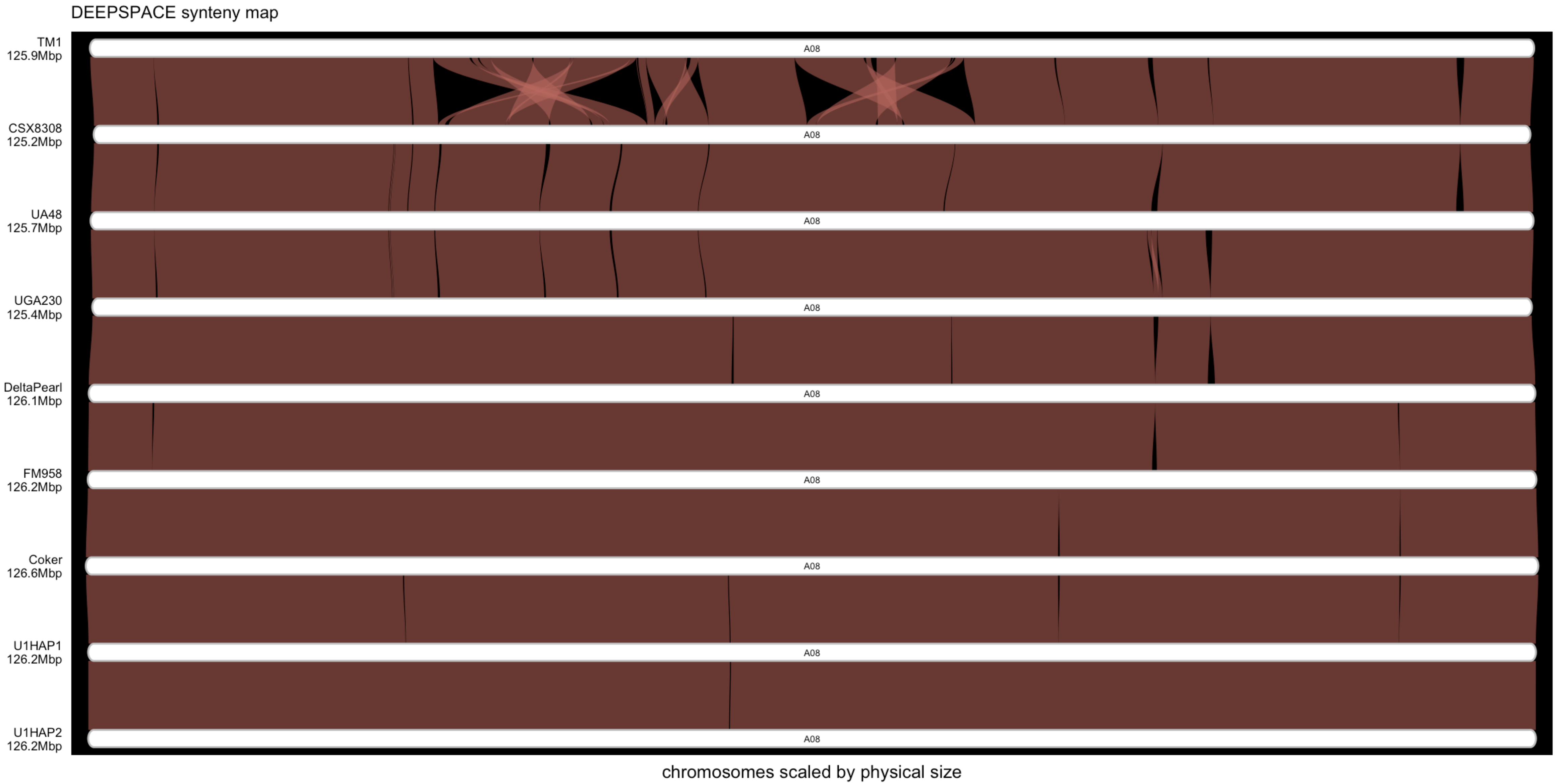




Sreedasyam et al. (2024)

Cotton genomes: evolution and quality

HUDSONALPHA GENOME SEQUENCING CENTER



Cotton genomes: evolution and quality

HUDSONALPHA
GENOME SEQUENCING
CENTER


G3 Genes | Genomes | Genetics
GSA Genetics Society of America
Issues More Content ▾ Series ▾ Submit ▾ Alerts About ▾

Article Navigation

JOURNAL ARTICLE ACCEPTED MANUSCRIPT

Nematode-resistance loci in Upland cotton genomes are associated with structural differences

Zachary P Cohen ✉, Lindsey C Perkin, Tanya A Wagner, Jinggao Liu, Alois A Bell, Mark A Arick II, Corrinne E Grover, John Z Yu, Joshua A Udall, Charles P -C Suh

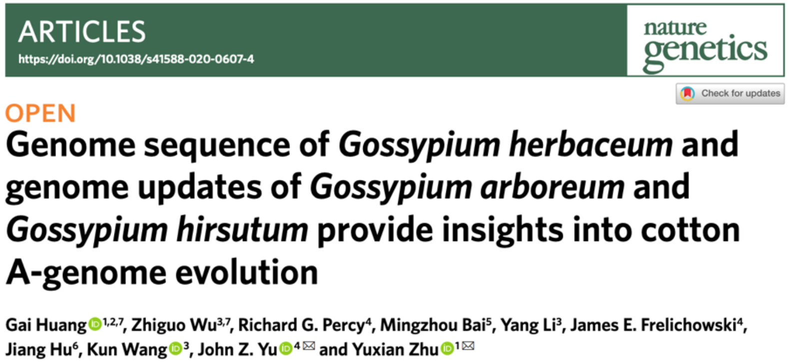

nature plants

Resource https://doi.org/10.1038/s41477-024-01713-z

Genome resources for three modern cotton lines guide future breeding efforts

Received: 27 October 2023
Accepted: 27 April 2024
Published online: 30 May 2024
Check for updates

Avinash Sreedasyam[1,2,17] ✉, John T. Lovell[1,2,17], Sujan Mamidi[1], Sameer Khanal[3], Jerry W. Jenkins[1], Christopher Plott[1], Kempton B. Bryan[4], Zhigang Li[4], Shengqiang Shu[2], Joseph Carlson[2], David Goodstein[2], Luis De Santiago[5], Ryan C. Kirkbride[5], Sebastian Calleja[6], Todd Campbell[7], Jenny C. Koebernick[8], Jane K. Dever[9,16], Jodi A. Scheffler[10], Duke Pauli[6], Johnie N. Jenkins[11], Jack C. McCarty[11], Melissa Williams[1], LoriBeth Boston[1], Jenell Webber[1], Joshua A. Udall[12], Z. Jeffrey Chen[5], Fred Bourland[13], Warwick N. Stiller[14], Christopher A. Saski[4], Jane Grimwood[1], Peng W. Chee[3], Don C. Jones[15] & Jeremy Schmutz[1,2] ✉


ARTICLES
https://doi.org/10.1038/s41588-020-0607-4
nature genetics
Check for updates

OPEN

Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution

Gai Huang[1,2,7], Zhiguo Wu[3,7], Richard G. Percy[4], Mingzhou Bai[5], Yang Li[3], James E. Frelichowski[4], Jiang Hu[6], Kun Wang[3], John Z. Yu[4] ✉ and Yuxian Zhu[1] ✉

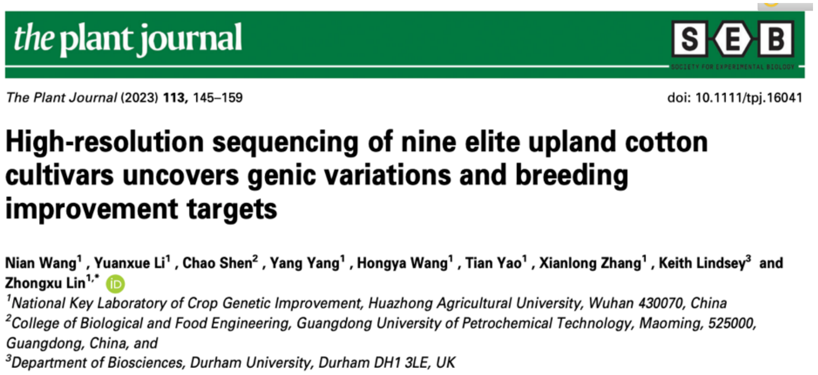

G3 Genes | Genomes | Genetics
GSA Genetics Society of America
Issues More Content ▾ Series ▾ Submit ▾ Alerts About ▾

Article Navigation

JOURNAL ARTICLE

Genome assembly of two nematode-resistant cotton lines (*Gossypium hirsutum* L.)

Lindsey C Perkin, Al Bell, Lori L Hinze, Charles P -C Suh, Mark A Arick, II, Daniel G Peterson, Joshua A Udall ✉


the plant journal SEB

The Plant Journal (2023) 113, 145–159 doi: 10.1111/tpj.16041

High-resolution sequencing of nine elite upland cotton cultivars uncovers genic variations and breeding improvement targets

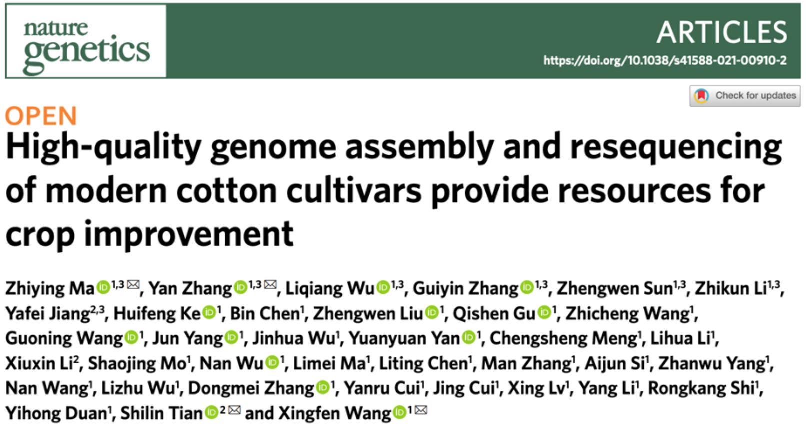
Nian Wang[1], Yuanxue Li[1], Chao Shen[2], Yang Yang[1], Hongya Wang[1], Tian Yao[1], Xianlong Zhang[1], Keith Lindsey[3] and Zhongxu Lin[1,*]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China
[2]College of Biological and Food Engineering, Guangdong University of Petrochemical Technology, Maoming, 525000, Guangdong, China, and
[3]Department of Biosciences, Durham University, Durham DH1 3LE, UK


PNAS RESEARCH ARTICLE | PLANT BIOLOGY OPEN ACCESS Check for updates

Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons

Renhai Peng[a,1], Yanchao Xu[b,1], Shilin Tian[c,1], Turgay Unver[d,1], Zhen Liu[a,1], Zhongli Zhou[b,1], Xiaoyan Cai[b,1], Kunbo Wang[b], Yangyang Wei[a], Yuling Liu[a], Heng Wang[b], Guanjing Hu[b,e], Zhongren Zhang[c], Corrinne E. Grover[f], Yuqing Hou[b], Yuhong Wang[b], Pengtao Li[a], Tao Wang[a], Quanwei Lu[a], Yuanyuan Wang[h], Justin L. Conover[f], Hassan Ghazal[g], Qinglian Wang[h], Baohong Zhang[i,2], Marc Van Montagu[j,k,2], Yves Van de Peer[j,k,l,m,2], Jonathan F. Wendel[f,2], and Fang Liu[b,2]

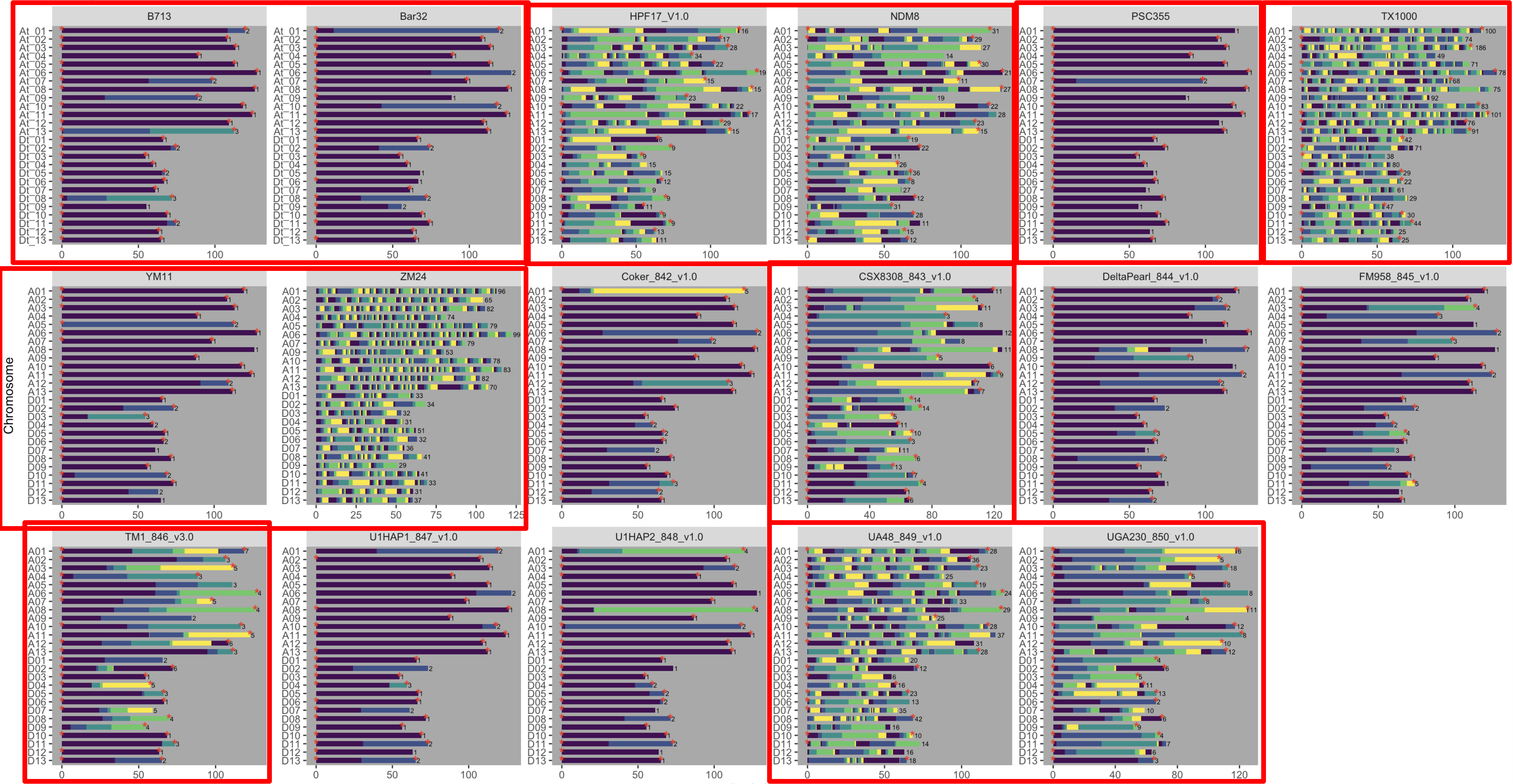

nature genetics
ARTICLES
https://doi.org/10.1038/s41588-021-00910-2
Check for updates

OPEN

High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement

Zhiying Ma[1,3] ✉, Yan Zhang[1,3] ✉, Liqiang Wu[1,3], Guiyin Zhang[1,3], Zhengwen Sun[1,3], Zhikun Li[1,3], Yafei Jiang[2,3], Huifeng Ke[1], Bin Chen[1], Zhengwen Liu[1], Qishen Gu[1], Zhicheng Wang[1], Guoning Wang[1], Jun Yang[1], Jinhua Wu[1], Yuanyuan Yan[1], Chengsheng Meng[1], Lihua Li[1], Xiuxin Li[2], Shaojing Mo[1], Nan Wu[1], Limei Ma[1], Liting Chen[1], Man Zhang[1], Aijun Si[1], Zhanwu Yang[1], Nan Wang[1], Lizhu Wu[1], Dongmei Zhang[1], Yanru Cui[1], Jing Cui[1], Xing Lv[1], Yang Li[1], Rongkang Shi[1], Yihong Duan[1], Shilin Tian[2] ✉ and Xingfen Wang[1] ✉

Contig positions: each cycle through colors is 5 contigs (4 gaps)



Physical position (Mb); * indicate telomere sequence

TM-1

Z. Jeffrey Chen
UT Austin

Coker, U1

Christopher A. Saski
Clemson University

CSX8308

Warwick N. Stiller
CSIRO, Australia

UGA230

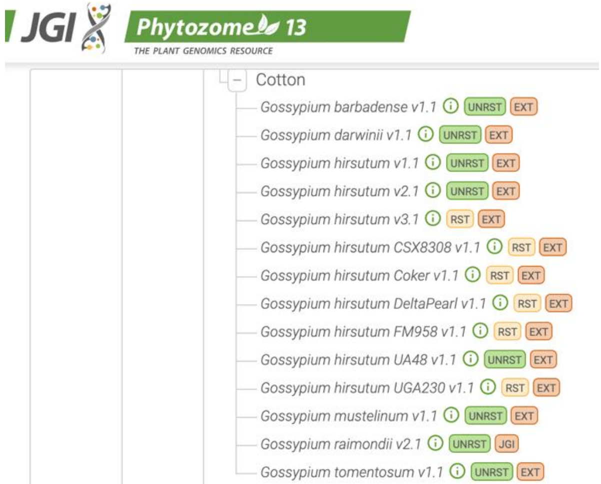
Peng W. Chee
U of Georgia, Tifton

UA48

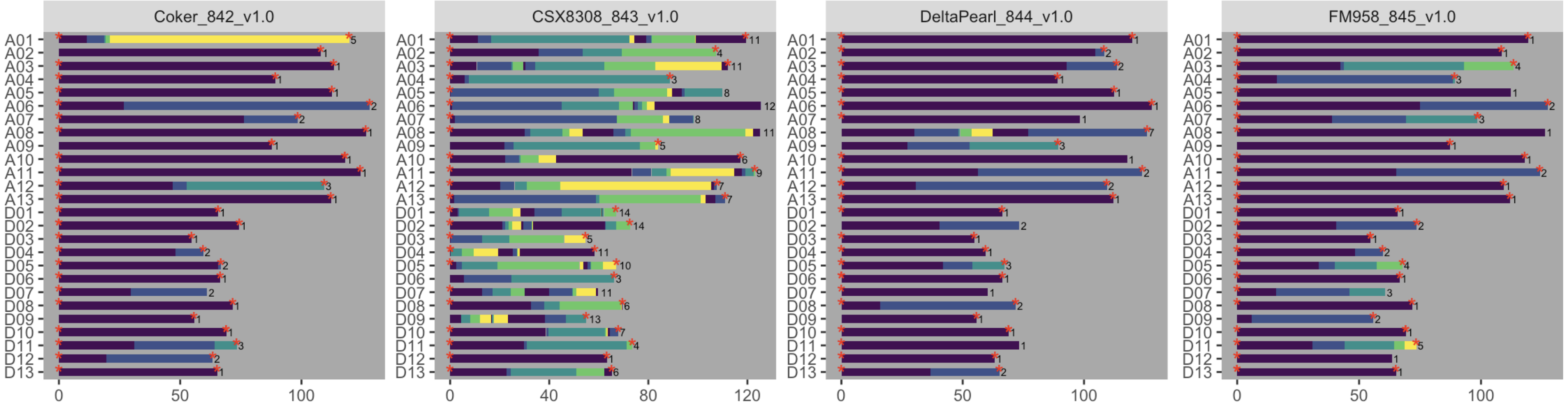
Fred Bourland
U of Arkansas

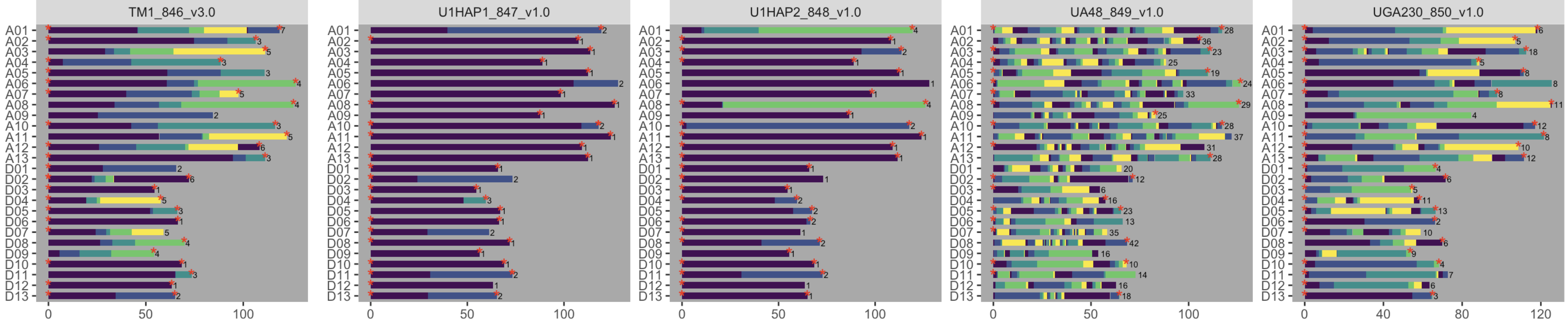
DeltaPearl:
Bayer

FM958:
BASF


JGI Phytozome 13
THE PLANT GENOMICS RESOURCE

Cotton
- *Gossypium barbadense v1.1* UNRST EXT
- *Gossypium darwinii v1.1* UNRST EXT
- *Gossypium hirsutum v1.1* UNRST EXT
- *Gossypium hirsutum v2.1* UNRST EXT
- *Gossypium hirsutum v3.1* RST EXT
- *Gossypium hirsutum CSX8308 v1.1* RST EXT
- *Gossypium hirsutum Coker v1.1* RST EXT
- *Gossypium hirsutum DeltaPearl v1.1* RST EXT
- *Gossypium hirsutum FM958 v1.1* RST EXT
- *Gossypium hirsutum UA48 v1.1* UNRST EXT
- *Gossypium hirsutum UGA230 v1.1* RST EXT
- *Gossypium mustelinum v1.1* UNRST EXT
- *Gossypium raimondii v2.1* UNRST JGI
- *Gossypium tomentosum v1.1* UNRST EXT

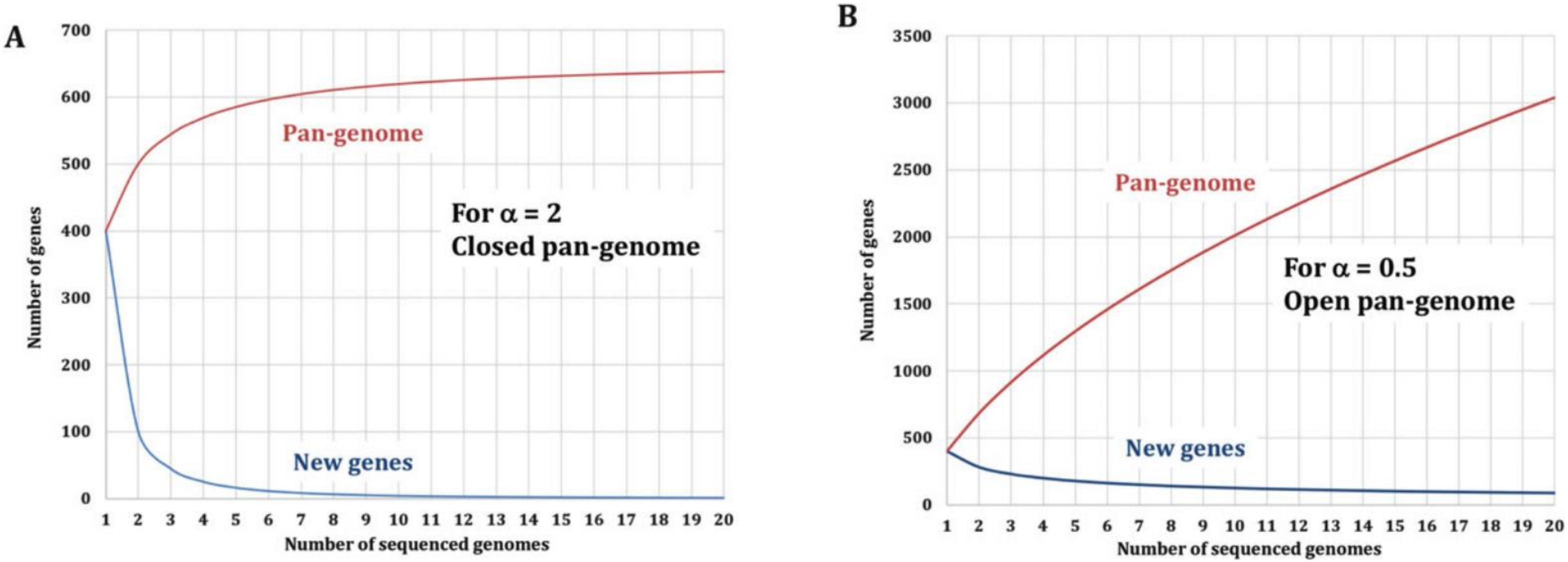




Physical position (Mb); * indicate telomere sequence

Why we need integrated pan-genome resources

- The idea of “open” vs. “closed” pangenomes comes from work in bacteria
  - Each lineage has relatively few genes
  - Huge influence of new genes from HGT

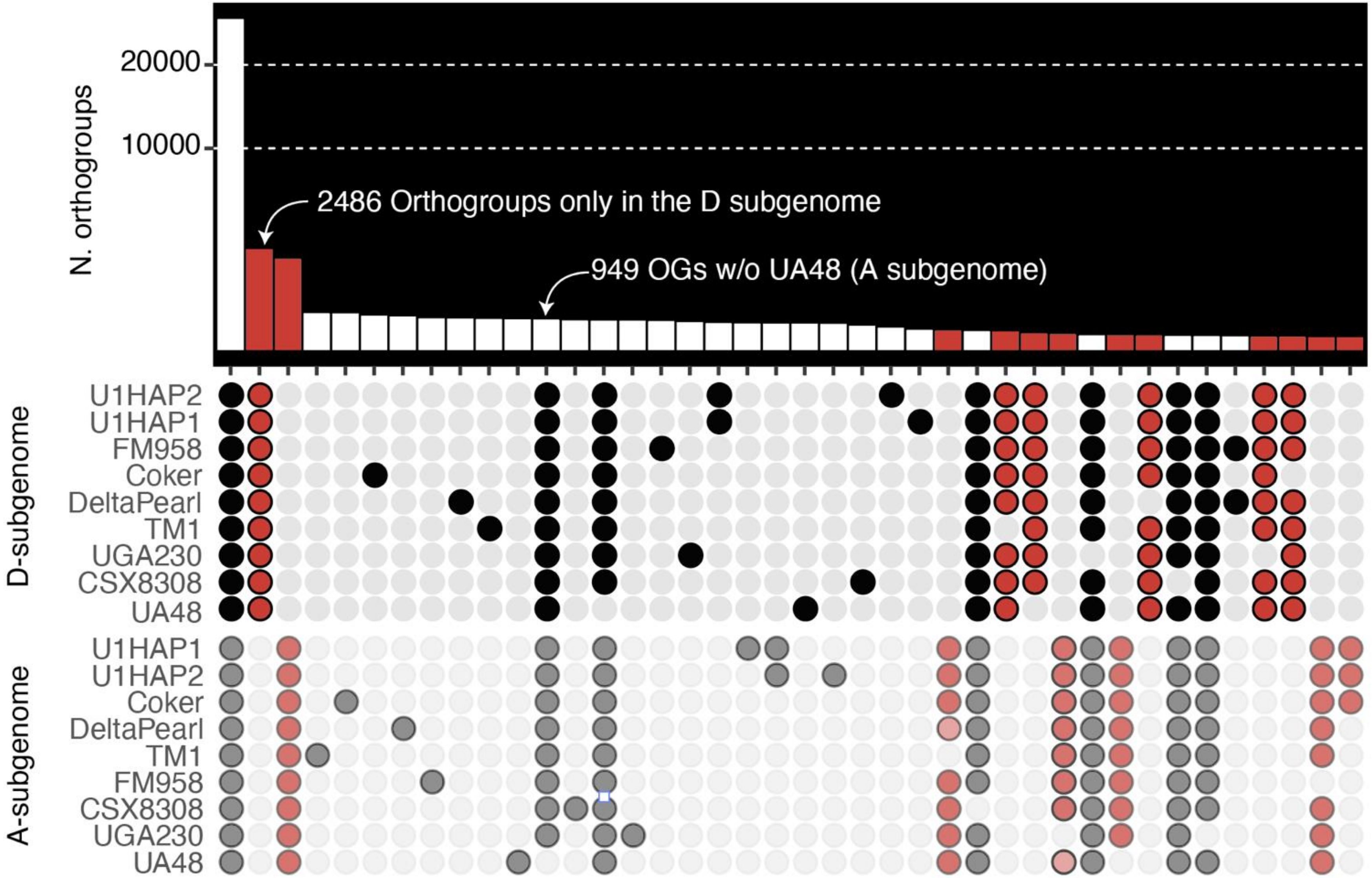


Richard (2020)

Why we need integrated pan-genome resources

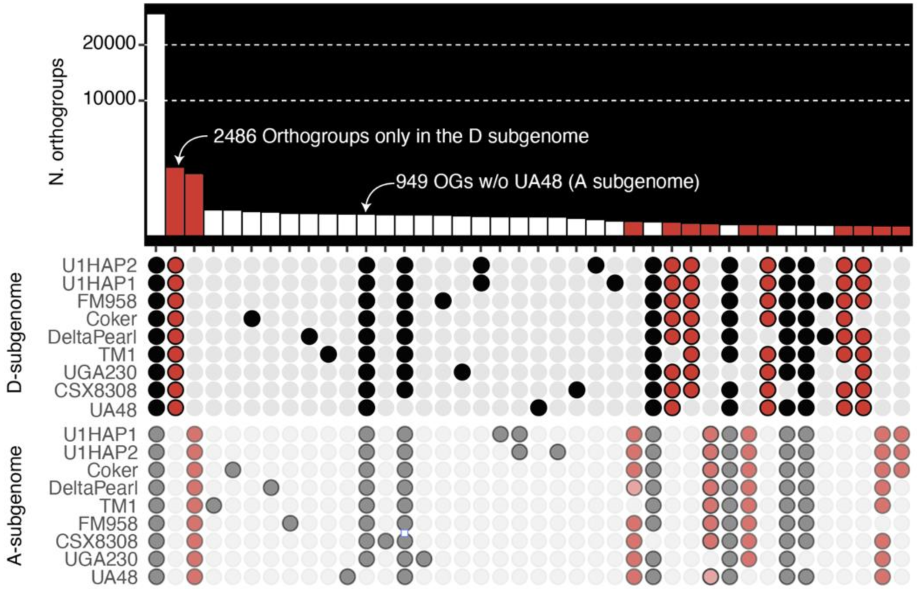
- The idea of “open” vs. “closed” pangenomes comes from work in bacteria
- In plants, genes come and go less frequently
  - But more frequently in polyploids

Why we need integrated pan-genome resources



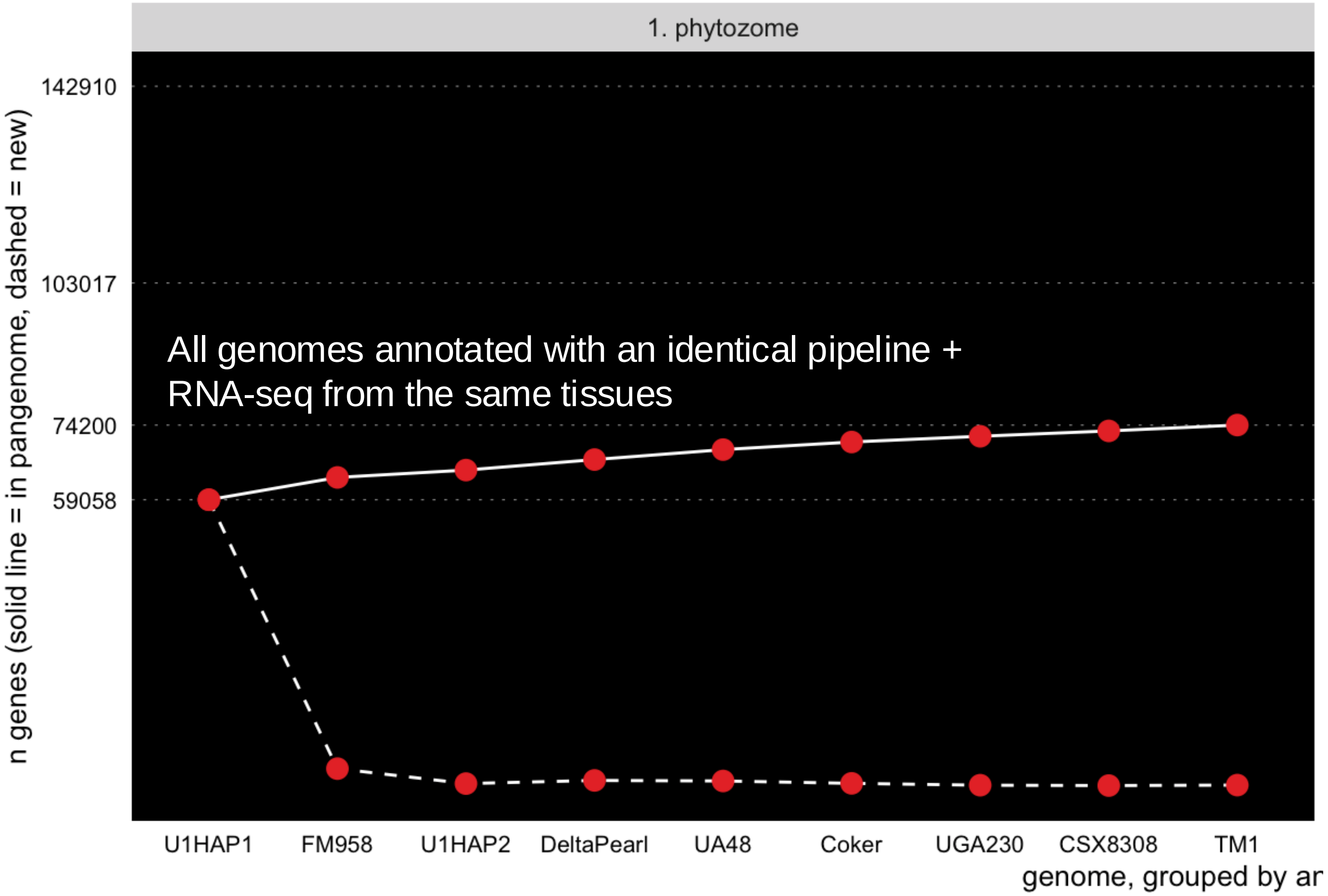
Why we need integrated pan-genome resources

- The idea of “open” vs. “closed” pangenomes comes from work in bacteria
- In plants, genes come and go less frequently
  - But more frequently in polyploids
  - So, species with whole-genome duplications (like cotton) might have more PAV than diploid genomes
  - *But … there might be other causes*

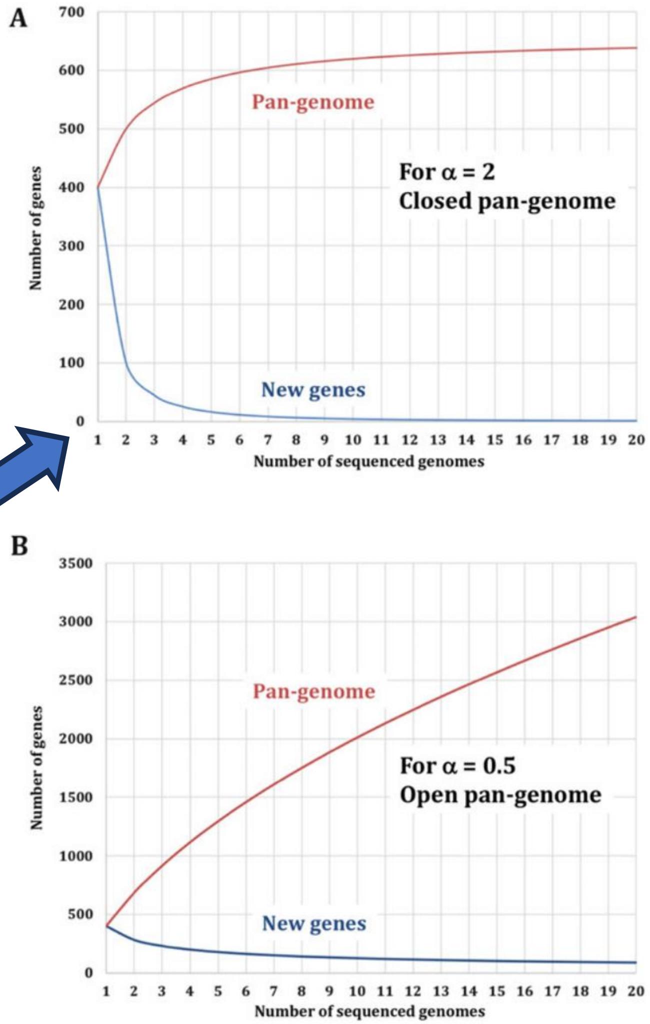


Why we need integrated pan-genome resources

- Causes of gene presence-absence variation (PAV) can be methodological



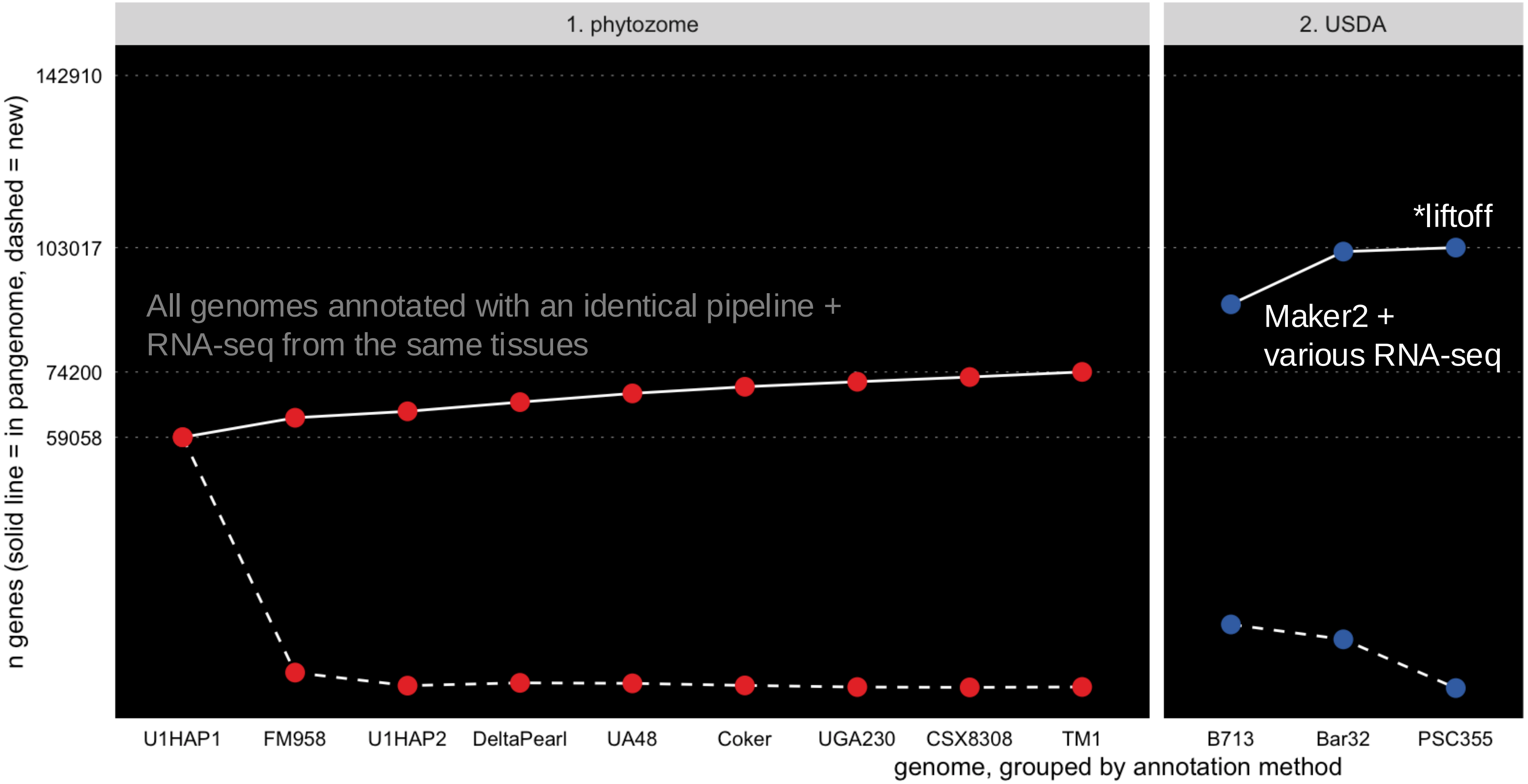
All genomes annotated with an identical pipeline + RNA-seq from the same tissues



Why we need integrated pan-genome resources

HUDSONALPHA GENOME SEQUENCING CENTER

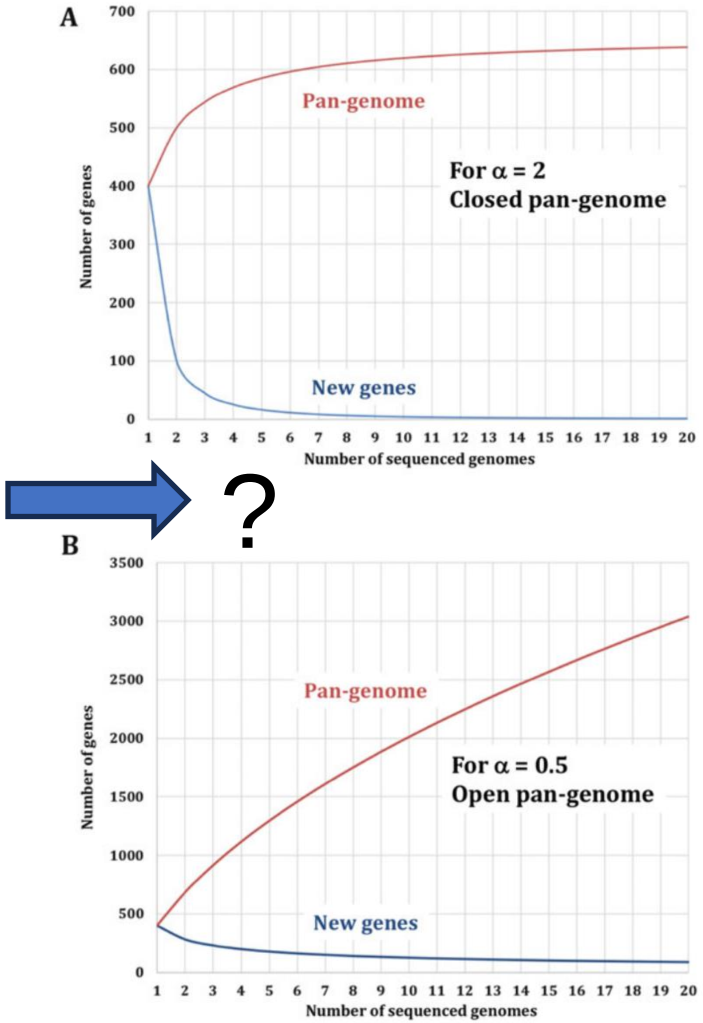
- Causes of gene presence-absence variation (PAV) can be methodological



All genomes annotated with an identical pipeline + RNA-seq from the same tissues

*liftoff

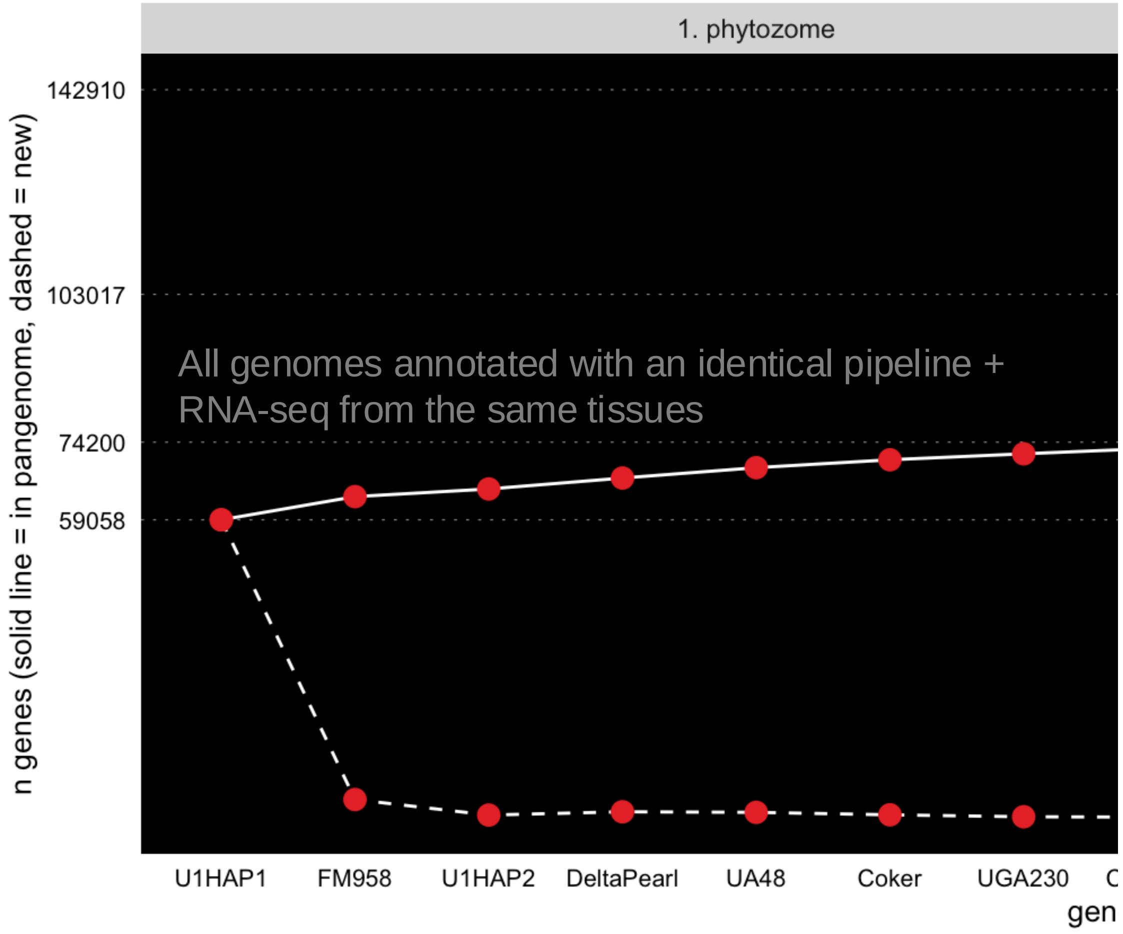
Maker2 + various RNA-seq



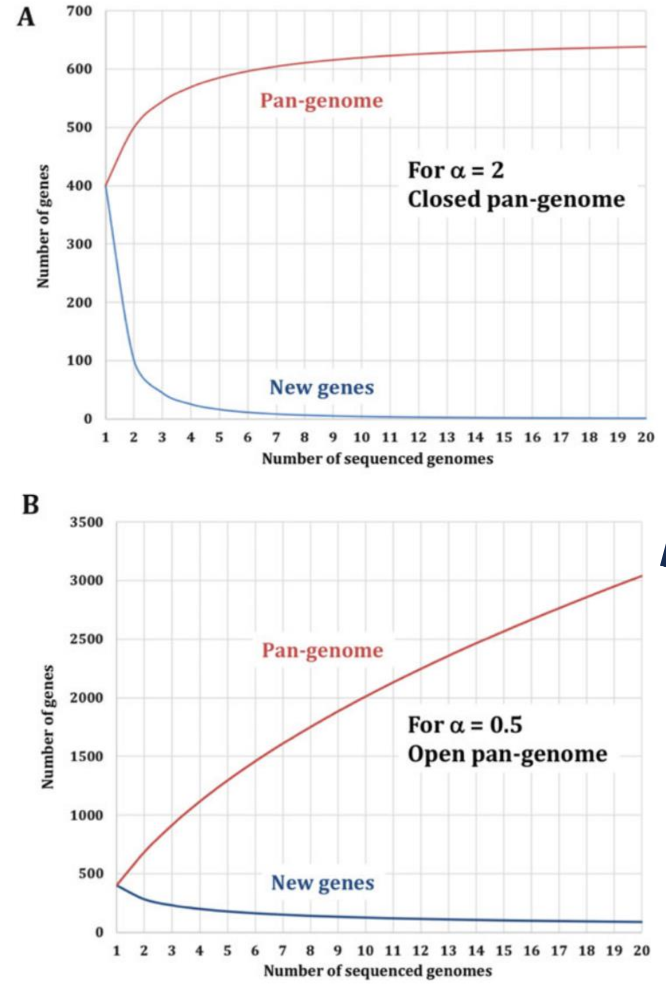
?

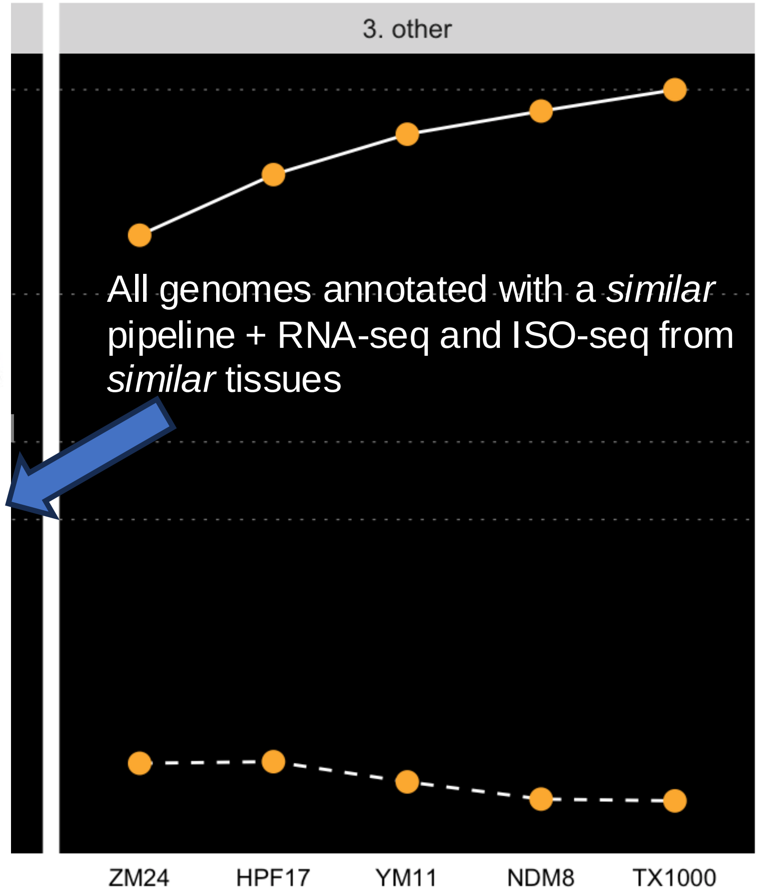
Why we need integrated pan-genome resources

- Causes of gene presence-absence variation (PAV) can be methodological



All genomes annotated with an identical pipeline + RNA-seq from the same tissues



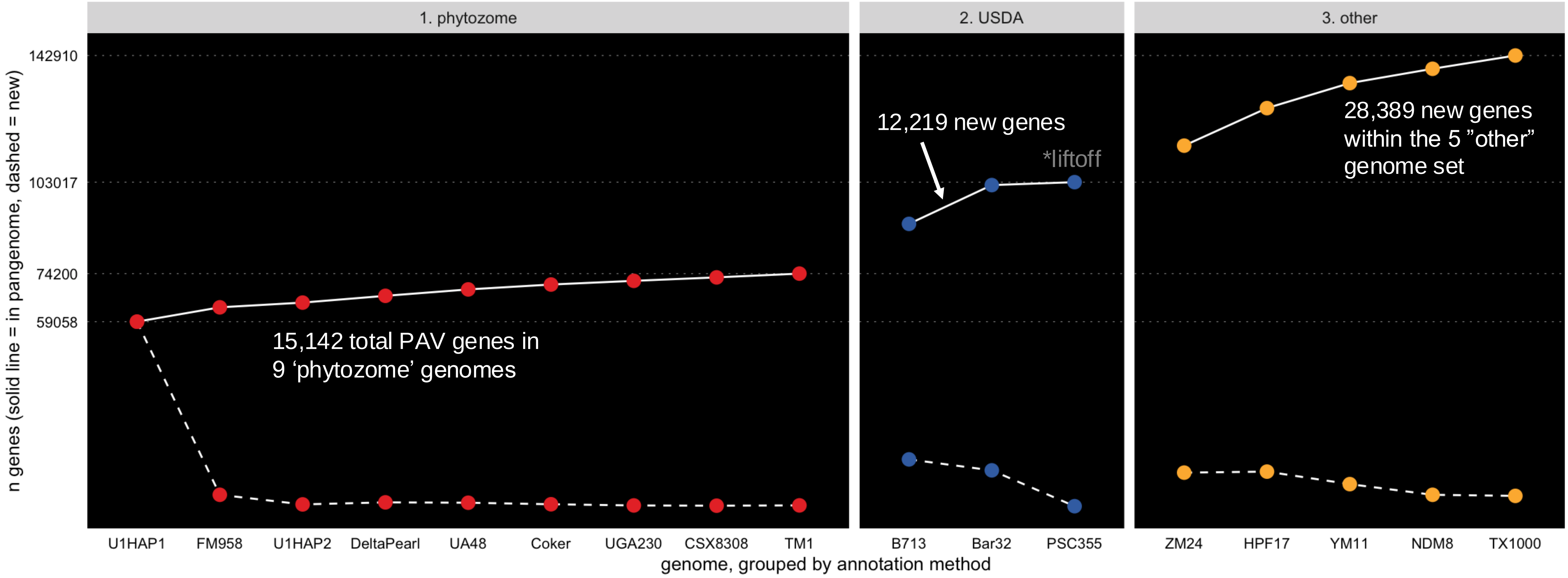


All genomes annotated with a *similar* pipeline + RNA-seq and ISO-seq from *similar* tissues

Why we need integrated pan-genome resources

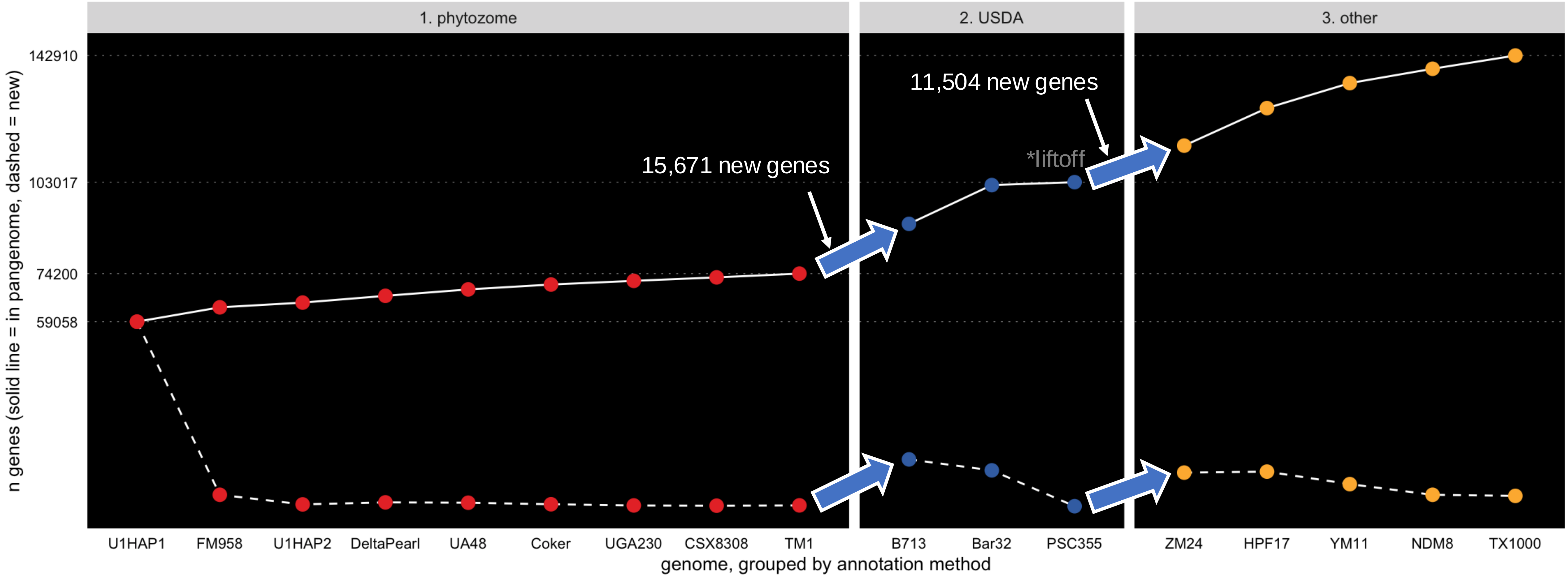
HUDSONALPHA
GENOME SEQUENCING
CENTER

- Causes of gene presence-absence variation (PAV) can be methodological



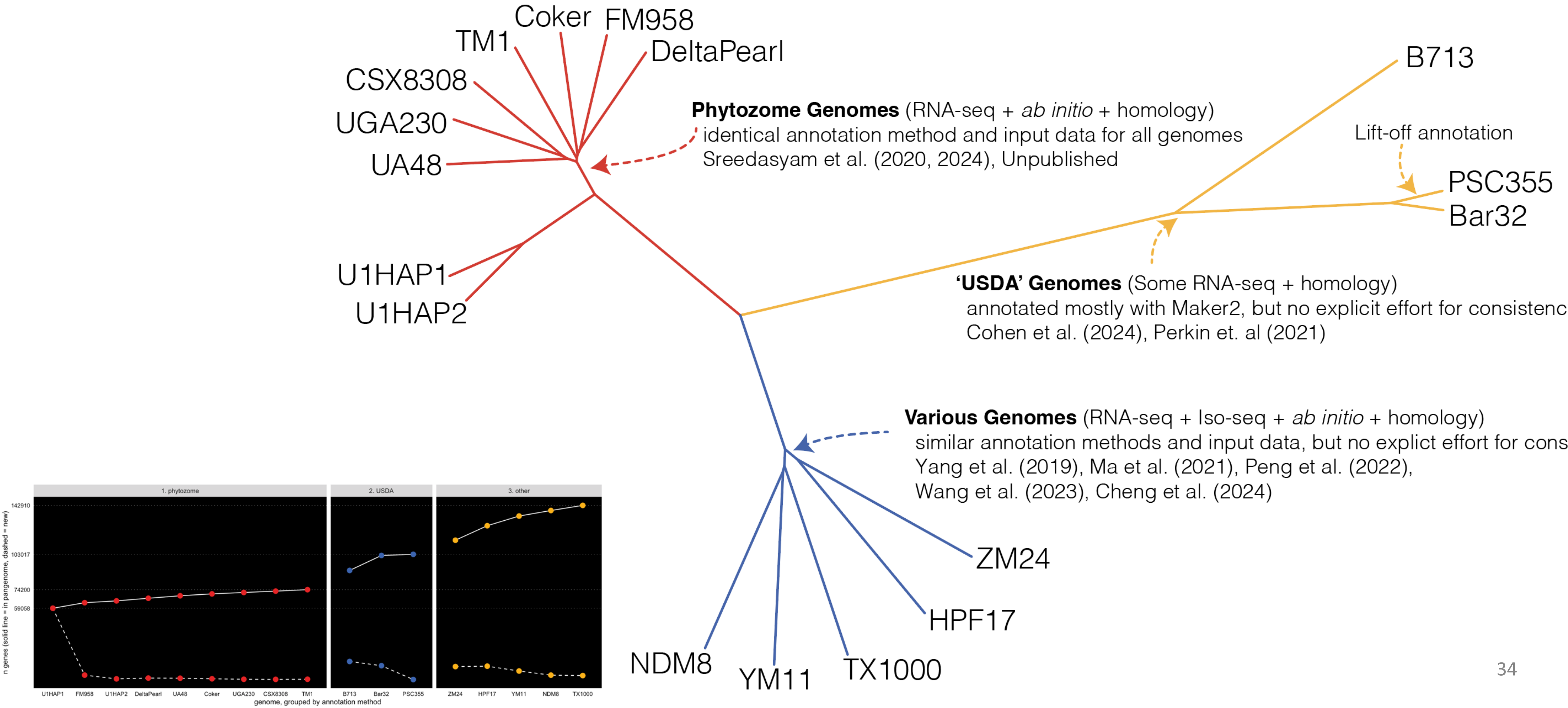
Why we need integrated pan-genome resources

- Causes of gene presence-absence variation (PAV) can be methodological



Why we need integrated pan-genome resources

Distance Tree based on PAV Similarity



Phytozome Genomes (RNA-seq + *ab initio* + homology)
identical annotation method and input data for all genomes
Sreedasyam et al. (2020, 2024), Unpublished

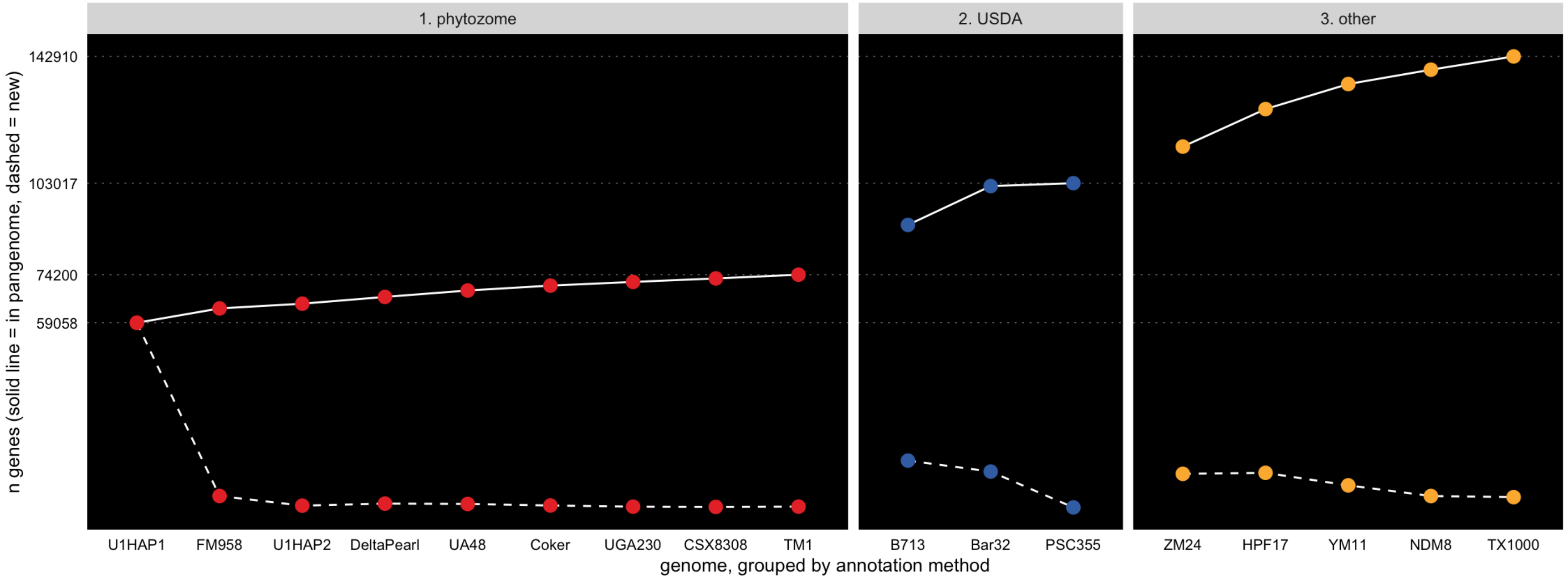
Lift-off annotation

'USDA' Genomes (Some RNA-seq + homology)
annotated mostly with Maker2, but no explicit effort for consistency
Cohen et al. (2024), Perkin et. al (2021)

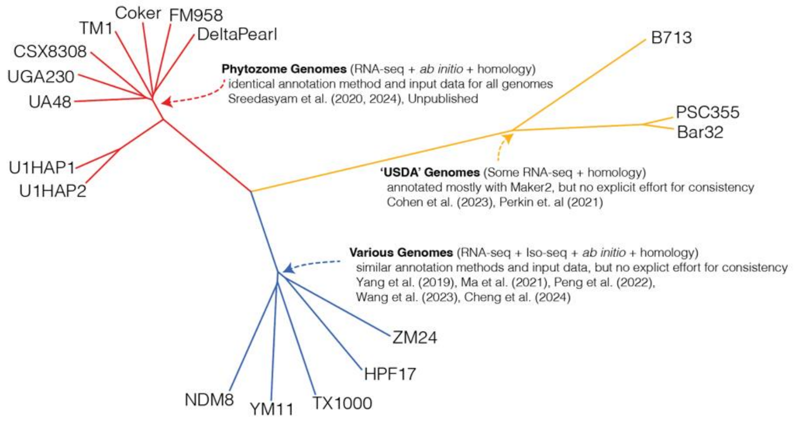
Various Genomes (RNA-seq + Iso-seq + *ab initio* + homology)
similar annotation methods and input data, but no explict effort for consistency
Yang et al. (2019), Ma et al. (2021), Peng et al. (2022),
Wang et al. (2023), Cheng et al. (2024)

Why we need integrated pan-genome resources

- In plants, genes come and go less frequently
- Causes of gene presence-absence variation (PAV) can be methodological
  - “open” pan-genomes might be driven by what we call a gene



Distance Tree based on PAV Similarity



Phytozome Genomes (RNA-seq + *ab initio* + homology)
identical annotation method and input data for all genomes
Sreedasyam et al. (2020, 2024), Unpublished

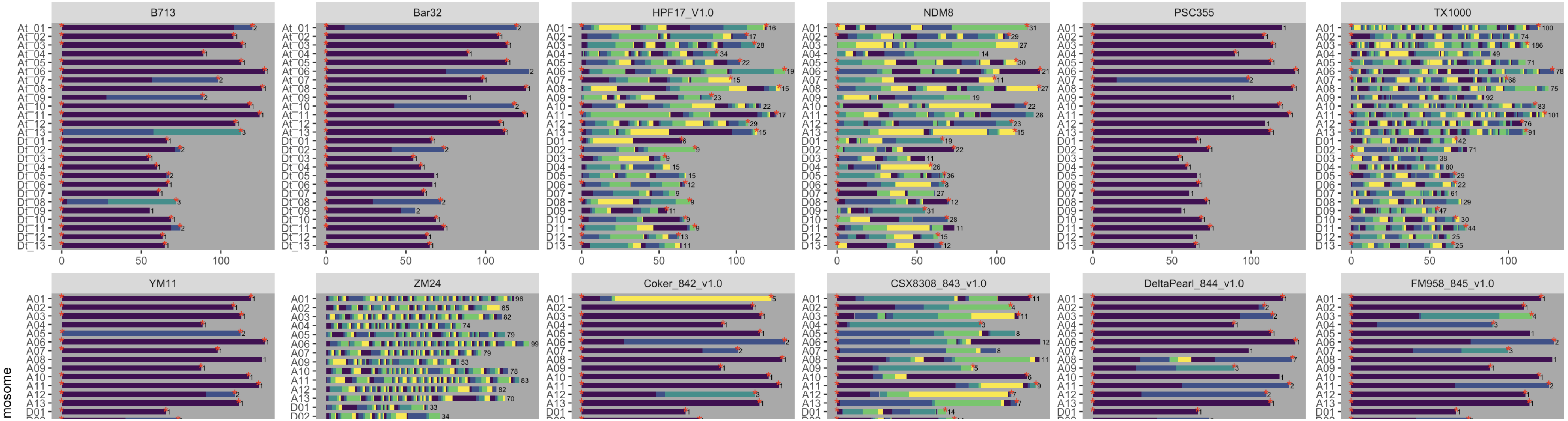
‘USDA’ Genomes (Some RNA-seq + homology)
annotated mostly with Maker2, but no explicit effort for consistency
Cohen et al. (2023), Perkin et. al (2021)

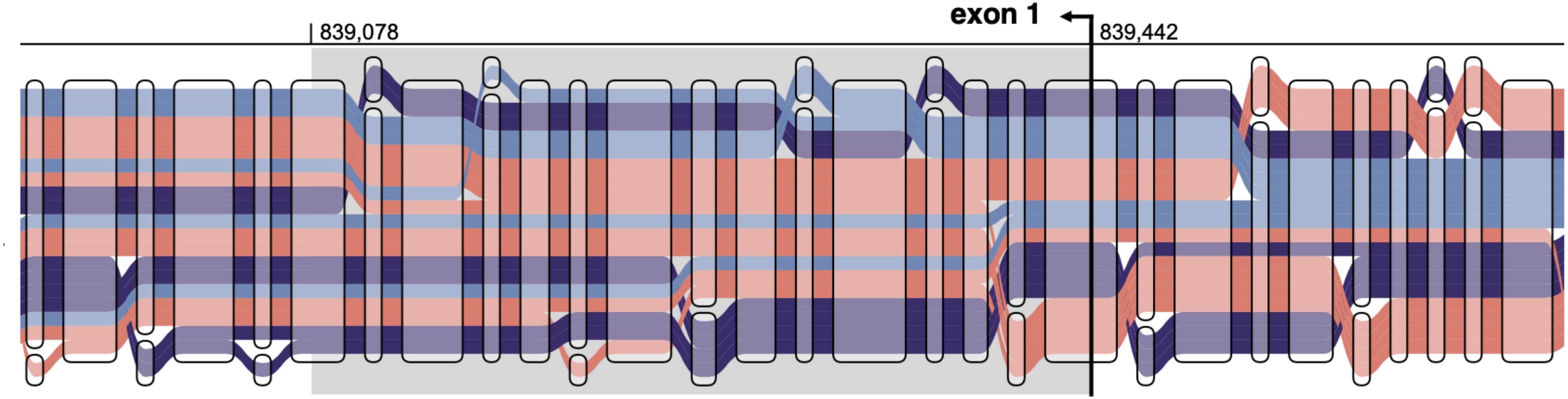
Various Genomes (RNA-seq + Iso-seq + *ab initio* + homology)
similar annotation methods and input data, but no explicit effort for consistency
Yang et al. (2019), Ma et al. (2021), Peng et al. (2022),
Wang et al. (2023), Cheng et al. (2024)

From multiple genomes to a “graph” pangenome

HUDSONALPHA
GENOME SEQUENCING
CENTER

Contig positions: each cycle through colors is 5 contigs (4 gaps)

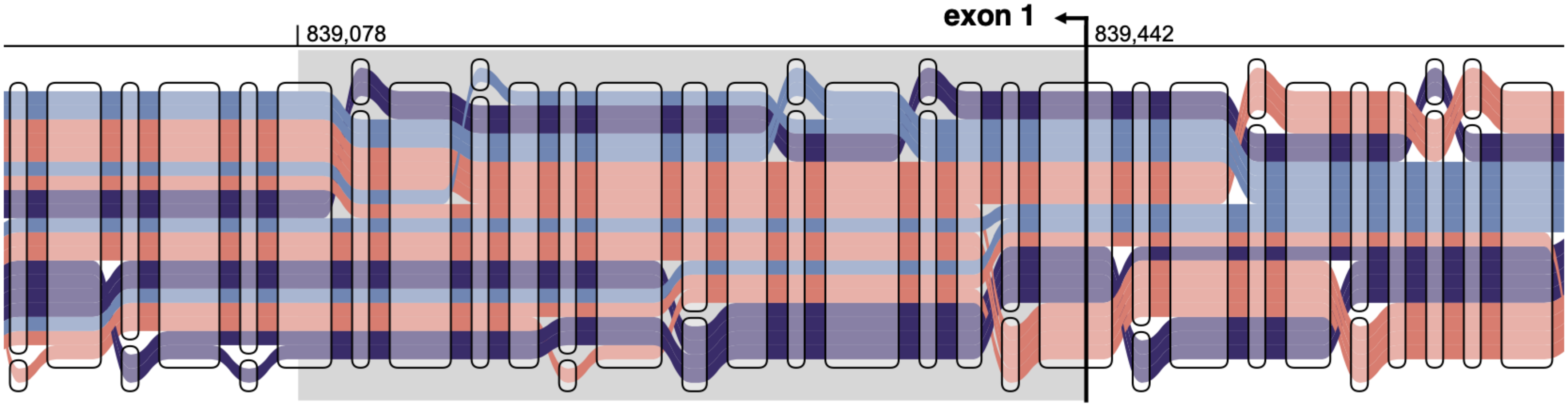




Considerations when constructing a pan-genome

- A pangenome graph is a static resource: all reads need to be re-mapped and analyses done from scratch if you want to add a new genome
  - Our group can do this … but, first the cotton community needs to come to a consensus about what genomes to include

-

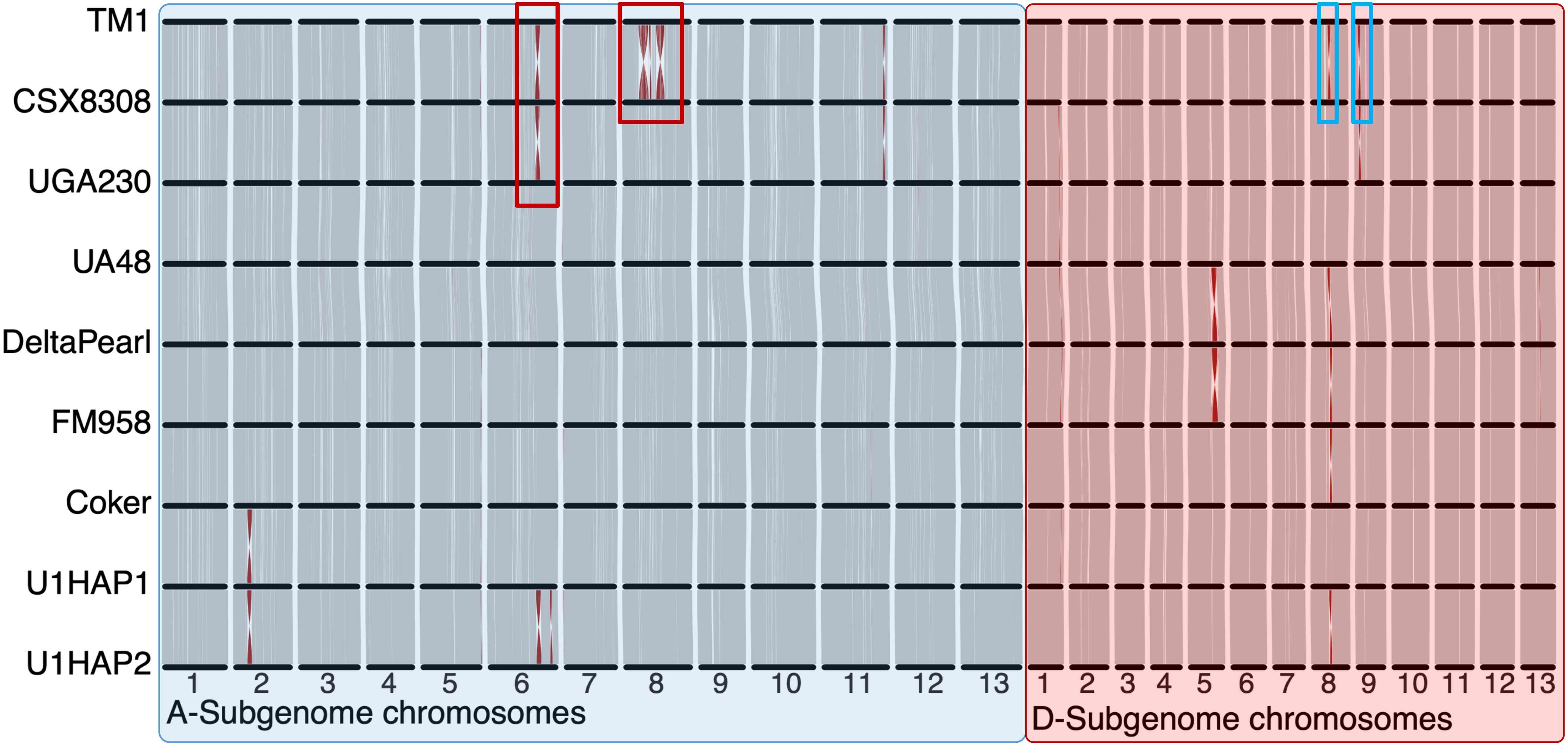


What can we learn from a ‘pangenome’?

- Our goal is to make the pan-genome *useful*
- What deliverables can we extract from the pangenome that are directly usable by breeders?
  - Large structural variants

What can we learn from a 'pangenome'?

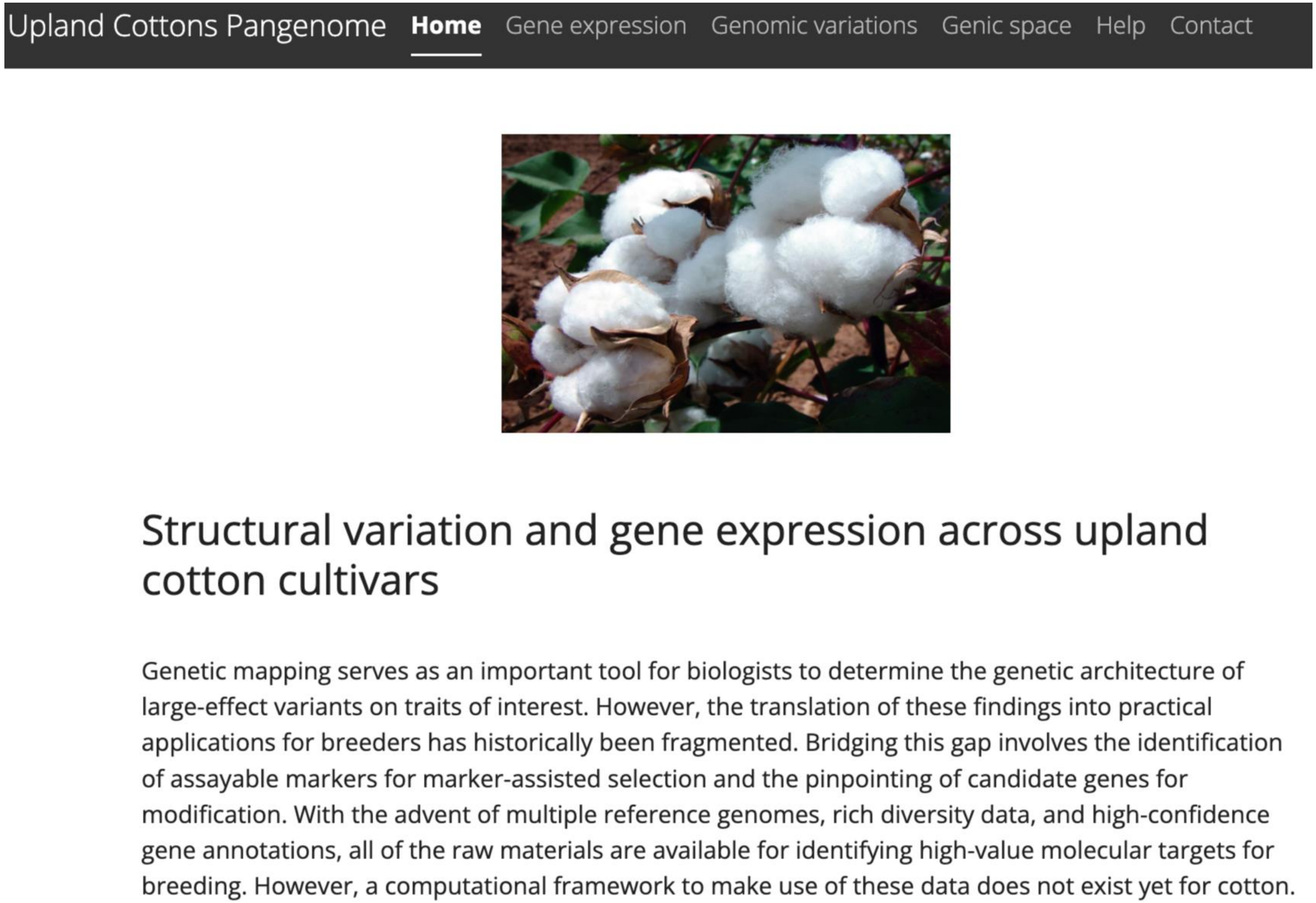
HUDSONALPHA
GENOME SEQUENCING
CENTER



What can we learn from a ‘pangenome’?

- Our goal is to make the pan-genome ***useful***
- What deliverables can we extract from the pangenome that are directly usable by breeders?
    - Large structural variants
    - Something that is easily queryable … not just a graph database

What can we learn from a ‘pangenome’?



Structural variation and gene expression across upland cotton cultivars

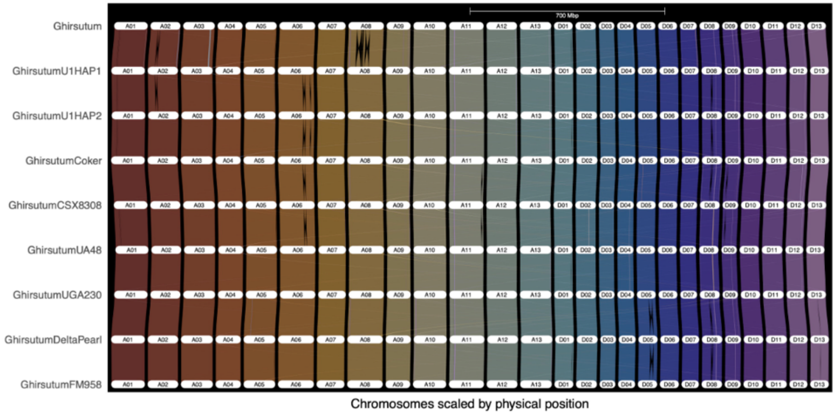
Genetic mapping serves as an important tool for biologists to determine the genetic architecture of large-effect variants on traits of interest. However, the translation of these findings into practical applications for breeders has historically been fragmented. Bridging this gap involves the identification of assayable markers for marker-assisted selection and the pinpointing of candidate genes for modification. With the advent of multiple reference genomes, rich diversity data, and high-confidence gene annotations, all of the raw materials are available for identifying high-value molecular targets for breeding. However, a computational framework to make use of these data does not exist yet for cotton.

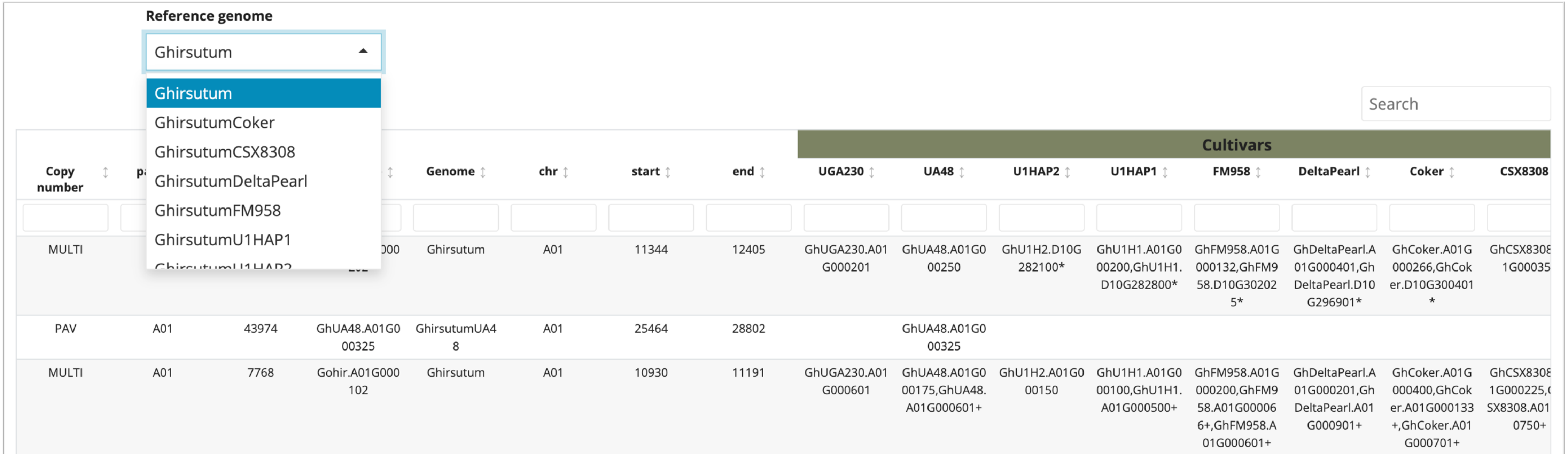
https://tinyurl.com/uplandcottonspangenome

What can we learn from a ‘pangenome’?

- Our goal is to make the pan-genome *useful*
- What deliverables can we extract from the pangenome that are directly usable by breeders?
    - Large structural variants
    - Access to “**pangene-sets**” … the group of genes across genomes that should recombine (be in synteny) and share a common ancestor (orthologs)

Multi-reference gene alignments and PAV





Reference genome

Ghirsutum

- Ghirsutum
- GhirsutumCoker
- GhirsutumCSX8308
- GhirsutumDeltaPearl
- GhirsutumFM958
- GhirsutumU1HAP1
- GhirsutumU1HAP2

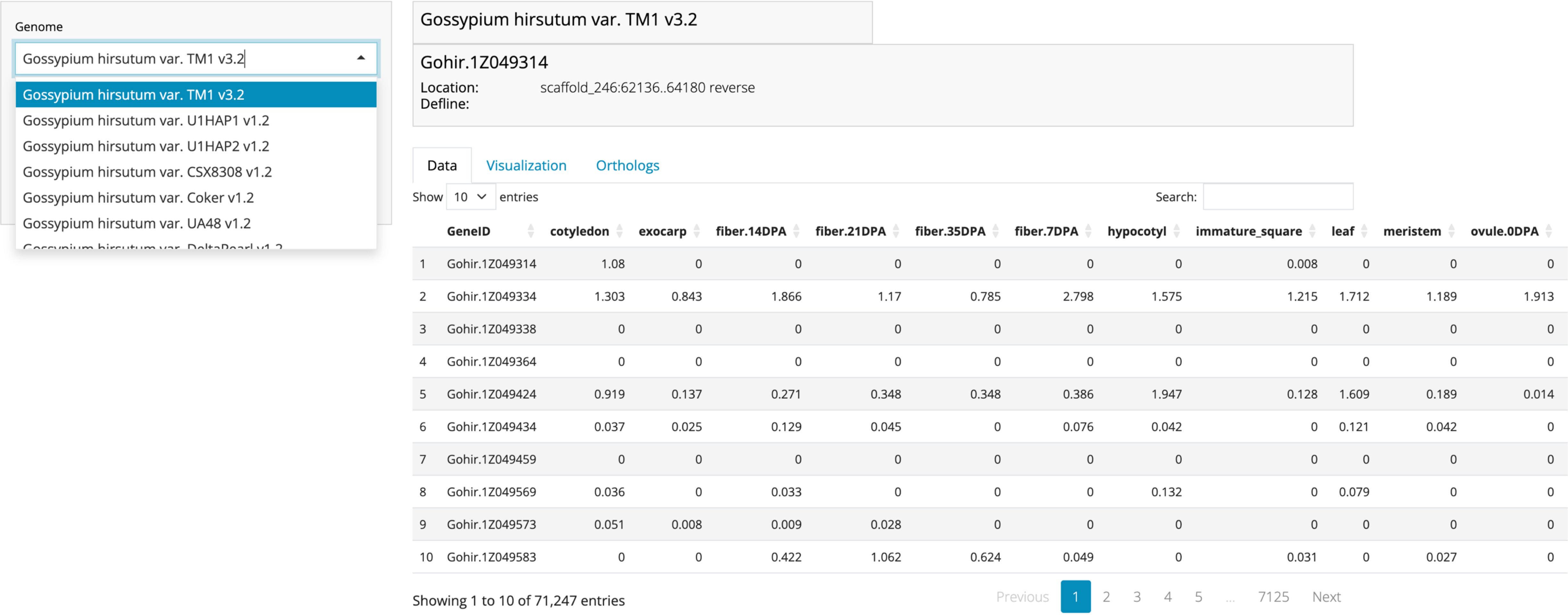
Search

								Cultivars							
Copy number	pa			Genome	chr	start	end	UGA230	UA48	U1HAP2	U1HAP1	FM958	DeltaPearl	Coker	CSX8308
MULTI			000	Ghirsutum	A01	11344	12405	GhUGA230.A01G000201	GhUA48.A01G000250	GhU1H2.D10G282100*	GhU1H1.A01G000200,GhU1H1.D10G282800*	GhFM958.A01G000132,GhFM958.D10G302025*	GhDeltaPearl.A01G000401,GhDeltaPearl.D10G296901*	GhCoker.A01G000266,GhCoker.D10G300401*	GhCSX83081G00035
PAV	A01	43974	GhUA48.A01G000325	GhirsutumUA48	A01	25464	28802		GhUA48.A01G000325						
MULTI	A01	7768	Gohir.A01G000102	Ghirsutum	A01	10930	11191	GhUGA230.A01G000601	GhUA48.A01G000175,GhUA48.A01G000601+	GhU1H2.A01G000150	GhU1H1.A01G000100,GhU1H1.A01G000500+	GhFM958.A01G000200,GhFM958.A01G000066+,GhFM958.A01G000601+	GhDeltaPearl.A01G000201,GhDeltaPearl.A01G000901+	GhCoker.A01G000400,GhCoker.A01G000133+,GhCoker.A01G000701+	GhCSX83081G000225,(SX8308.A010750+

What can we learn from a ‘pangenome’?

- Our goal is to make the pan-genome *useful*
- What deliverables can we extract from the pangenome that are directly usable by breeders?
  - Large structural variants
  - Access to “**pan-gene sets**” … the group of genes across genomes that should recombine (be in synteny) and share a common ancestor (orthologs)
  - **Compare gene expression** among pan-gene sets

Multi-reference gene expression



Genome: Gossypium hirsutum var. TM1 v3.2

Gossypium hirsutum var. TM1 v3.2

Gohir.1Z049314

Location: scaffold_246:62136..64180 reverse

Defline:

Show 10 entries

Search:

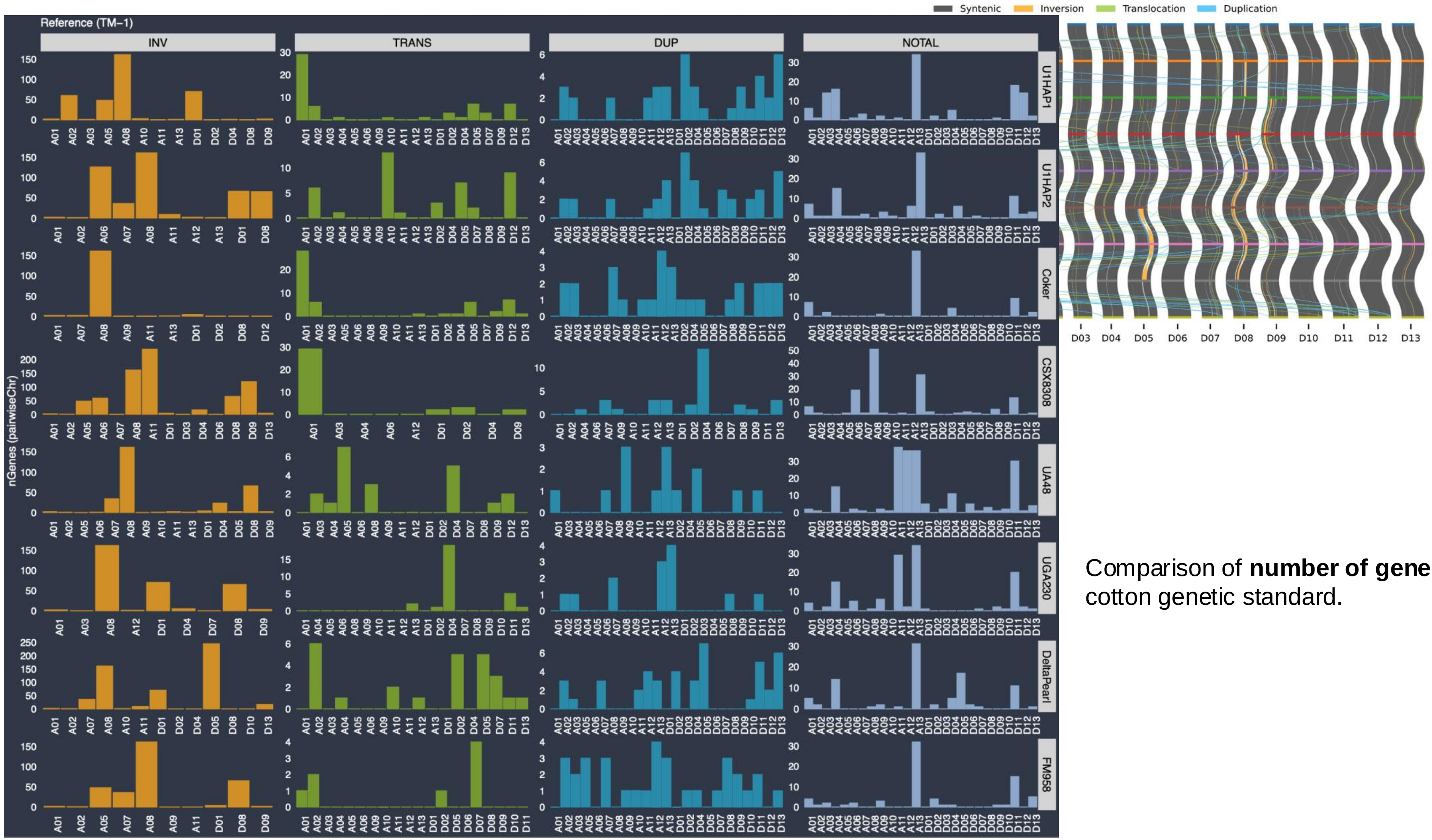
	GeneID	cotyledon	exocarp	fiber.14DPA	fiber.21DPA	fiber.35DPA	fiber.7DPA	hypocotyl	immature_square	leaf	meristem	ovule.0DPA
1	Gohir.1Z049314	1.08	0	0	0	0	0	0	0.008	0	0	0
2	Gohir.1Z049334	1.303	0.843	1.866	1.17	0.785	2.798	1.575	1.215	1.712	1.189	1.913
3	Gohir.1Z049338	0	0	0	0	0	0	0	0	0	0	0
4	Gohir.1Z049364	0	0	0	0	0	0	0	0	0	0	0
5	Gohir.1Z049424	0.919	0.137	0.271	0.348	0.348	0.386	1.947	0.128	1.609	0.189	0.014
6	Gohir.1Z049434	0.037	0.025	0.129	0.045	0	0.076	0.042	0	0.121	0.042	0
7	Gohir.1Z049459	0	0	0	0	0	0	0	0	0	0	0
8	Gohir.1Z049569	0.036	0	0.033	0	0	0	0.132	0	0.079	0	0
9	Gohir.1Z049573	0.051	0.008	0.009	0.028	0	0	0	0	0	0	0
10	Gohir.1Z049583	0	0	0.422	1.062	0.624	0.049	0	0.031	0	0.027	0

Showing 1 to 10 of 71,247 entries

What can we learn from a 'pangenome'?

- Our goal is to make the pan-genome *useful*
- What deliverables can we extract from the pangenome that are directly usable by breeders?
  - Large structural variants
  - Access to "**pan-gene sets**" … the group of genes across genomes that should recombine (be in synteny) and share a common ancestor (orthologs)
  - **Compare gene expression** among pan-gene sets
  - Find putative **functional DNA sequence variants** among genomes

Multi-reference functional variants



Comparison of **number of genes** affected by LSVs relative to TM-1, cotton genetic standard.

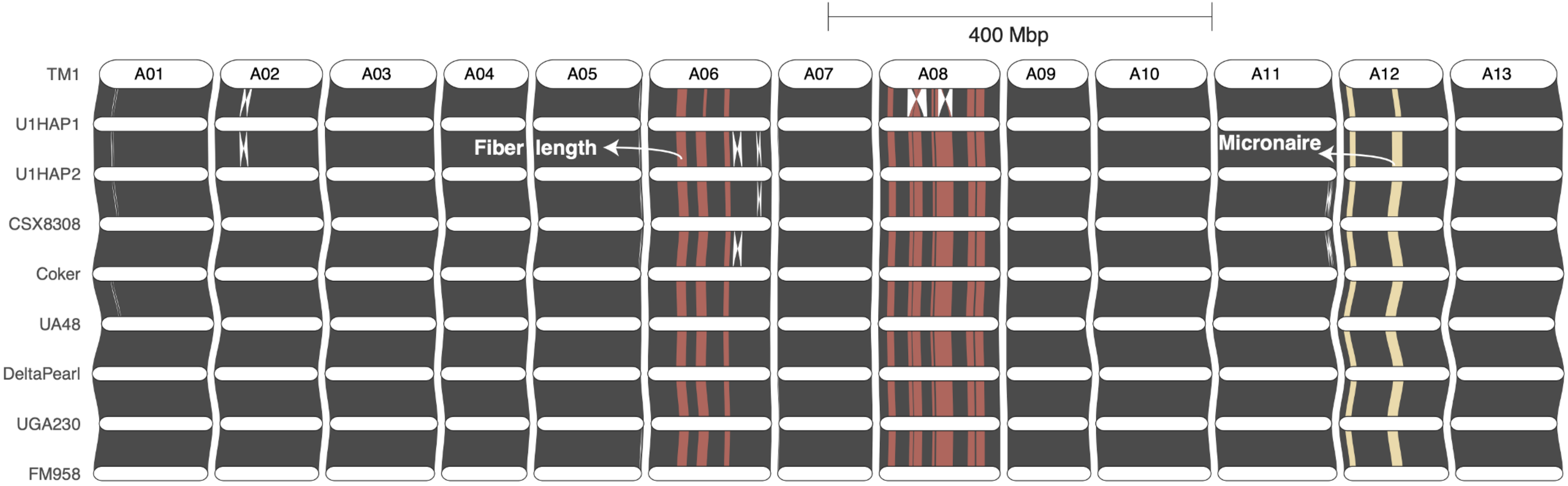
What can we learn from a ‘pangenome’?

HUDSONALPHA GENOME SEQUENCING CENTER

- Our goal is to make the pan-genome *useful*
- What deliverables can we extract from the pangenome that are directly usable by breeders?
  - Large structural variants
  - Access to “**pan-gene sets**” … the group of genes across genomes that should recombine (be in synteny) and share a common ancestor (orthologs)
  - **Compare gene expression** among pan-gene sets
  - Find putative **functional DNA sequence variants** among genomes

  - **Explore QTL** and other genomic regions of interest (ROI)

Tracking QTL across multiple genomes



Chromosomes scaled by physical position

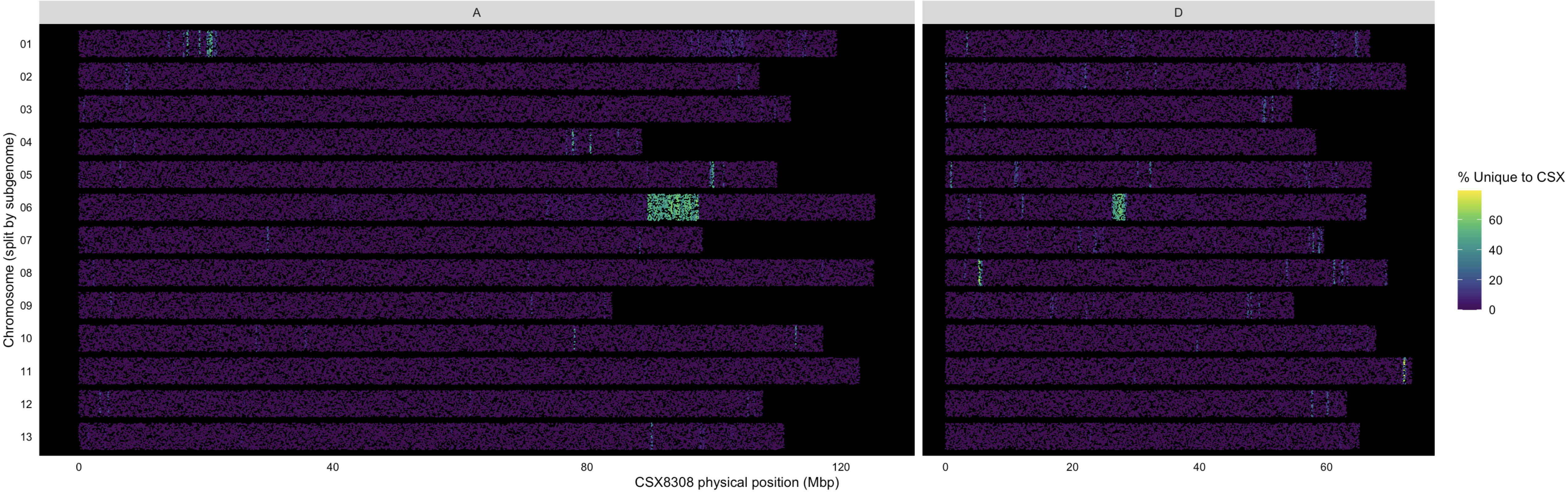
What can we learn from a ‘pangenome’?

- Our goal is to make the pan-genome *useful*
- What deliverables can we extract from the pangenome that are directly usable by breeders?
  - Large structural variants
  - Access to “**pan-gene sets**” … the group of genes across genomes that should recombine (be in synteny) and share a common ancestor (orthologs)
  - **Compare gene expression** among pan-gene sets
  - Find putative **functional DNA sequence variants** among genomes

  - **Explore QTL** and other genomic regions of interest (ROI)
  - Test for **sequences that are unique** to one or more genomes

K-mer genotyping for fast tracking plant breeding

HUDSONALPHA GENOME SEQUENCING CENTER



What can we learn from a ‘pangenome’?

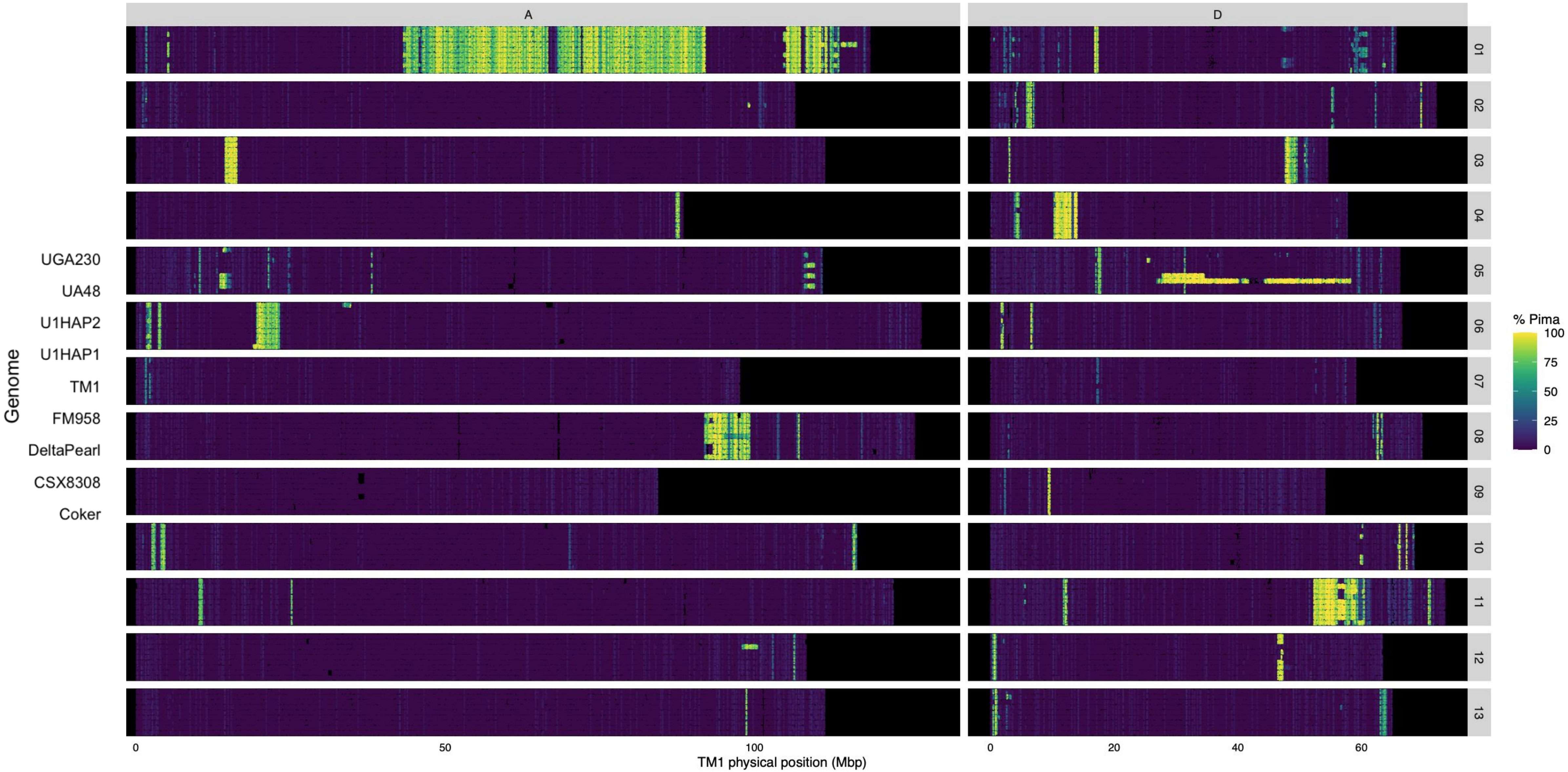
HUDSONALPHA GENOME SEQUENCING CENTER

- Our goal is to make the pan-genome *useful*
- What deliverables can we extract from the pangenome that are directly usable by breeders?
  - Large structural variants
  - Access to “**pan-gene sets**” … the group of genes across genomes that should recombine (be in synteny) and share a common ancestor (orthologs)
  - **Compare gene expression** among pan-gene sets
  - Find putative **functional DNA sequence variants** among genomes

  - **Explore QTL** and other genomic regions of interest (ROI)
  - Test for **sequences that are unique** to one or more genomes
  - Test for **regions of co-ancestry** among genomes: example Pima cotton introgressions

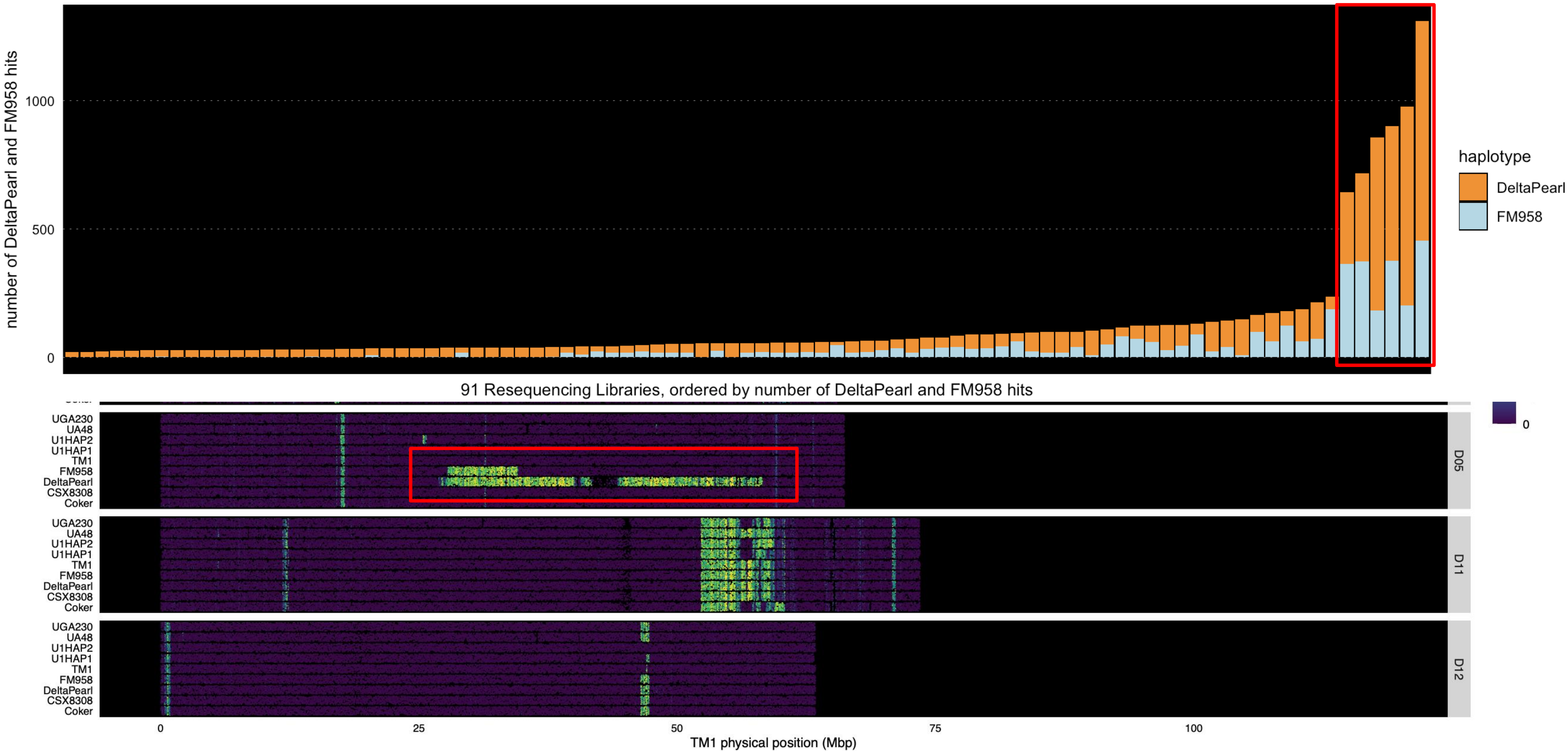
Tracking Pima introgressions in upland cotton

UGA230



Tracking Pima introgressions in upland cotton

UGA230



Next steps

HUDSONALPHA GENOME SEQUENCING CENTER

What deliverables can we extract from the pangenome that are directly usable by breeders?

- Access to “pan-gene sets”
- Compare gene expression
- Define functional DNA sequence variants
- Explore genomic regions of interest (ROI)
- Find sequences that are unique to one or more genomes
- Find Pima (and any other) introgressions

- What else?
  - Potential to integrate public and private resources? (private data delivery, anonymous naming, etc.)
  - …