

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.



Shengqiang Shu

JGI ANNOTATION / PHYTOZOME



Tomáš Brůna



Cotton
Incorporated



National Science Foundation
Directorate for Biological Sciences (BIO)



JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY



Z. Jeffrey Chen
UT Austin



Peng W. Chee
U of Georgia, Tifton



Christopher A. Saski
Clemson University



HudsonAlpha Genome Sequencing Center



Fred Bourland
U of Arkansas



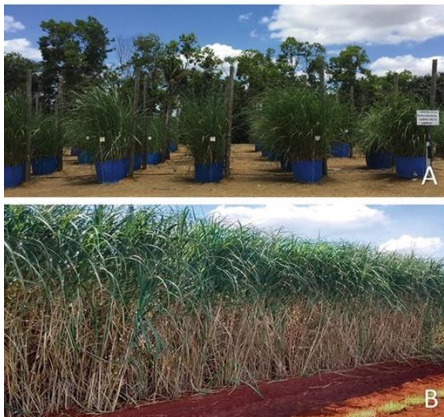
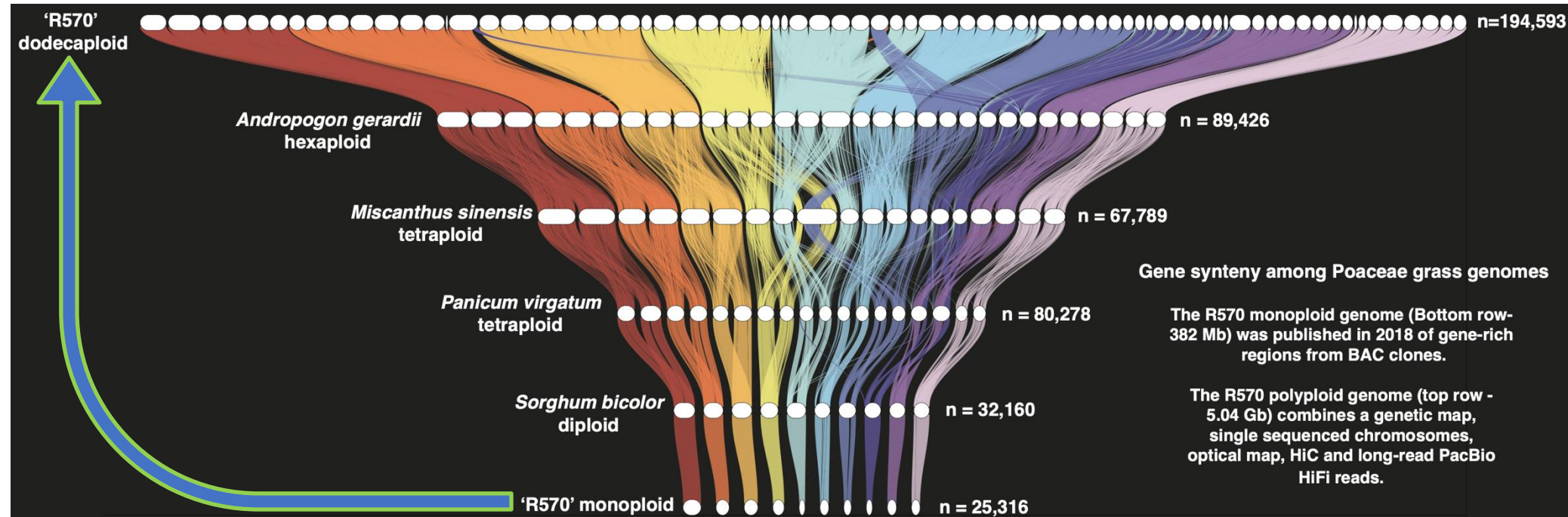
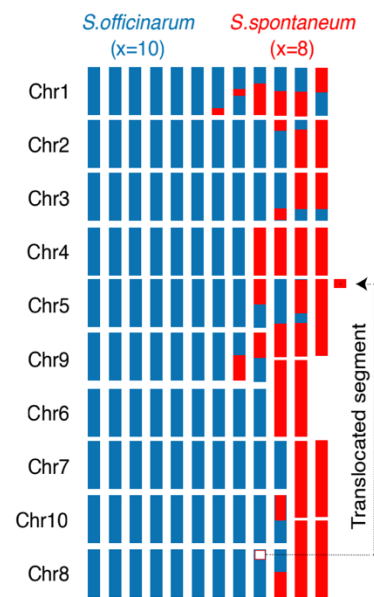
Warwick N. Stiller
CSIRO, Australia

HAGSC Genome Analysis

**HUDSONALPHA
GENOME SEQUENCING
CENTER**



We build complex reference genome assemblies



Cursi et al. (2021)



TOOLS AND RESOURCES



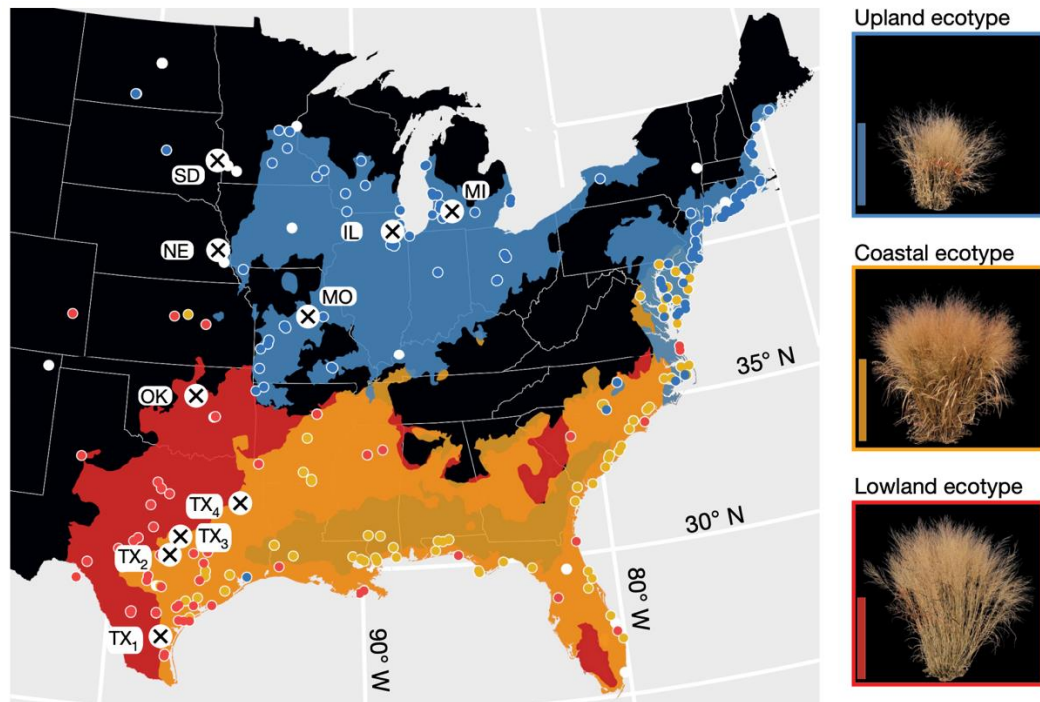
Healey et al. (2024) Nature

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

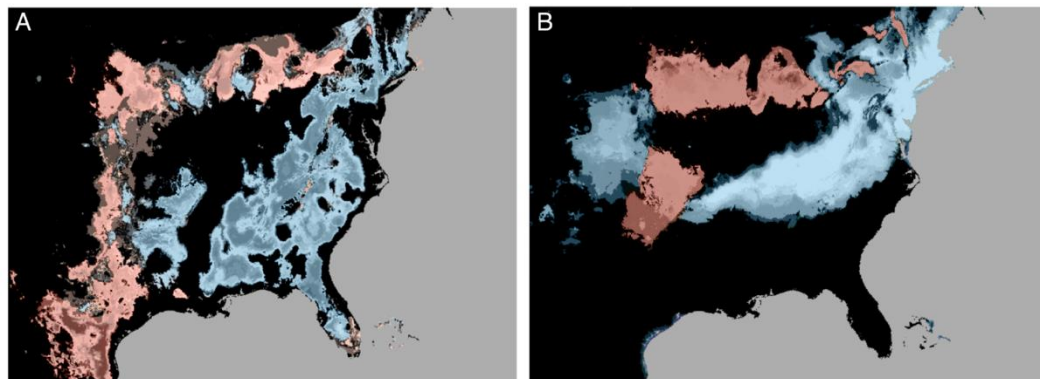
John T Lovell^{1,2*}, Avinash Sreedasyam¹, M Eric Schranz³, Melissa Wilson⁴, Joseph W Carlson², Alex Harkess^{1,5}, David Emms⁶, David M Goodstein², Jeremy Schmutz^{1,2}

Custom evolutionary analysis to optimize complex breeding outcomes

**HUDSONALPHA
GENOME SEQUENCING
CENTER**



Lovell et al. (2021) Nature

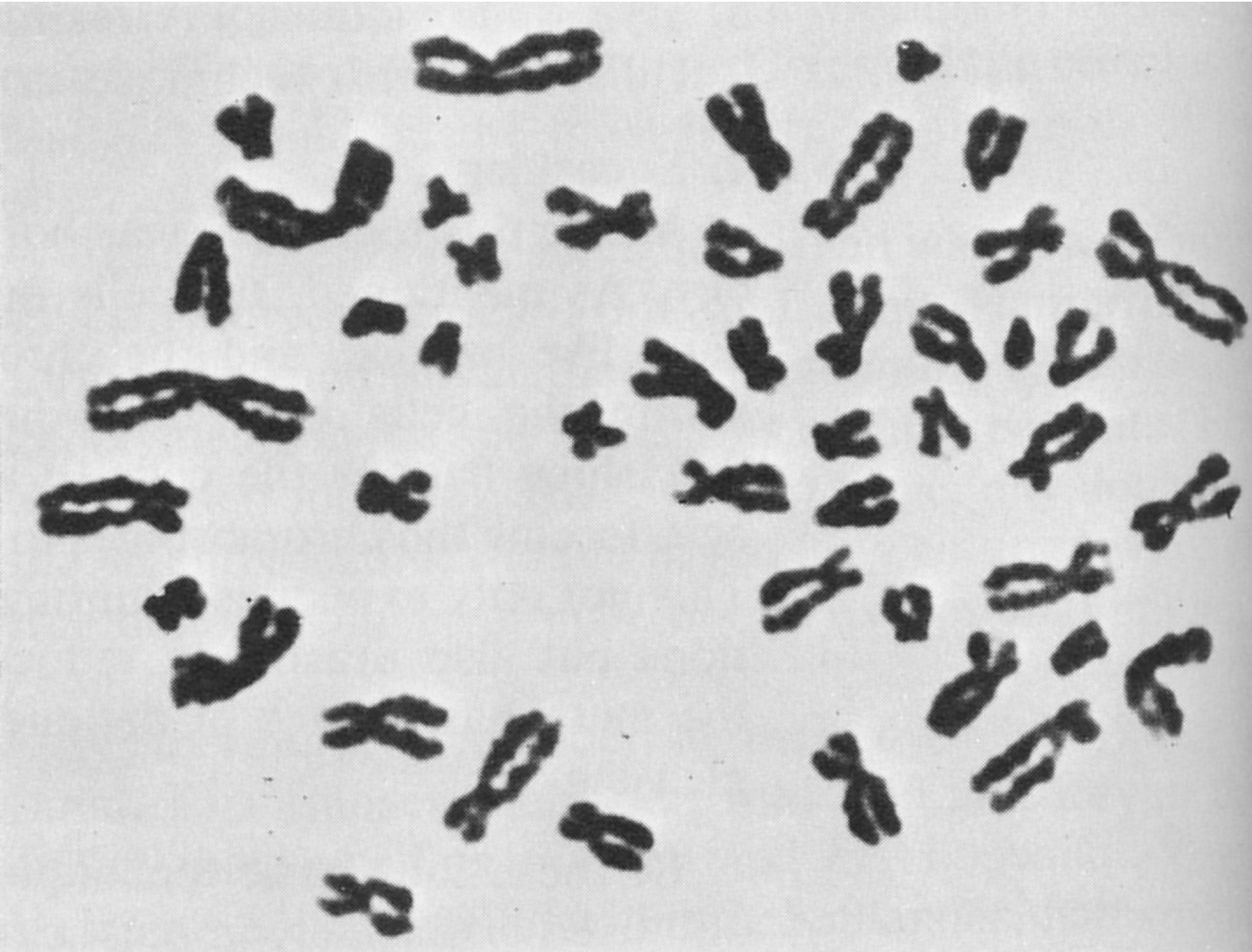


Napier et al. (2022) PNAS

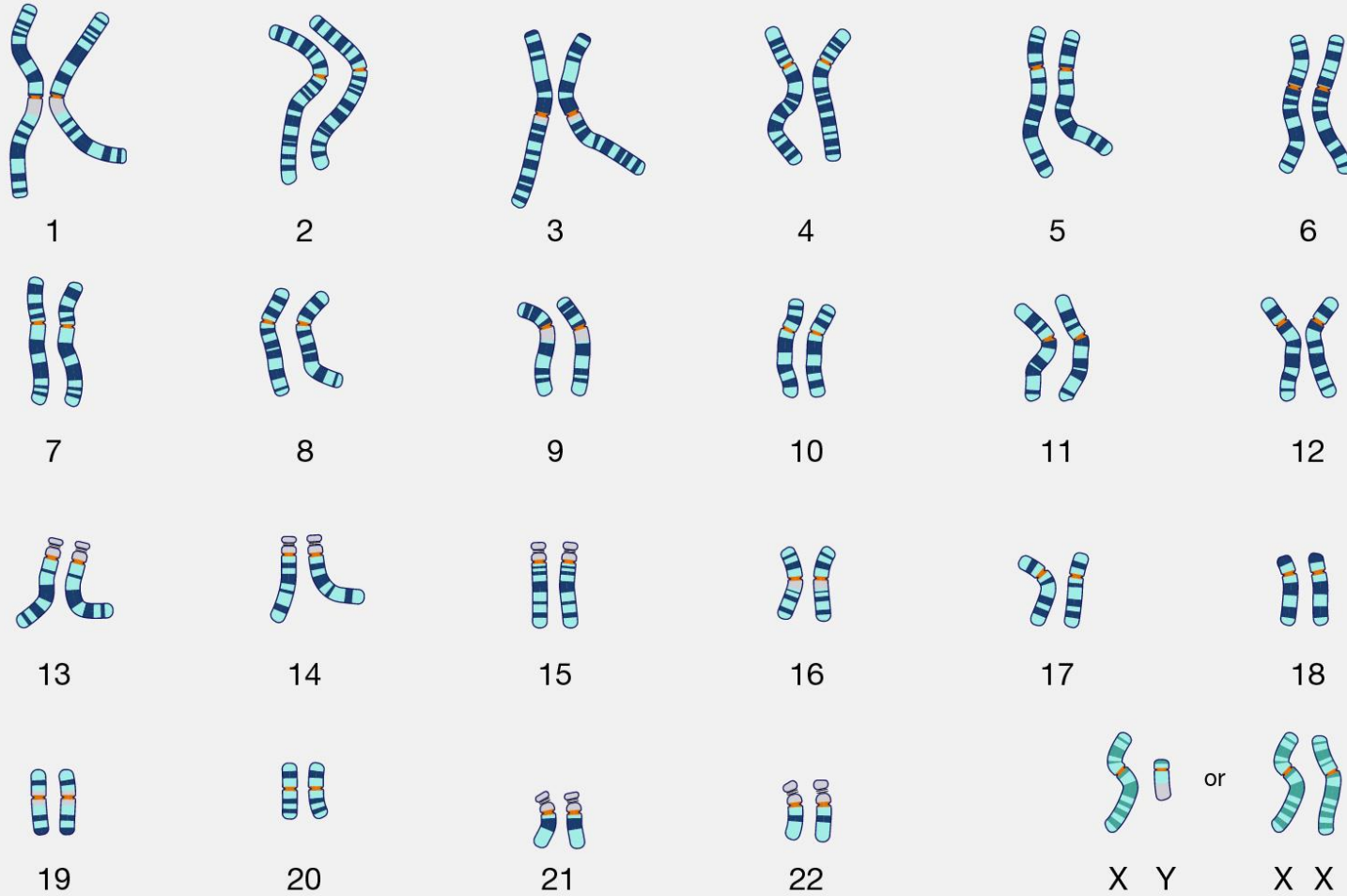


- **Genomics**: why it's important and what progress has been made
- **Cotton genomes**: what has been done, and what is left to do
- **Conceptual issues with pan-genomes**: why we need integrated resources
- Current **state of the cotton pan-genome**: genes, variants and tools
- **Using the pan-genome**: 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?



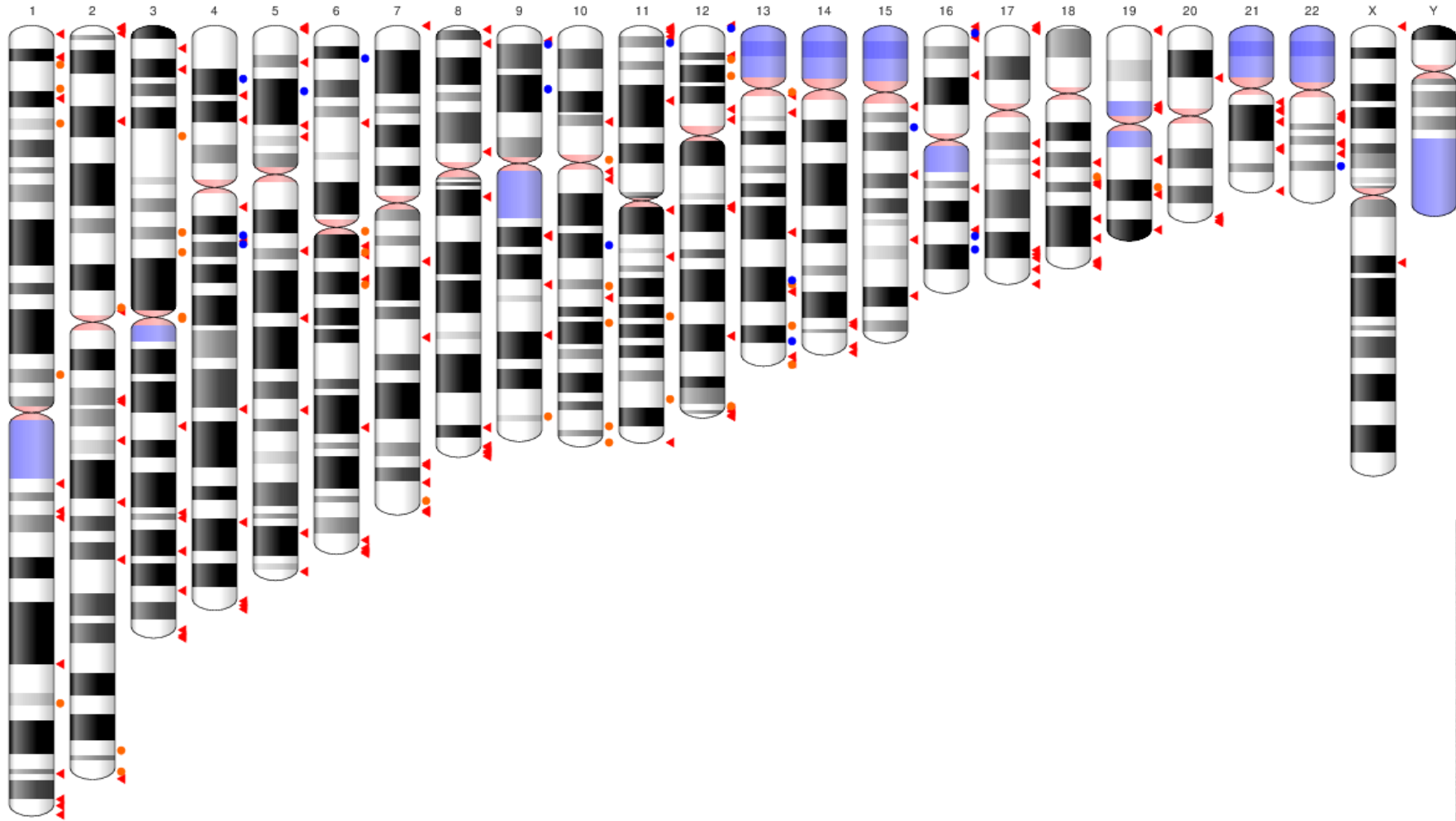
What is a genome?



Shown here are human chromosomes



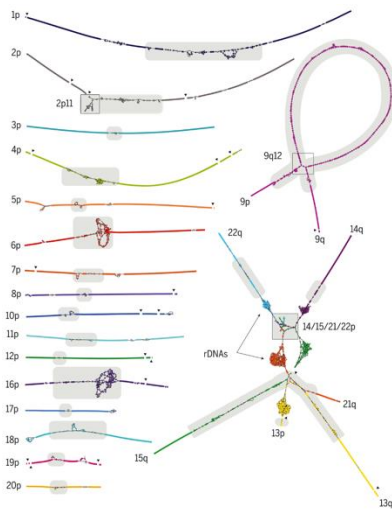
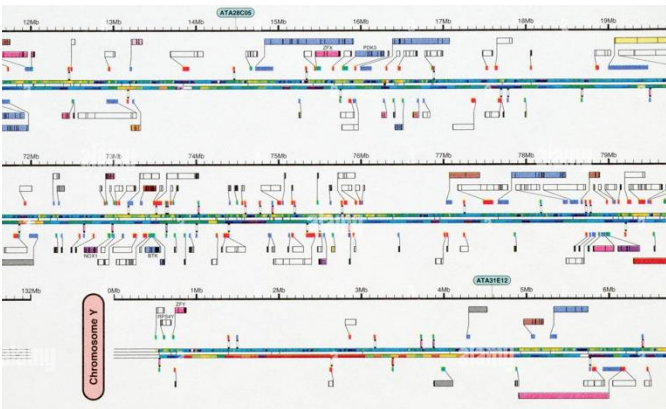
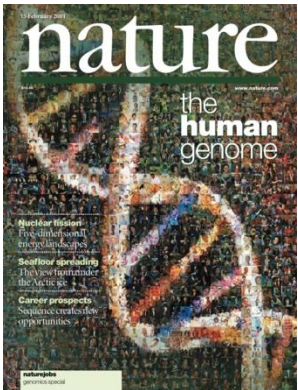
What is a genome?



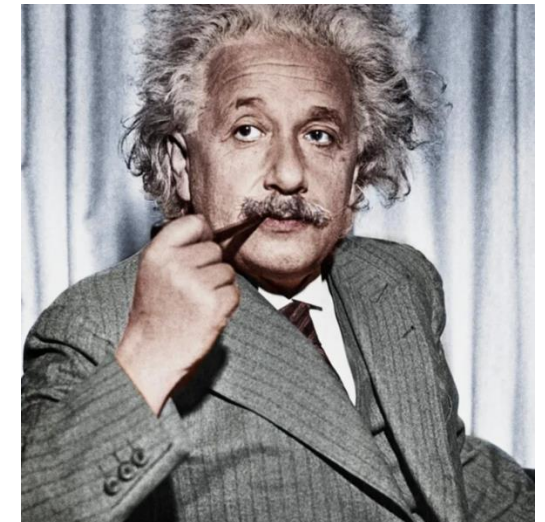
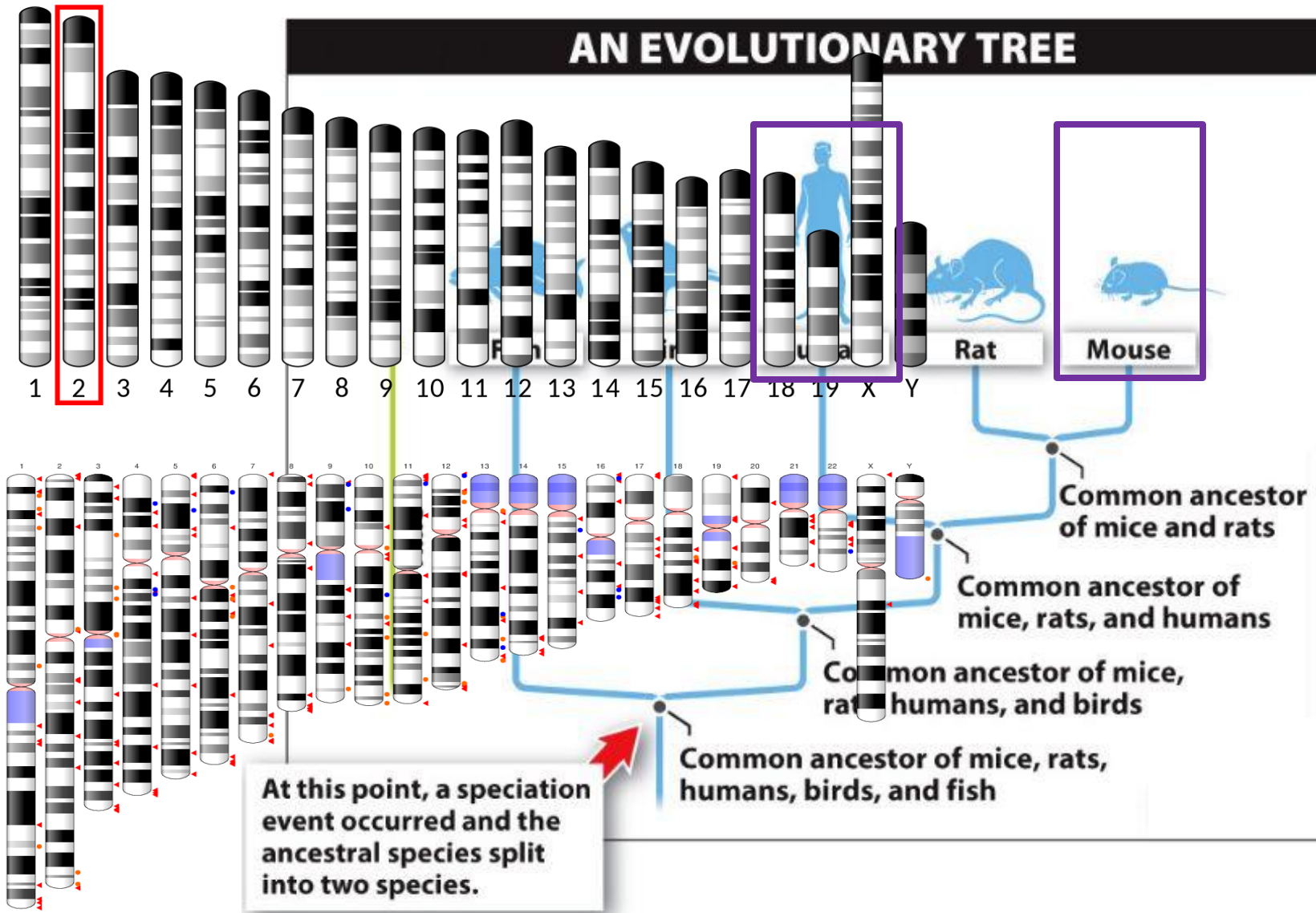
- ◄ Region containing alternate loci
- Region containing fix patches
- Region containing novel patches



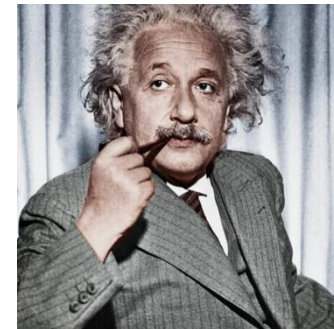
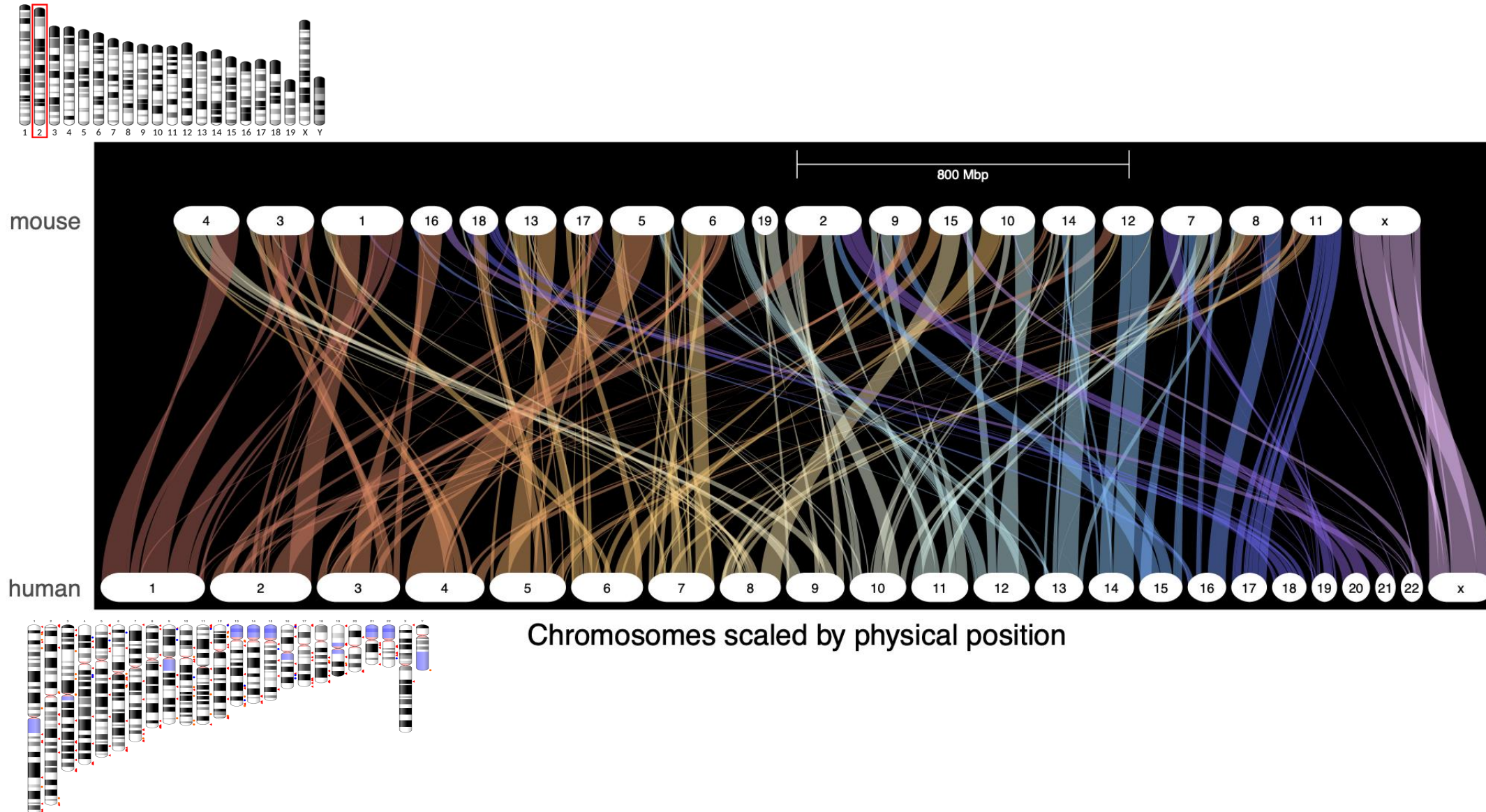
State of the art in genomics



Comparing genomes



Comparing genomes



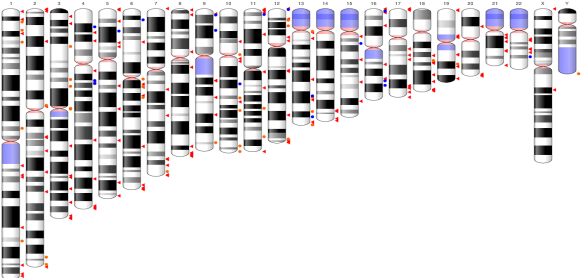
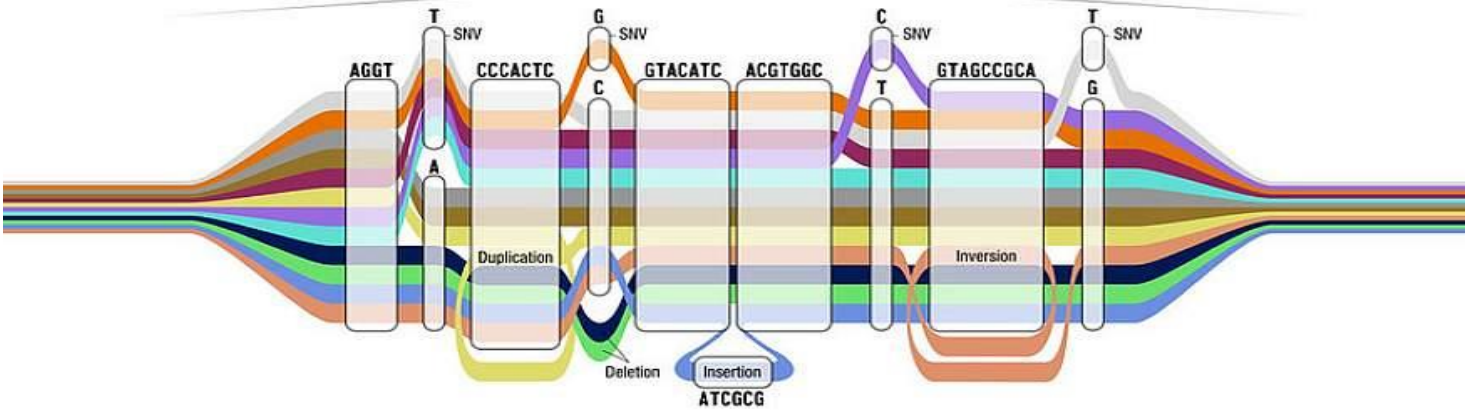
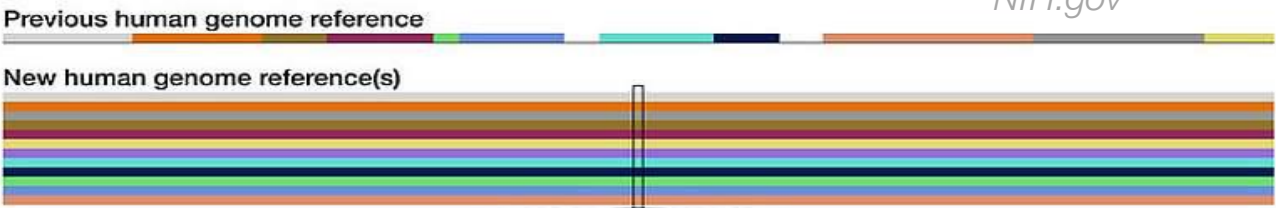
The concept of a pan-genome graph

A draft human pangenome reference

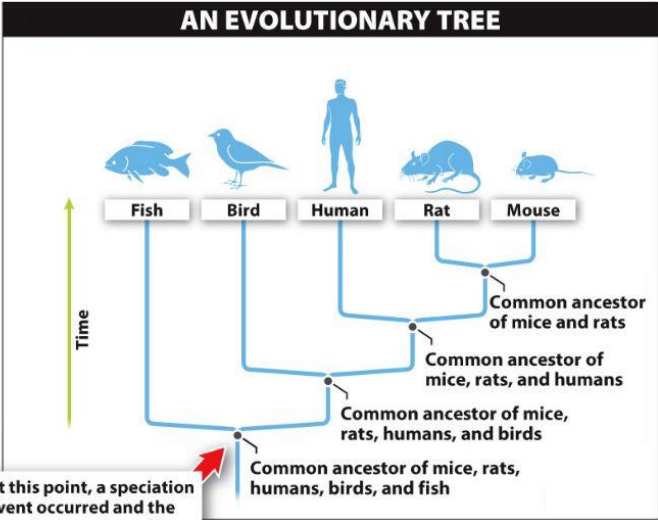
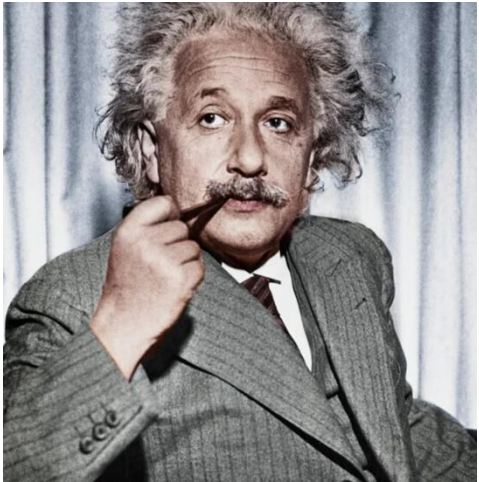
<https://doi.org/10.1038/s41586-023-05896-x>
Received: 9 July 2022
Accepted: 28 February 2023
Published online: 10 May 2023
Open access
 Check for updates

Wen-Wei Liao^{1,2,3,60}, Mobin Asri^{4,60}, Jana Ebler^{5,6,60}, Daniel Doerr^{5,6}, Marina Haukness⁴, Glenn Hickey⁴, Shuangjia Lu^{1,2}, Julian K. Lucas⁴, Jean Montlong⁴, Haley J. Abel⁷, Silvia Buonaiuto⁸, Xian H. Chang⁴, Haoyu Cheng^{8,10}, Justin Chu⁸, Vincenza Colonna^{8,11}, Jordan M. Eizenga⁴, Xiaowen Feng^{8,10}, Christian Fischer¹¹, Robert S. Fulton^{12,13}, Shilpa Garg¹⁴, Cristian Groza¹⁵, Andrea Guarracino^{11,16}, William T. Harvey¹⁷, Simon Heumos^{18,19}, Kerstin Howe²⁰, Miten Jain²¹, Tsung-Yu Lu²², Charles Markello⁴, Fergal J. Martin²³, Matthew W. Mitchell²⁴, Katherine M. Munson¹⁷, Moses Njagi Mwaniki²⁵, Adam M. Novak⁴, Hugh E. Olsen⁴, Trevor Pesout⁴, David Porubsky¹⁷, Piotr Prins¹¹, Jonas A. Sibbesen²⁶, Jouni Sirén⁴, Chad Tomlinson¹⁵, Flavia Villani¹¹, Mitchell R. Vollger^{17,27}, Lucinda L. Antonacci-Fulton¹⁵, Gunjan Baid²⁸, Carl A. Baker¹⁷, Anastasiya Belyaeva²⁸, Konstantinos Billis²³, Andrew Carroll²⁸, Pi-Chuan Chang²⁸, Sarah Cody¹², Daniel E. Cook²⁸, Robert M. Cook-Deegan²⁹, Omar E. Cornejo³⁰, Mark Diekhans⁴, Peter Ebert^{5,6,31}, Susan Fairley³², Olivier Fedrigo³², Adam L. Felsenfeld³³, Giulio Formenti³², Adam Frankish²³, Yan Gao³⁴, Nanibaa' A. Garrison^{35,36,37}, Carlos Garcia Giron²³, Richard E. Green^{38,39}, Leanne Haggerty²³, Kendra Hoekzema¹⁷, Thibaut Hourlier²³, Hanlee P. Ji¹⁰, Eimear E. Kenny⁴¹, Barbara A. Koenig⁴², Alexey Kolesnikov²⁸, Jan O. Korbel^{23,43}, Jennifer Kordosky¹⁷, Sergey Koren⁴⁴, HoJoon Lee⁴⁰, Alexandra P. Lewis¹⁷, Hugo Magalhães^{5,6}, Santiago Marco-Sola^{45,46}, Pierre Marjion^{5,6}, Ann McCartney⁴⁴, Jennifer McDaniel⁴⁷, Jacquelyn Mountcastle³², Maria Nattestad²⁸, Sergey Nurk⁴⁴, Nathan D. Olson¹⁷, Alice B. Popejoy⁴⁸, Daniela Puiu⁴⁹, Mikko Rautiainen¹⁴, Allison A. Regier¹⁵, Arang Rhie⁴⁴, Samuel Sacco¹⁰, Ashley D. Sanders⁵⁰, Valerie A. Schneider⁵¹, Baergen I. Schultz³³, Kishwar Shafin²⁸, Michael W. Smith³³, Heidi J. Sofia³³, Ahmad N. Abou Tayoun^{52,53}, Françoise Thibaud-Nissen⁵¹, Francesca Floriana Tricomi²³, Justin Wagner⁴⁷, Brian Walenz⁴⁴, Jonathan M. D. Wood⁵⁰, Aleksey V. Zimin^{49,54}, Guillaume Bourque^{55,56,57}, Mark J. P. Chaisson²², Paul Flicek²³, Adam M. Phillippy¹⁴, Justin M. Zook⁴⁷, Evan E. Eichler^{17,58}, David Haussler^{4,58}, Ting Wang^{12,13}, Erich D. Jarvis^{32,58,59}, Karen H. Miga⁴, Erik Garrison^{11,52}, Tobias Marschall^{5,6,52}, Ira M. Hall^{1,2,52}, Heng Li^{19,10,52} & Benedict Paten^{4,52}

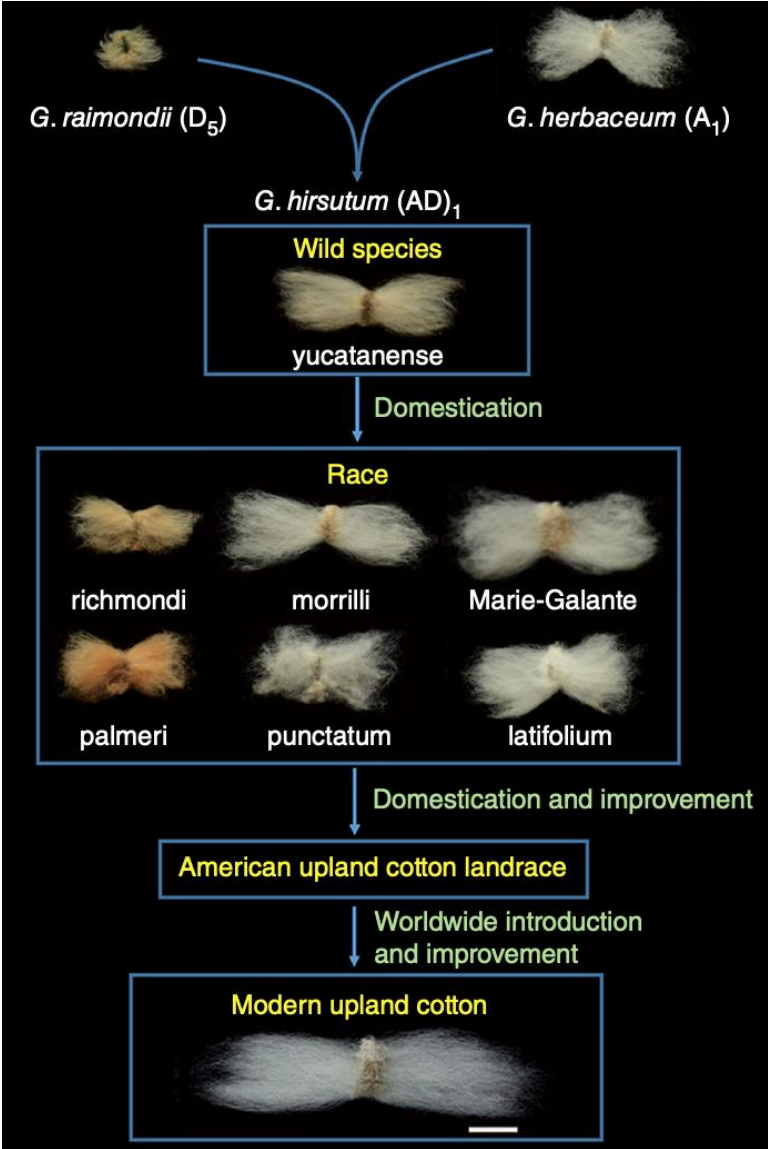
NIH.gov



Structure of the tetraploid cotton genomes

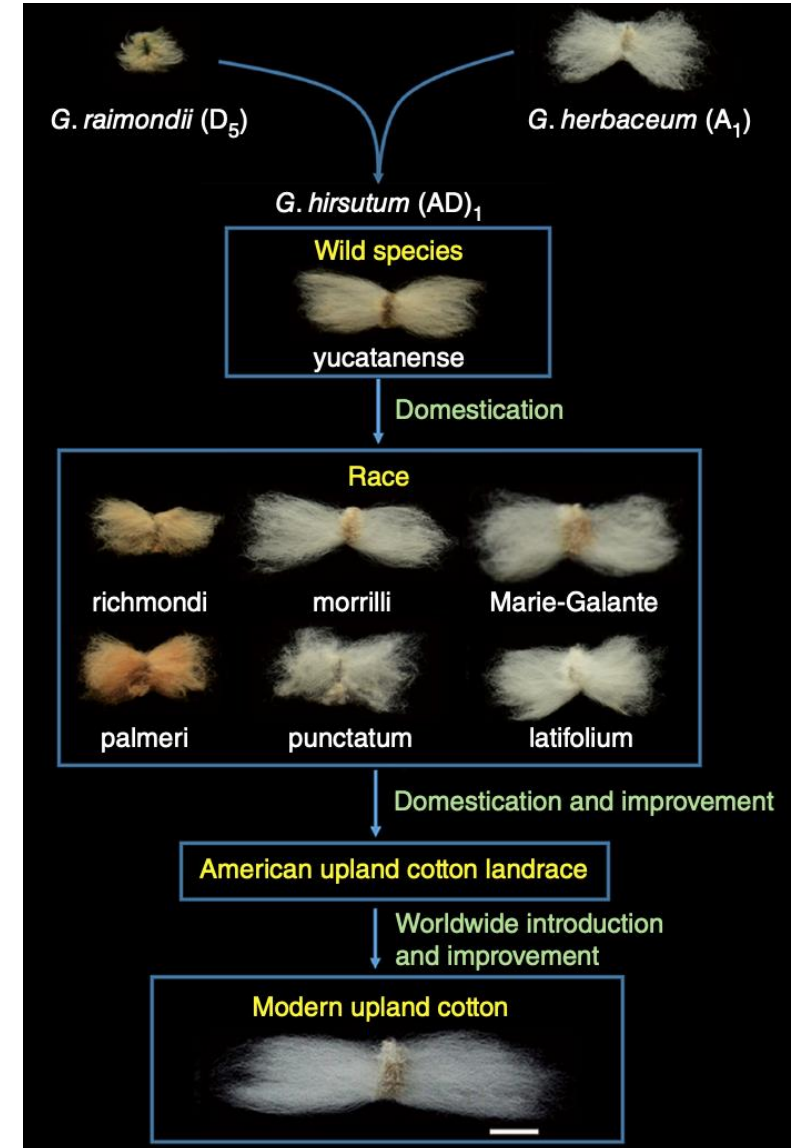
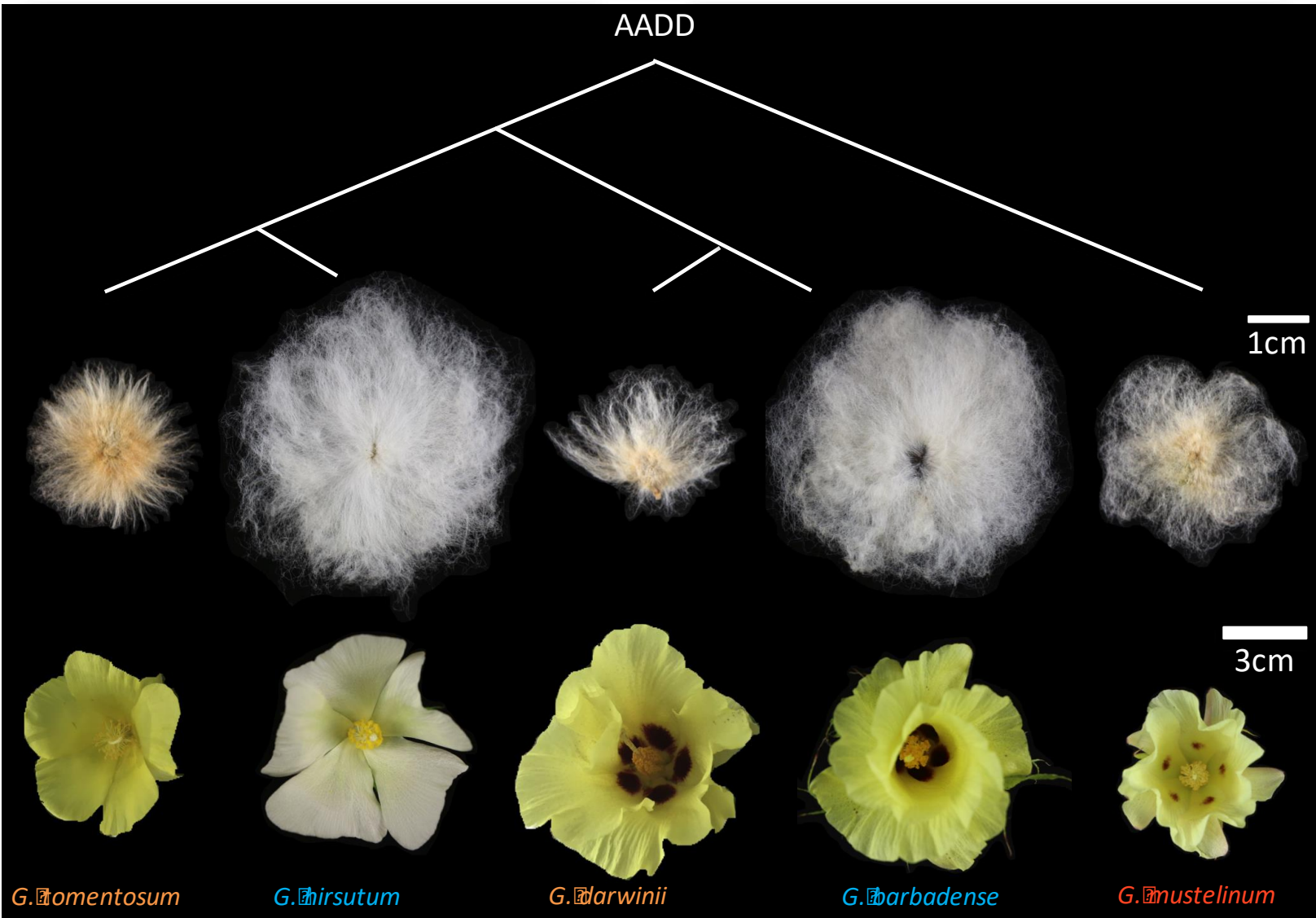


≠



Evolutionary history of cotton

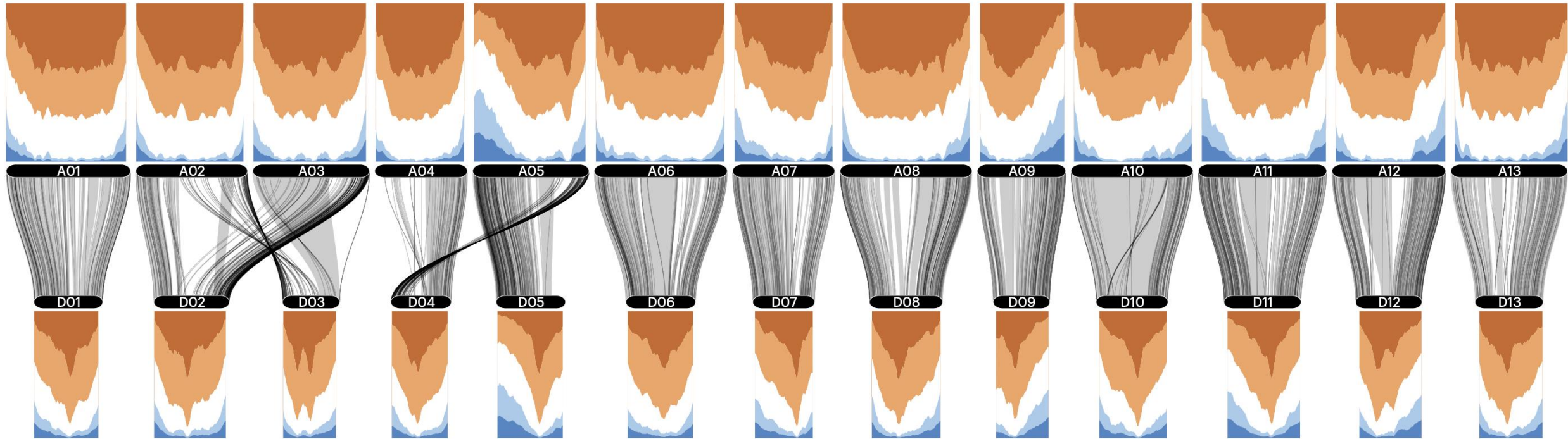
HUDSONALPHA
GENOME SEQUENCING
CENTER



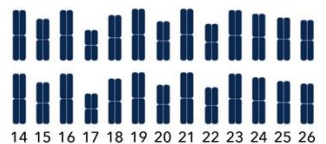
Structure of the tetraploid cotton genomes

b Cotton genome structure and subgenome evolution

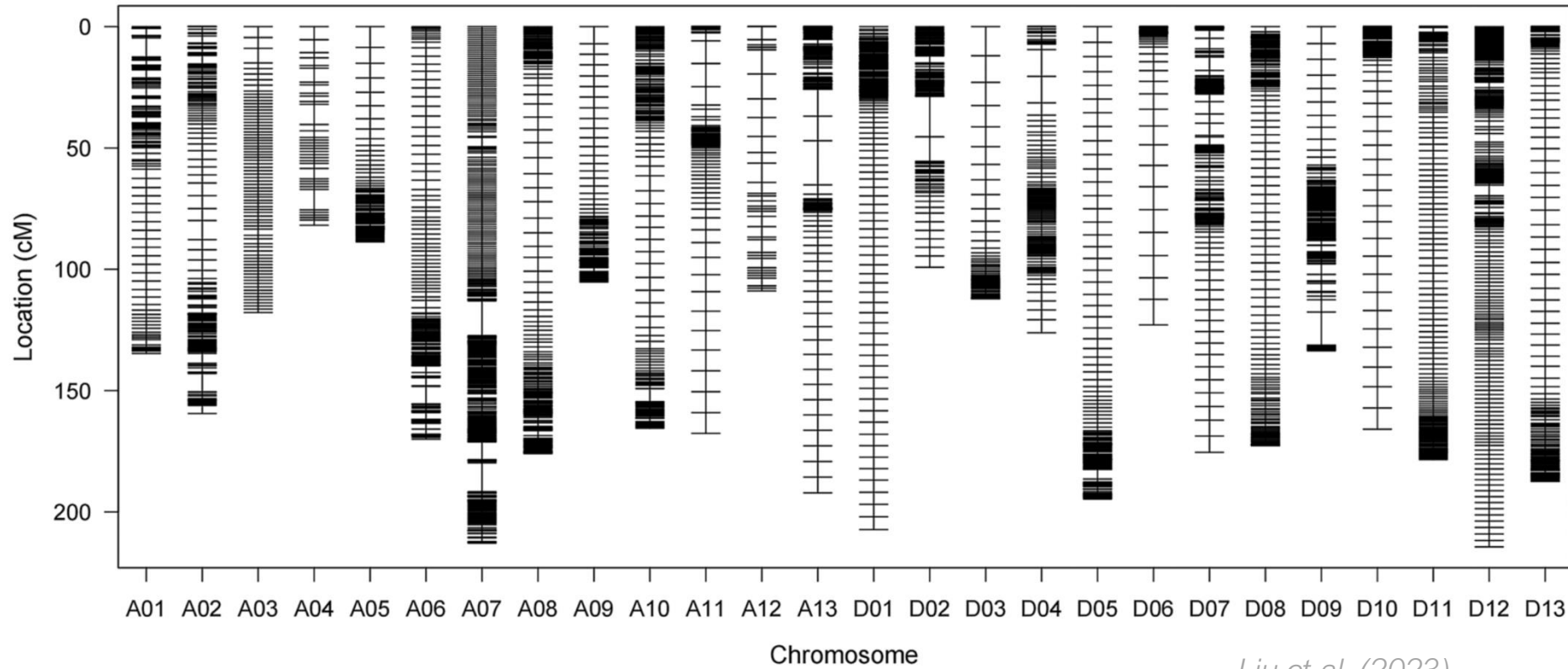
Repetitive sequence: Ty3 LTR Other Genic sequence: Exons Introns



Sreedasyam et al. (2024)

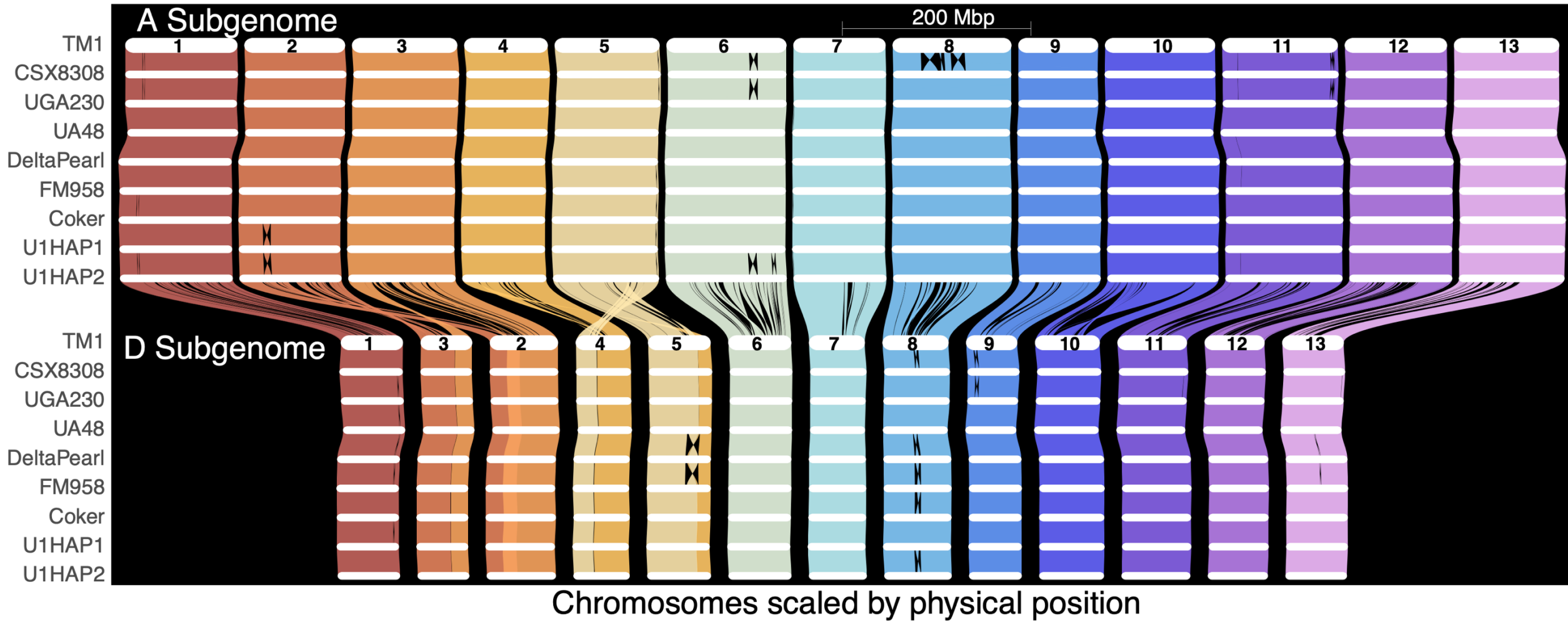


Cotton genomes: evolution and quality



Liu et al. (2023)

The emerging cotton “pan-genome”



Cotton genomes: evolution and quality

a Genome contiguity improvements from v2 to v3
v2 | contig N50 = 0.79 Mb, contigs per chr = 220.1

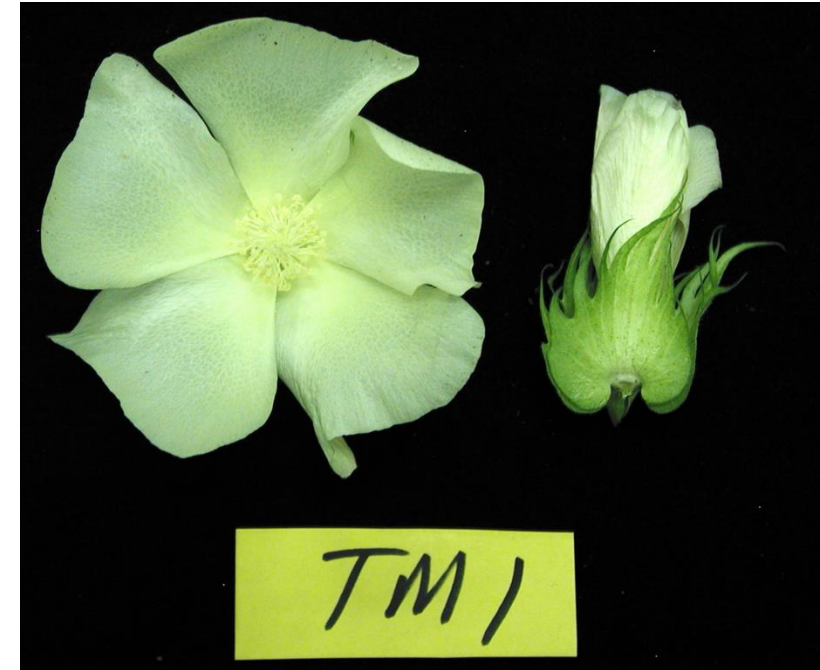
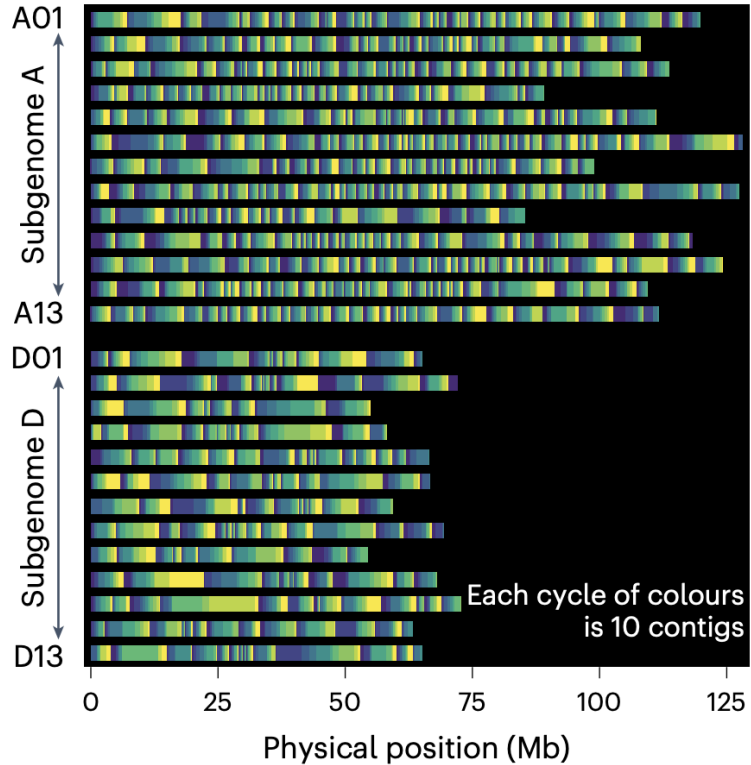
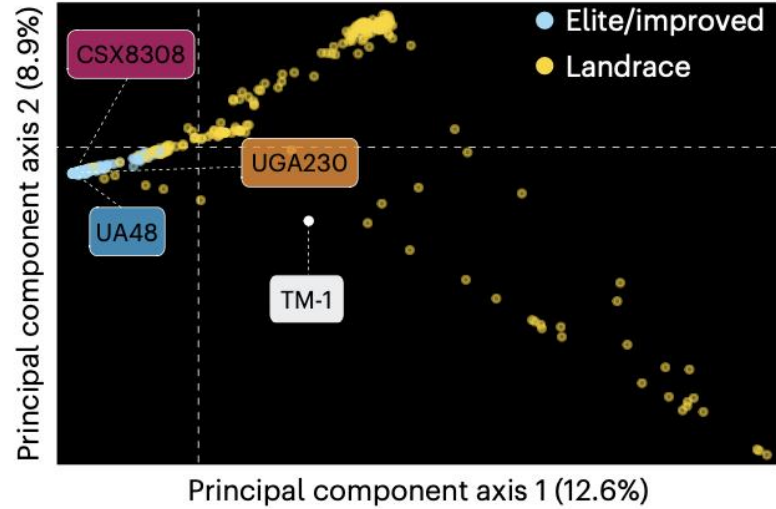


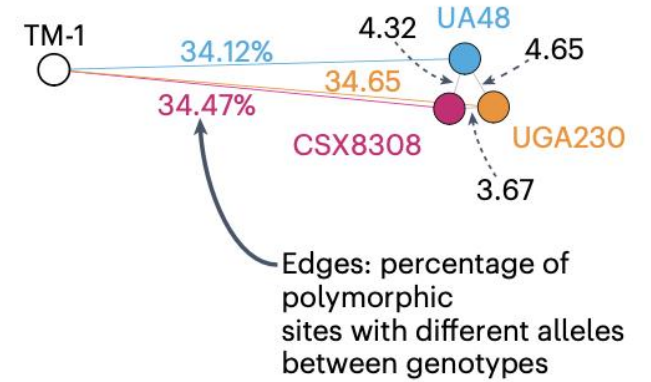
Photo courtesy of Z.J. Chen

Cotton genomes: evolution and quality

a Cultivated and improved germplasm PCA

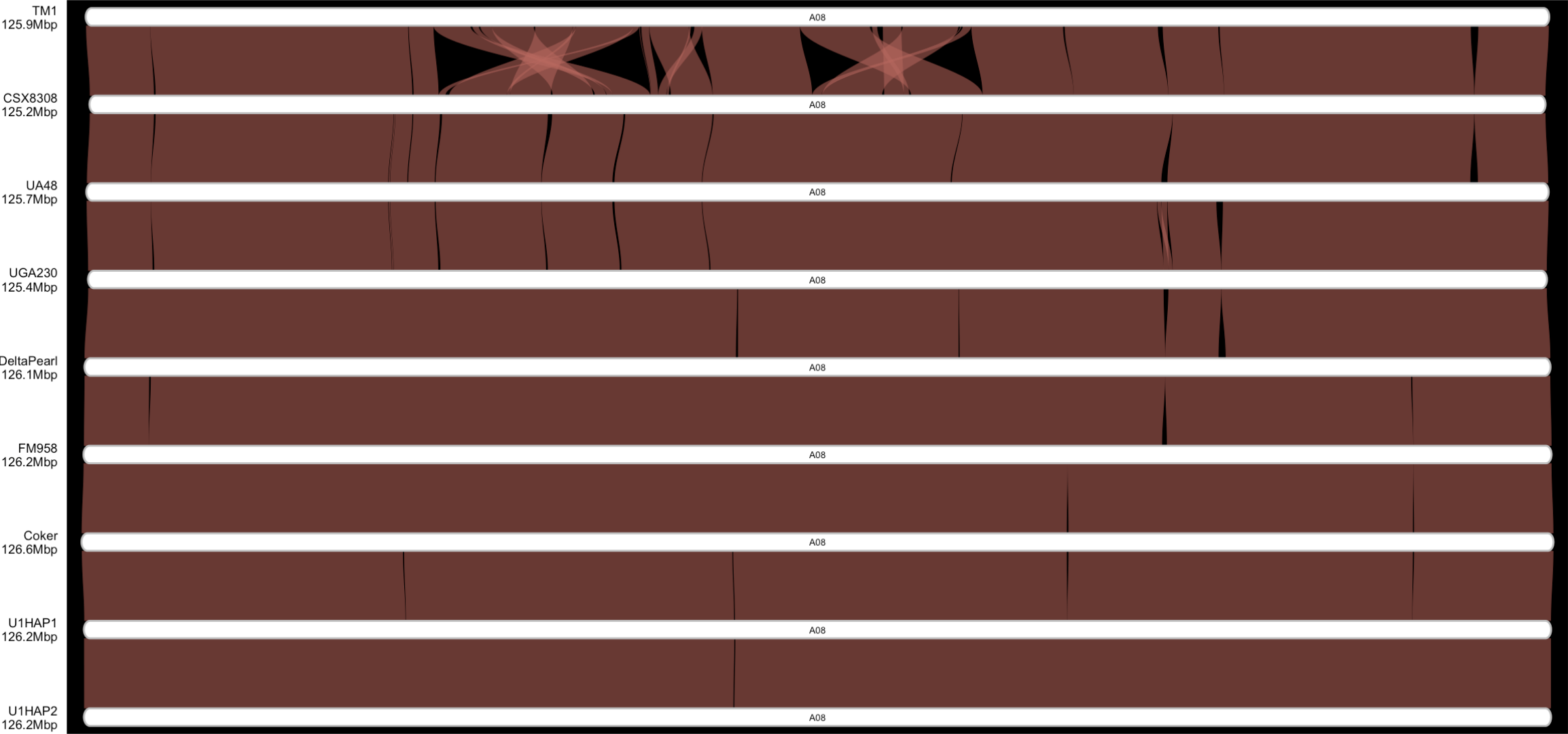


c Genetic similarity graph



Cotton genomes: evolution and quality

DEEPSPACE syntenic map



chromosomes scaled by physical size

Cotton genomes: evolution and quality

G3

Genes | Genomes | Genetics

GSA

Genetics Society of America

IssuesMore ContentSeriesSubmitAlertsAbout

Article Navigation

JOURNAL ARTICLEACCEPTED 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

Resourcehttps://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 Mamidi¹, Sameer Khanal³, Jerry W. Jenkins¹, Christopher Plott¹, Kempton B. Bryan⁴, Zhigang Li⁴, Shengqiang Shu², Joseph Carlson², David Goodstein², Luis De Santiago⁵, Ryan C. Kirkbride⁵, Sebastian Calleja⁶, Todd Campbell⁷, Jenny C. Koebernick⁸, Jane K. Dever^{9,16}, Jodi A. Scheffler¹⁰, Duke Pauli¹¹, Johnnie N. Jenkins¹¹, Jack C. McCarty¹¹, Melissa Williams¹, LoriBeth Boston¹, Jenell Webber¹, Joshua A. Udall¹², Z. Jeffrey Chen⁵, Fred Bourland¹³, Warwick N. Stiller¹⁴, Christopher A. Saski¹, Jane Grimwood¹, Peng W. Chee¹, Don C. Jones¹⁵ & Jeremy Schmutz¹²✉

ARTICLESnature geneticshttps://doi.org/10.1038/s41588-020-0607-4

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. Percy⁴, Mingzhou Bai⁵, Yang Li³, James E. Frelichowski⁴, Jiang Hu⁶, Kun Wang^{10,3}, John Z. Yu¹⁴✉ and Yuxian Zhu¹⁰✉

G3

Genes | Genomes | Genetics

GSA

Genetics Society of America

IssuesMore ContentSeriesSubmitAlertsAbout

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 journalS E B

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

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

Nian Wang¹, Yuanxue Li¹, Chao Shen², Yang Yang¹, Hongya Wang¹, Tian Yao¹, Xianlong Zhang¹, Keith Lindsey³ and Zhongxu Lin^{1,2}✉

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

PNASRESEARCH ARTICLEPLANT BIOLOGYOPEN ACCESS

Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons

Renhai Peng¹, Yanchao Xu^{1,2}✉, Shilin Tian^{1,2}✉, Turgay Unver^{1,2}✉, Zhen Liu^{1,2}✉, Zhongli Zhou^{1,2}✉, Xiaoyan Cai^{1,2}✉, Kunbo Wang¹, Yangyang Wei¹, Yuling Liu¹, Heng Wang¹, Guanjing Hu^{1,2}✉, Zhongren Zhang¹✉, Corrinne E. Grover¹✉, Yuqing Hou¹, Yuhong Wang¹, Pengtao Li¹, Tao Wang¹, Quanwei Lu¹, Yuanyuan Wang¹, Justin L. Conover¹✉, Hassan Ghazal¹✉, Qinglian Wang¹✉, Baohong Zhang^{1,2}✉, Marc Van Montagu^{1,2}✉, Yves Van de Peer^{1,2,3}✉, Jonathan F. Wendel^{1,2}✉, and Fang Liu^{1,2}✉

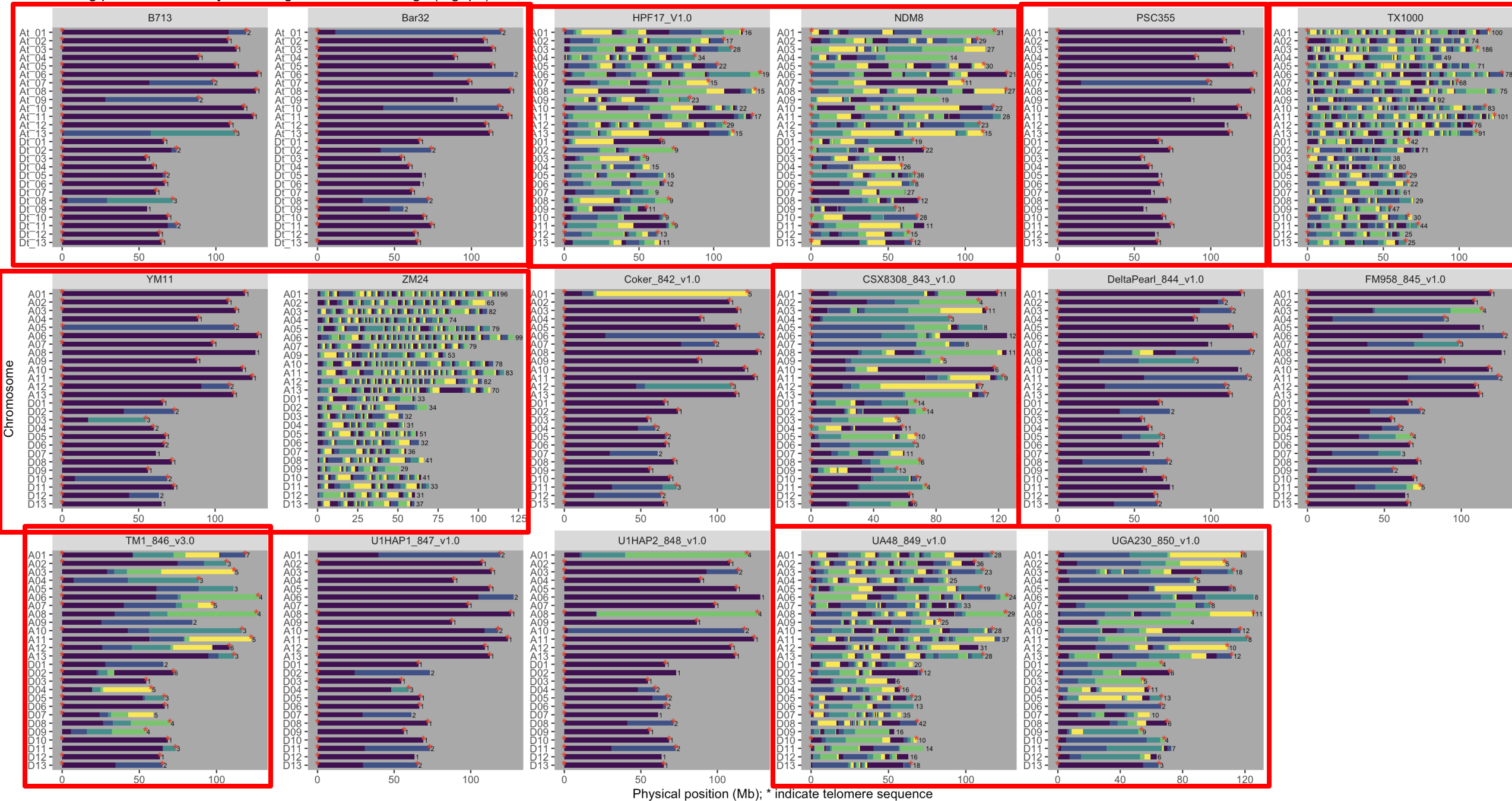
nature geneticsARTICLEShttps://doi.org/10.1038/s41588-021-00910-2

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 Ke¹✉, Bin Chen¹✉, Zhengwen Liu¹✉, Qishen Gu¹✉, Zhicheng Wang¹✉, Guoning Wang¹✉, Jun Yang¹✉, Jinhua Wu¹✉, Yuanyuan Yan¹✉, Chengsheng Meng¹✉, Lihua Li¹✉, Xiuxin Li²✉, Shaojing Mo¹✉, Nan Wu¹✉, Limei Ma¹✉, Liting Chen¹✉, Man Zhang¹✉, Aijun Si¹✉, Zhanwu Yang¹✉, Nan Wang¹✉, Lizhu Wu¹✉, Dongmei Zhang¹✉, Yanru Cui¹✉, Jing Cui¹✉, Xing Lv¹✉, Yang Li¹✉, Rongkang Shi¹✉, Yihong Duan¹✉, Shilin Tian^{1,2}✉ and Xingfen Wang¹✉

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



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

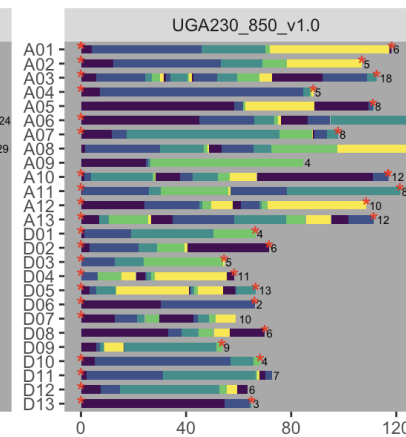
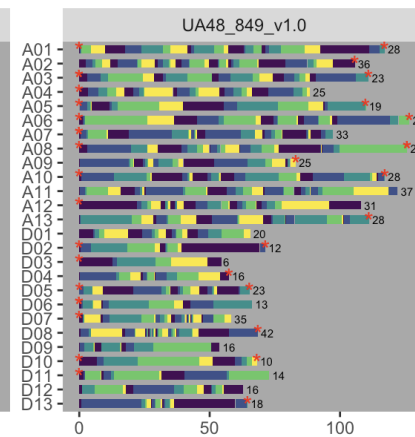
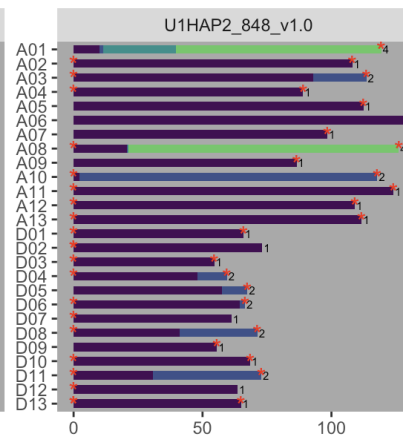
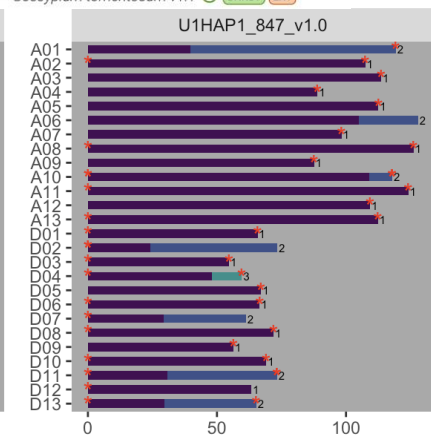
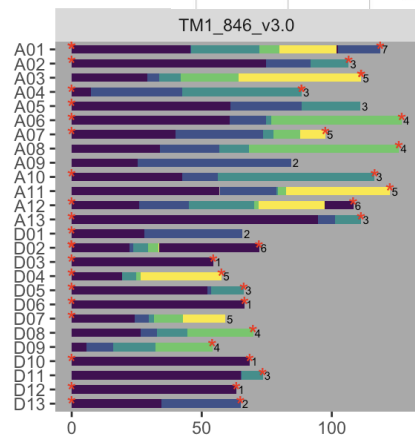
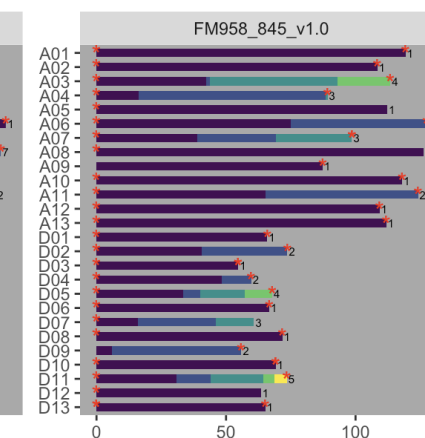
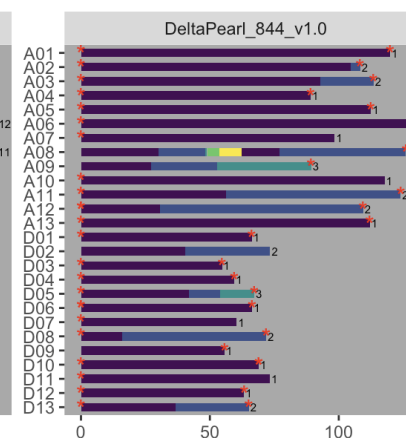
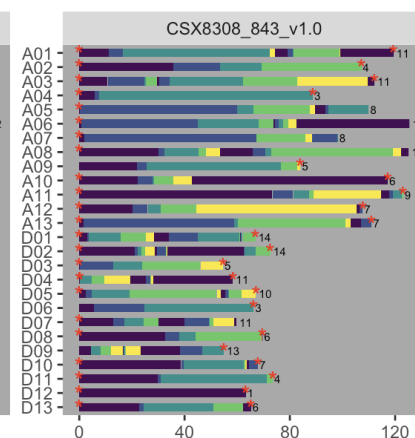
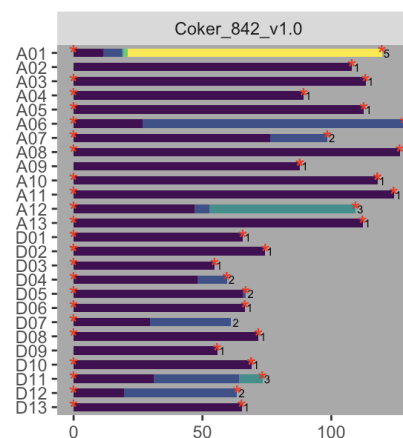
DeltaPearl:
Bayer

FM958:
BASF



Phytozome 13
THE PLANT GENOMICS RESOURCE

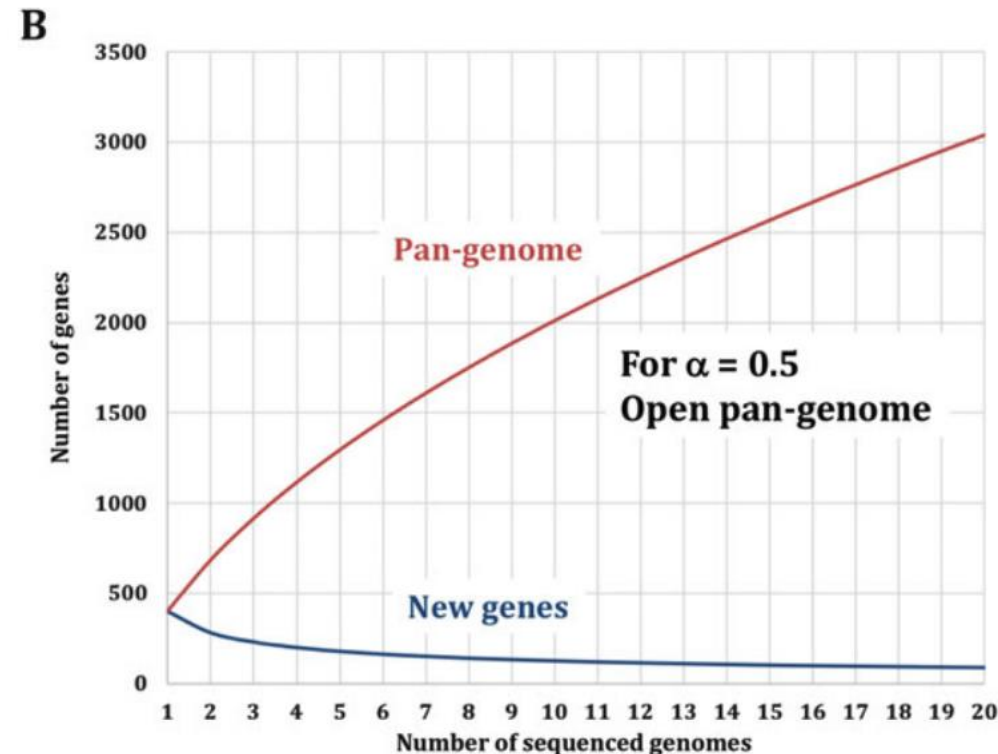
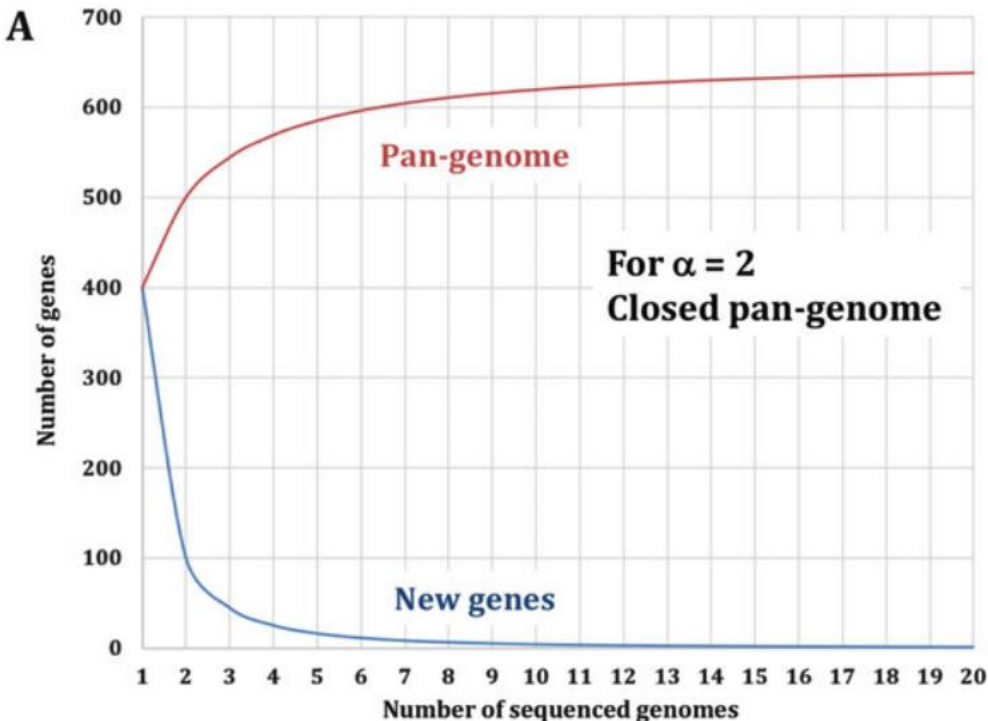
- Cotton
- Gossypium barbadense v1.1 (UNRST EXT)
 - Gossypium darwinii v1.1 (UNRST EXT)
 - Gossypium hirsutum v1.1 (UNRST EXT)
 - Gossypium hirsutum v2.1 (UNRST EXT)
 - Gossypium hirsutum v3.1 (RST EXT)
 - Gossypium hirsutum CSX8308 v1.1 (RST EXT)
 - Gossypium hirsutum Coker v1.1 (RST EXT)
 - Gossypium 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

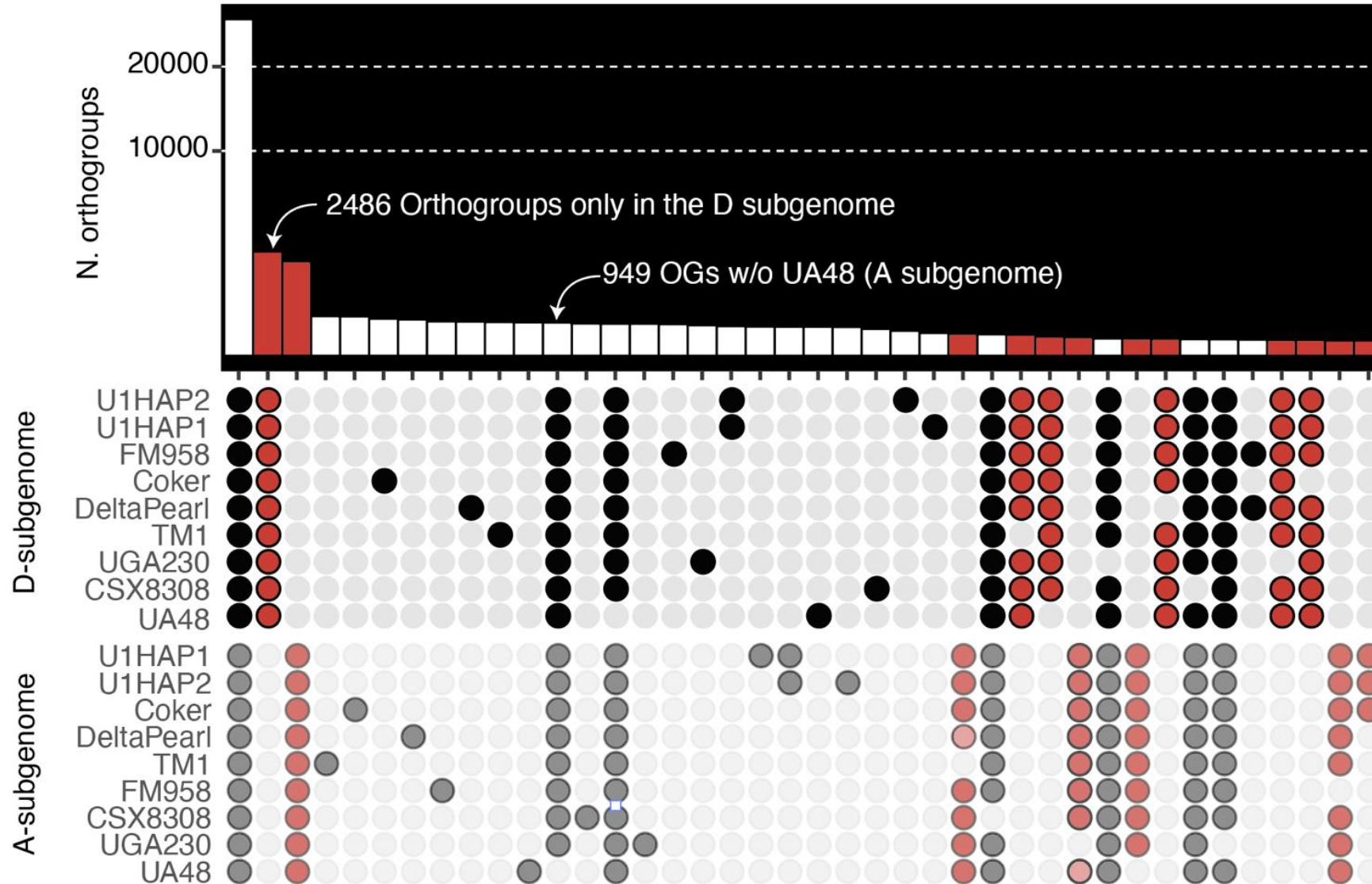


Richard (2020)

Why we need integrated pan-genome resources

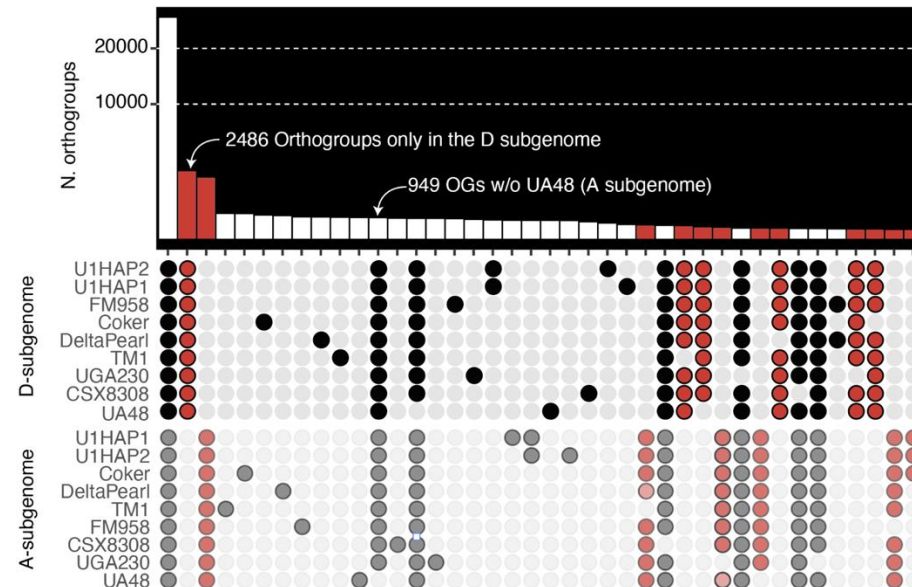
- The idea of “open” vs. “closed” pangenomes comes from work in bacteria
- In plants, genes come and go less frequently
 - But more frequently in polyploids

Why we need integrated pan-genome resources



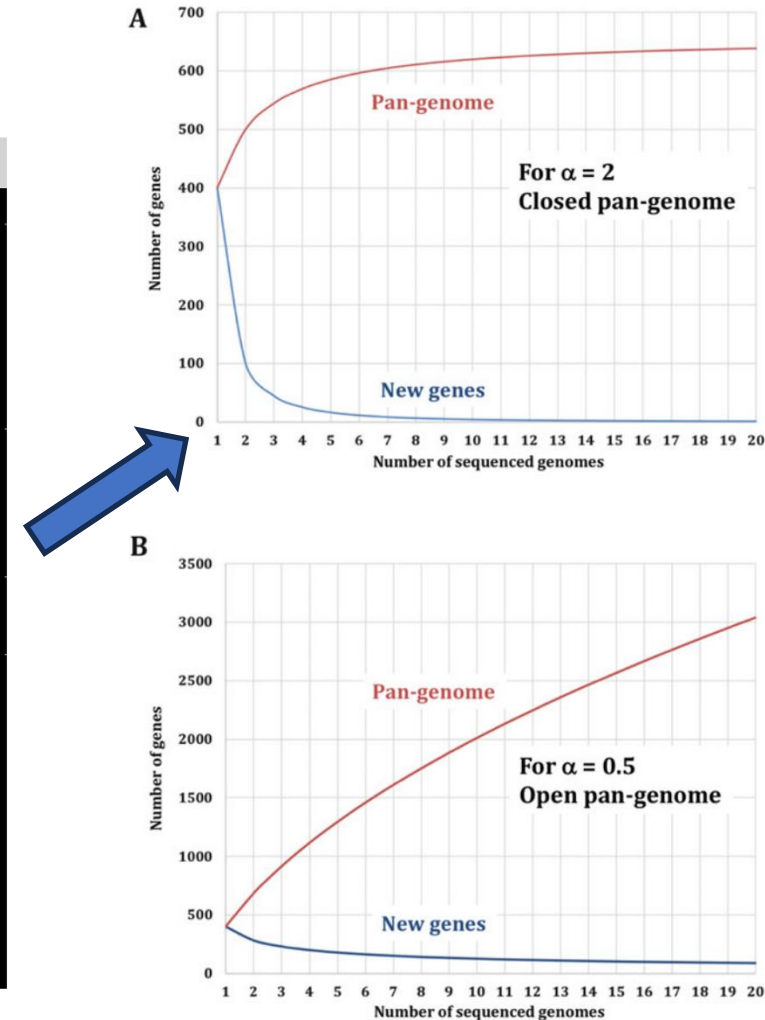
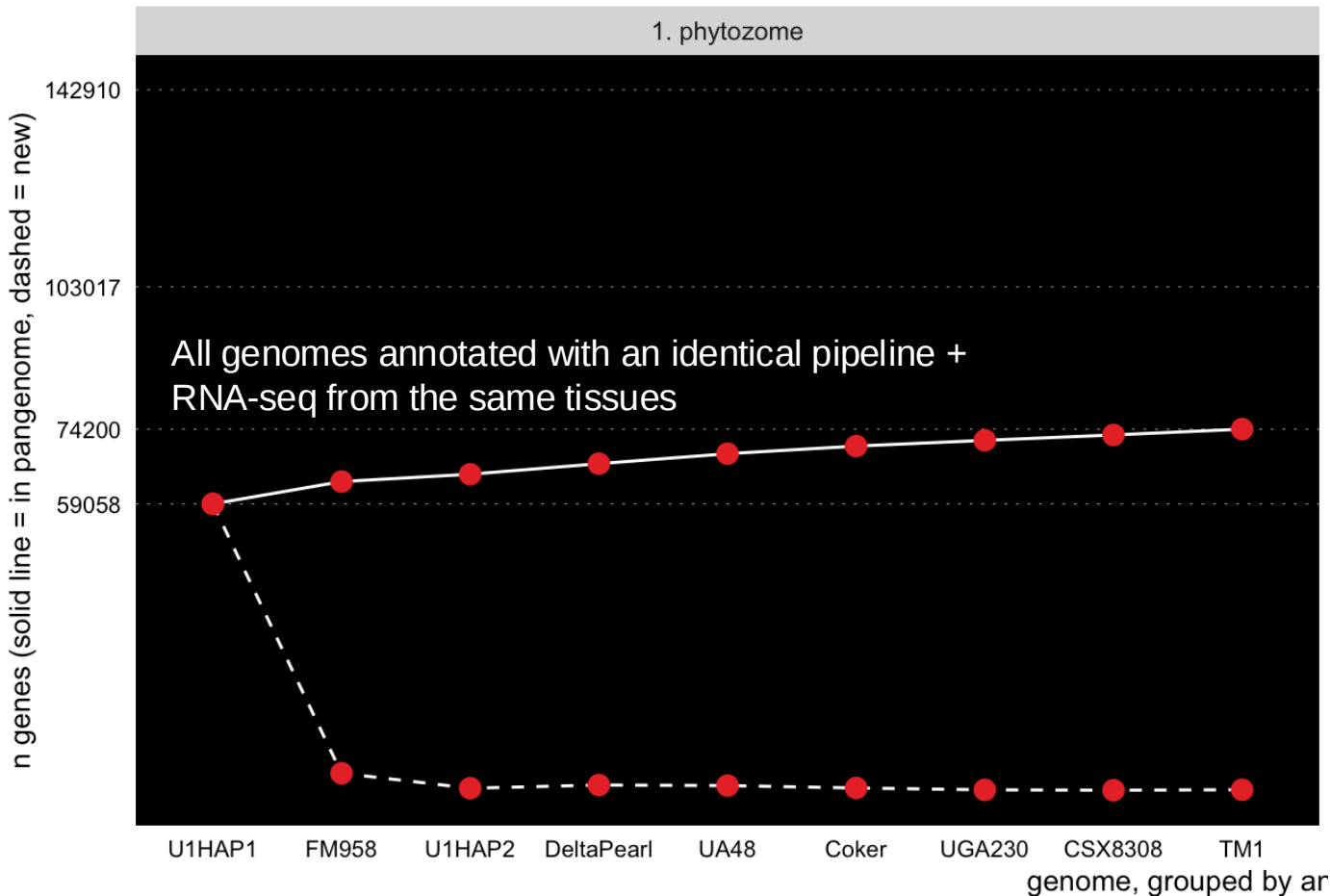
Why we need integrated pan-genome resources

- 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*



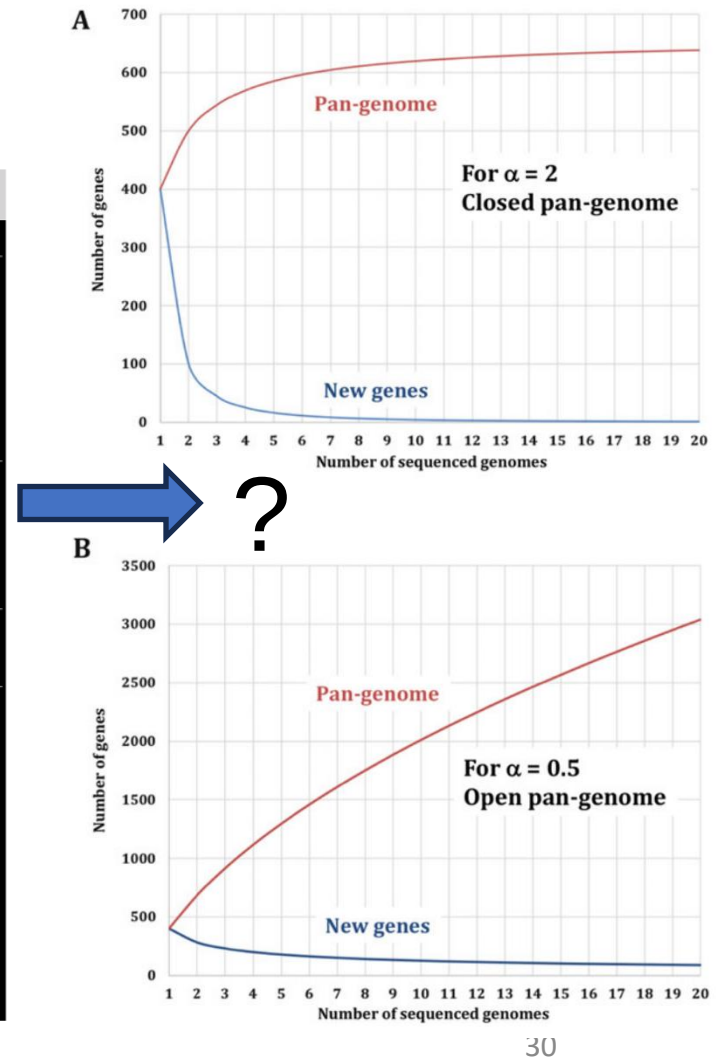
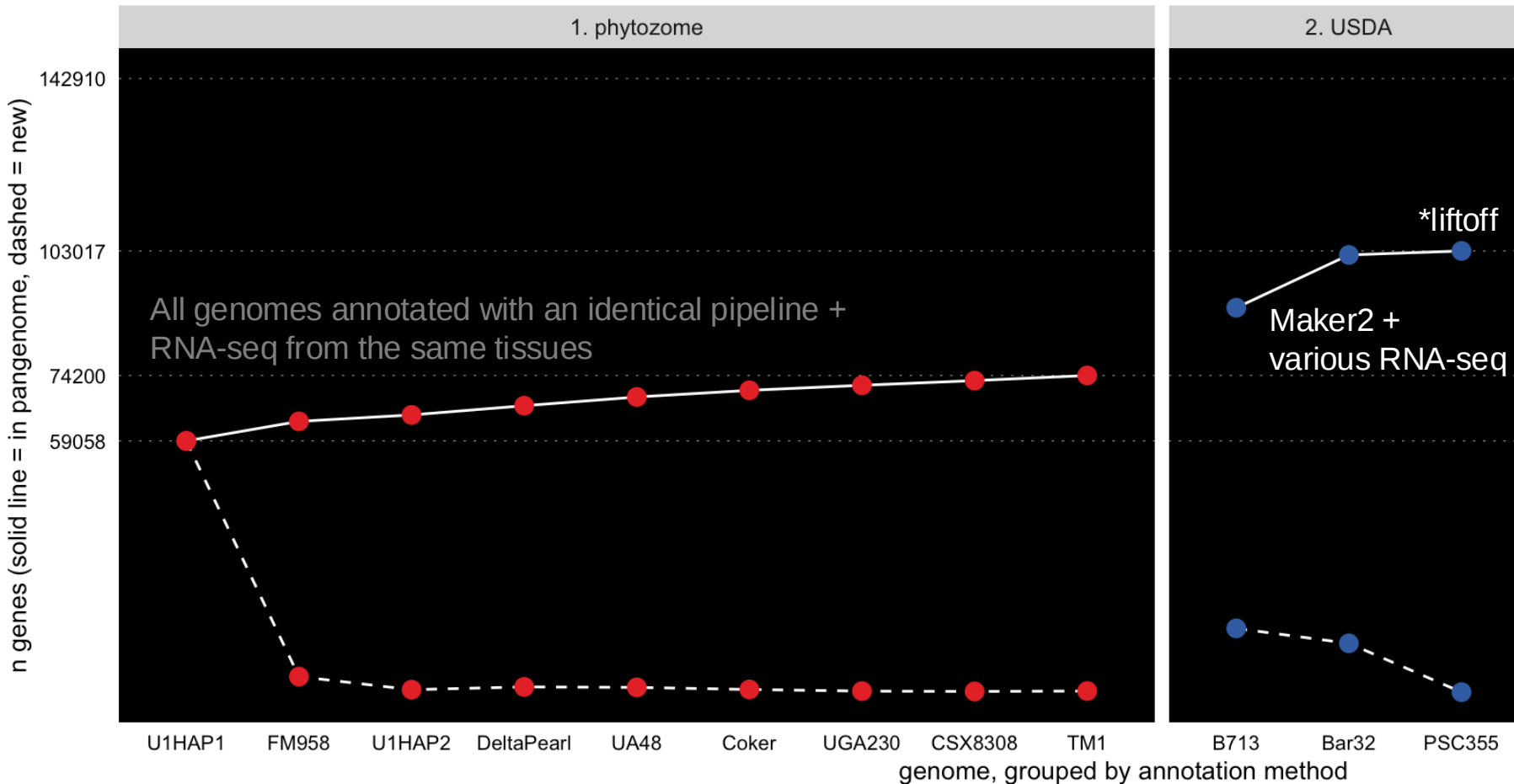
Why we need integrated pan-genome resources

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



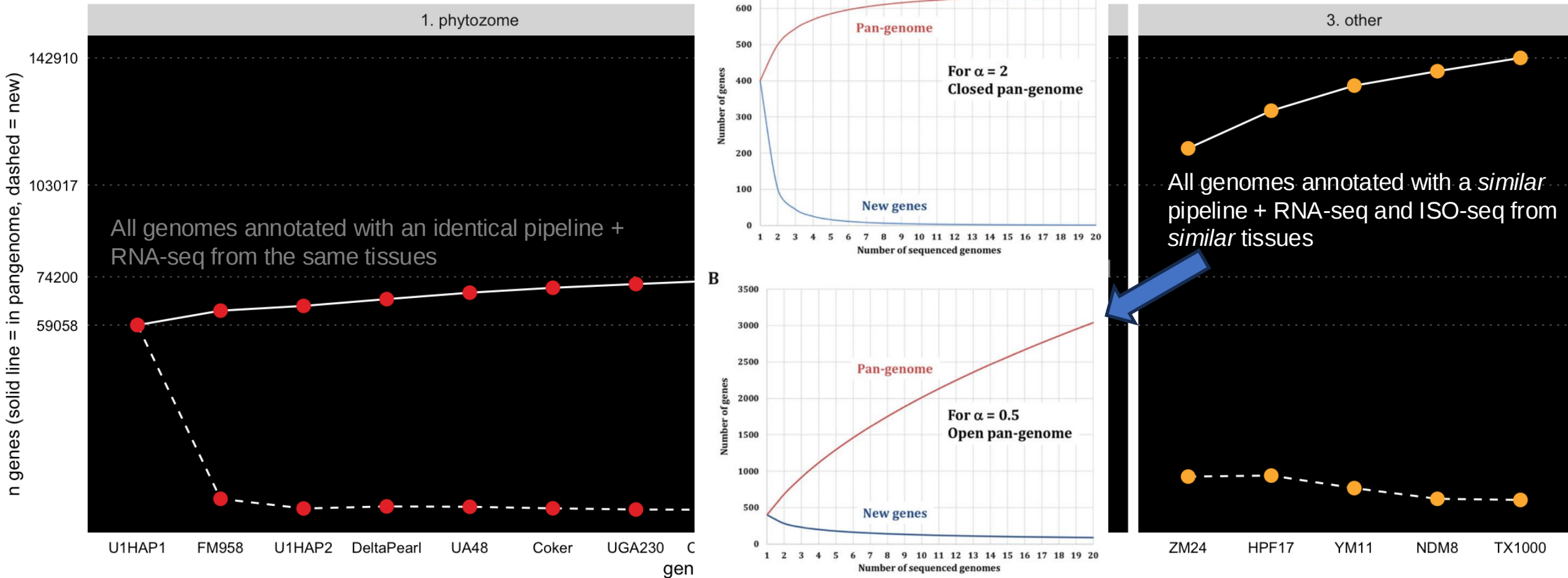
Why we need integrated pan-genome resources

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



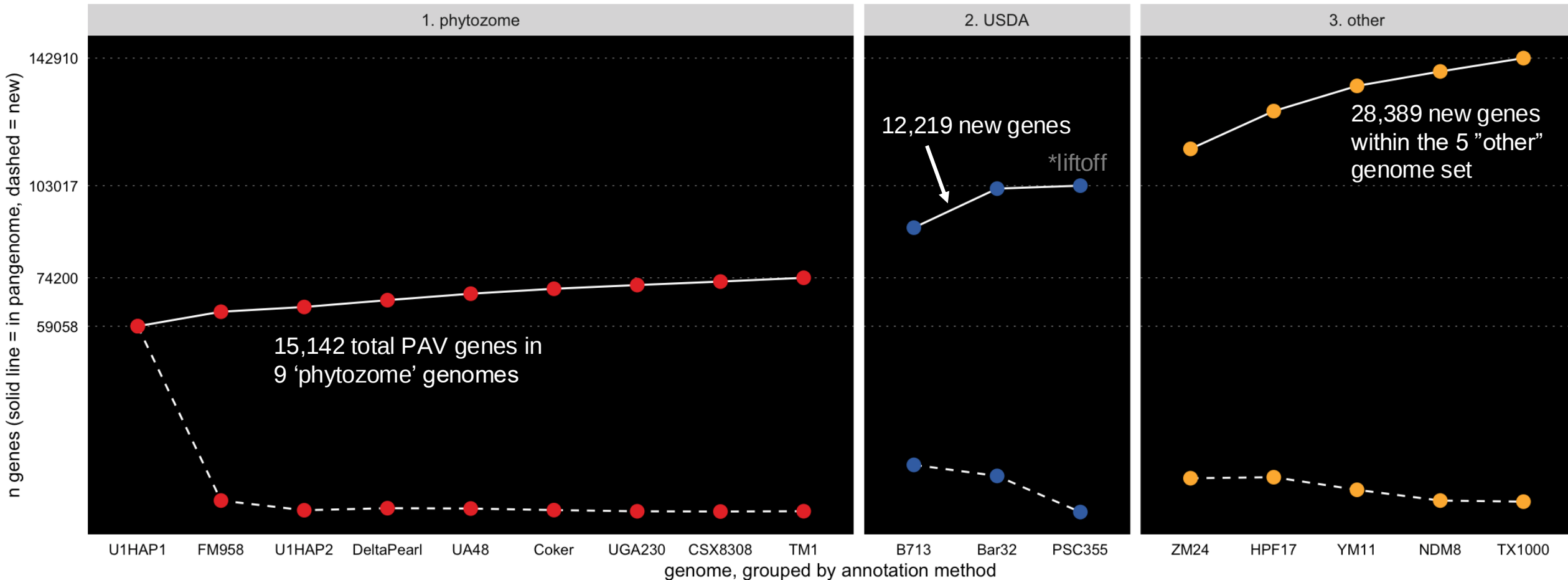
Why we need integrated pan-genome resources

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



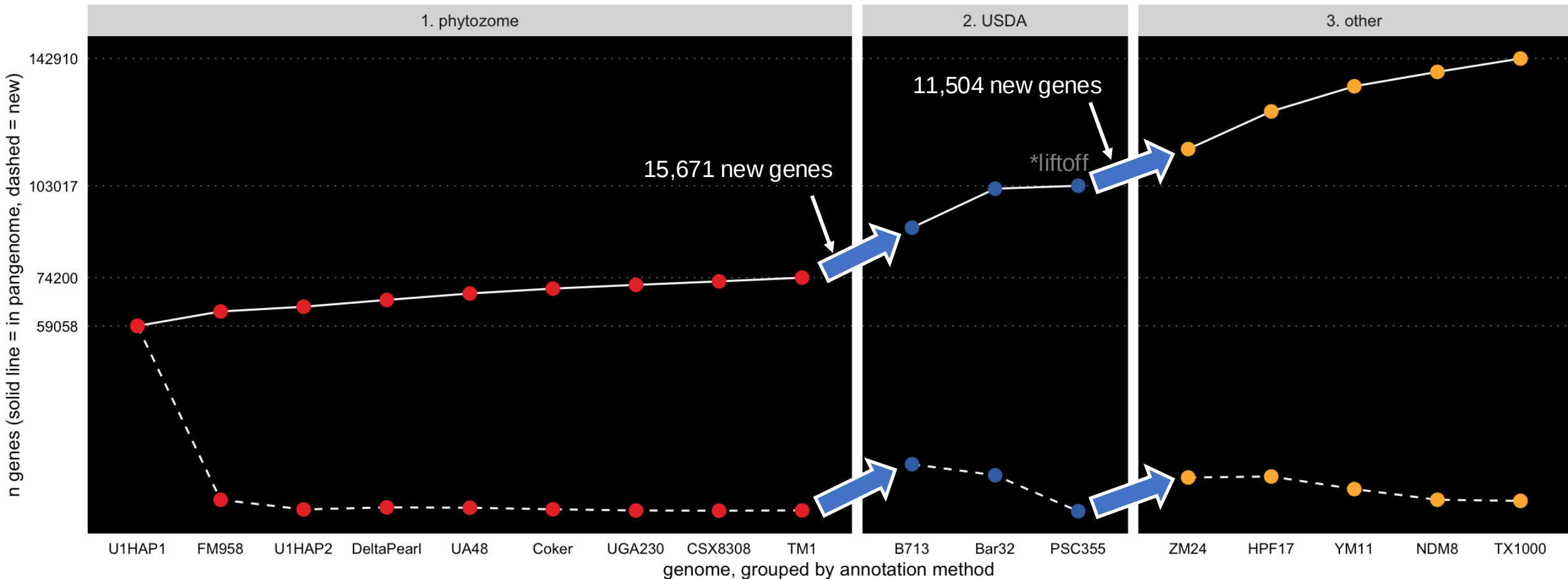
Why we need integrated pan-genome resources

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



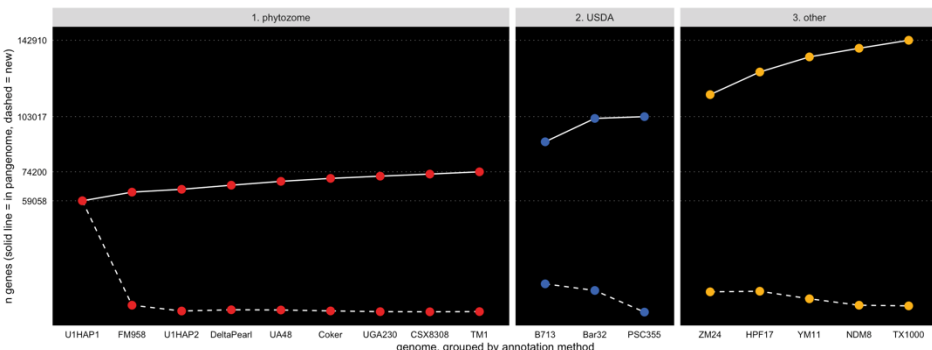
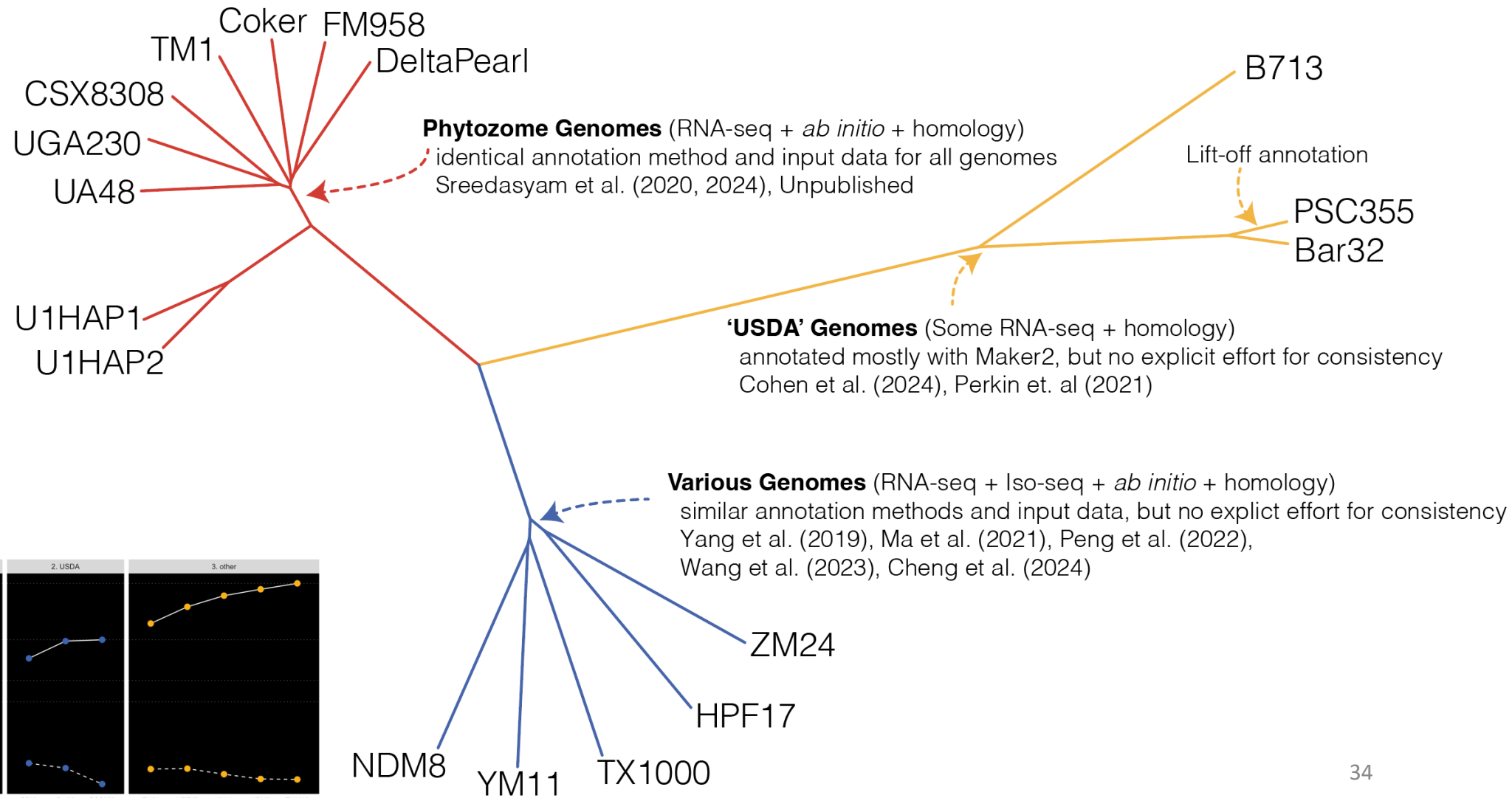
Why we need integrated pan-genome resources

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



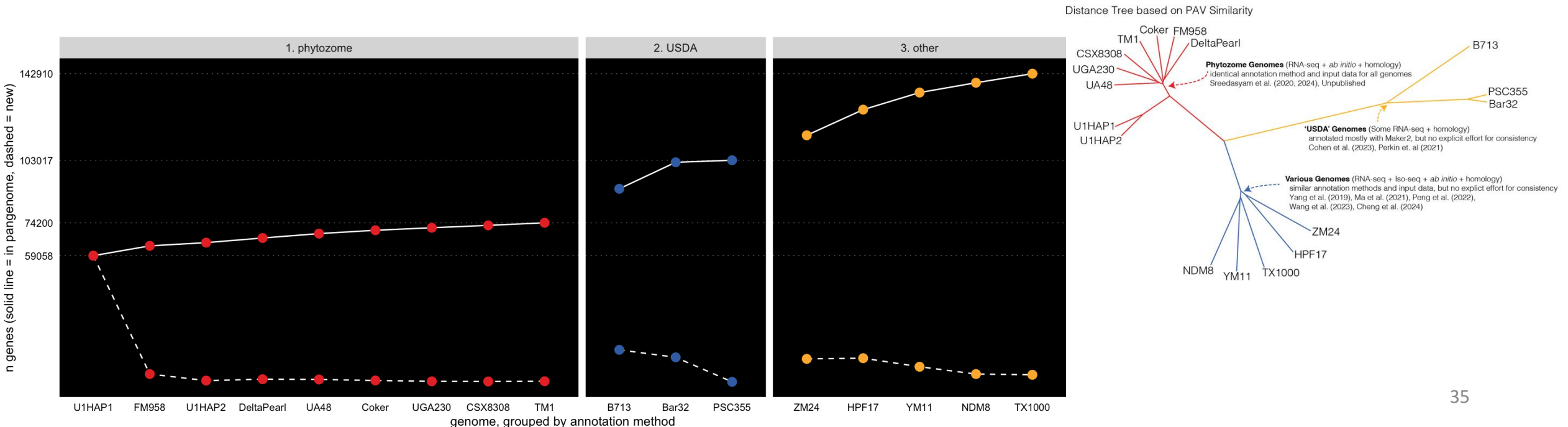
Why we need integrated pan-genome resources

Distance Tree based on PAV Similarity



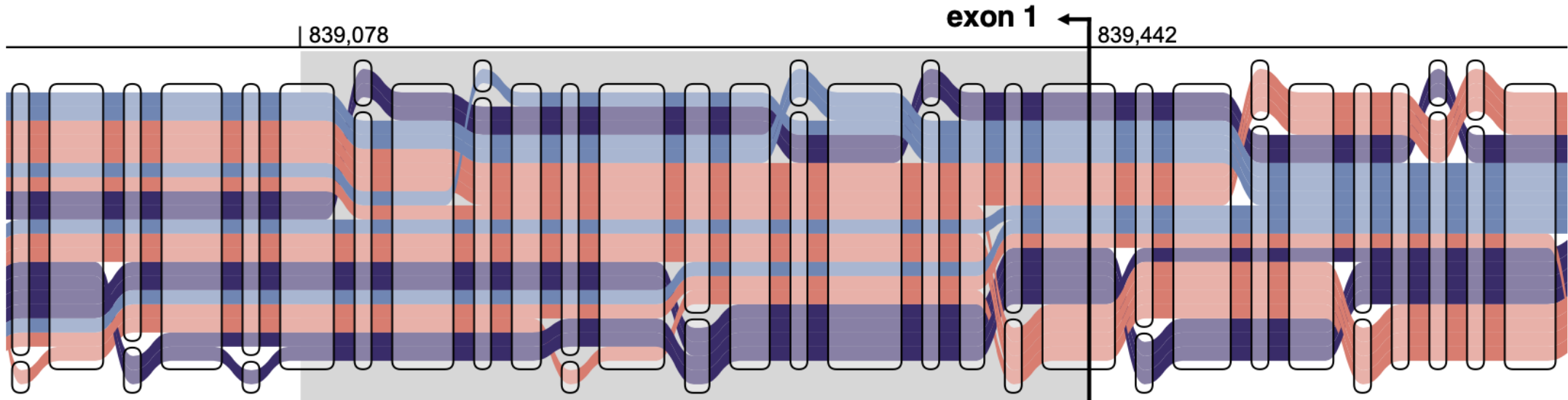
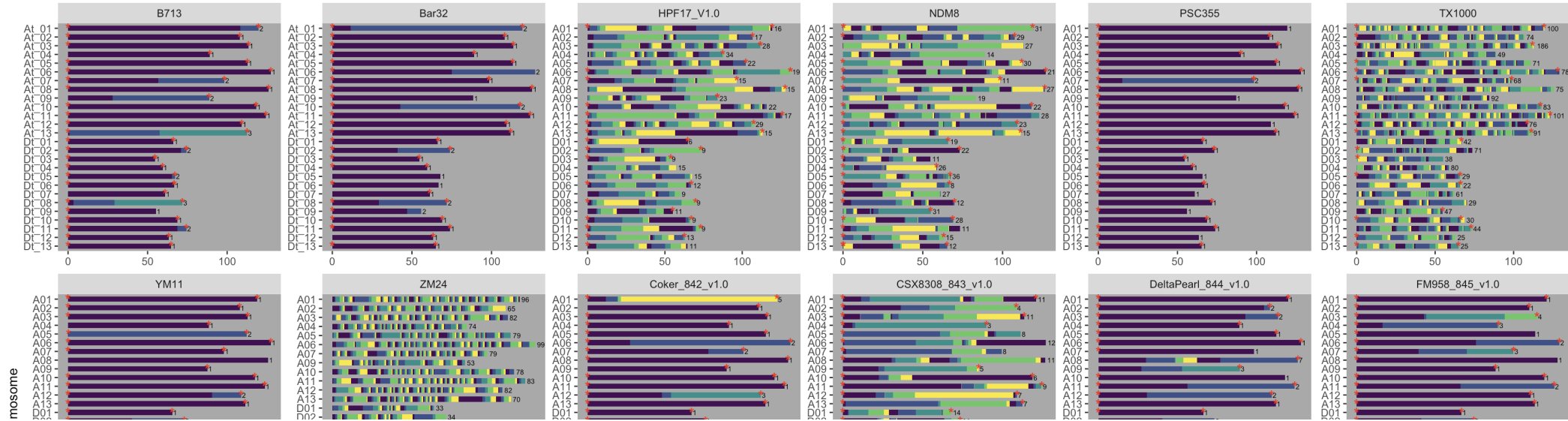
Why we need integrated pan-genome resources

- 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



From multiple genomes to a “graph” pangenome

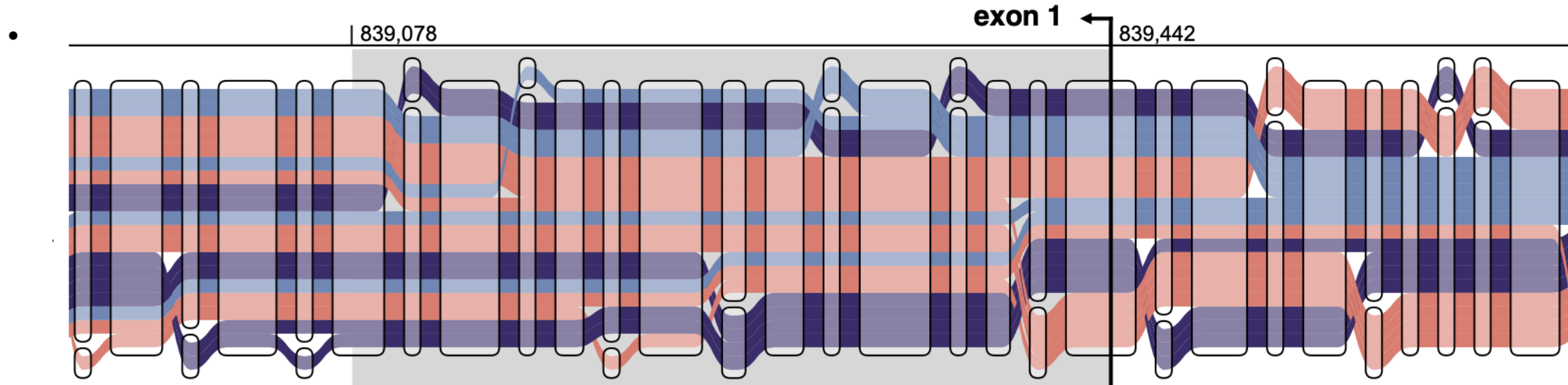
Contig positions: each cycle through colors is 5 contigs (4 gaps)



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

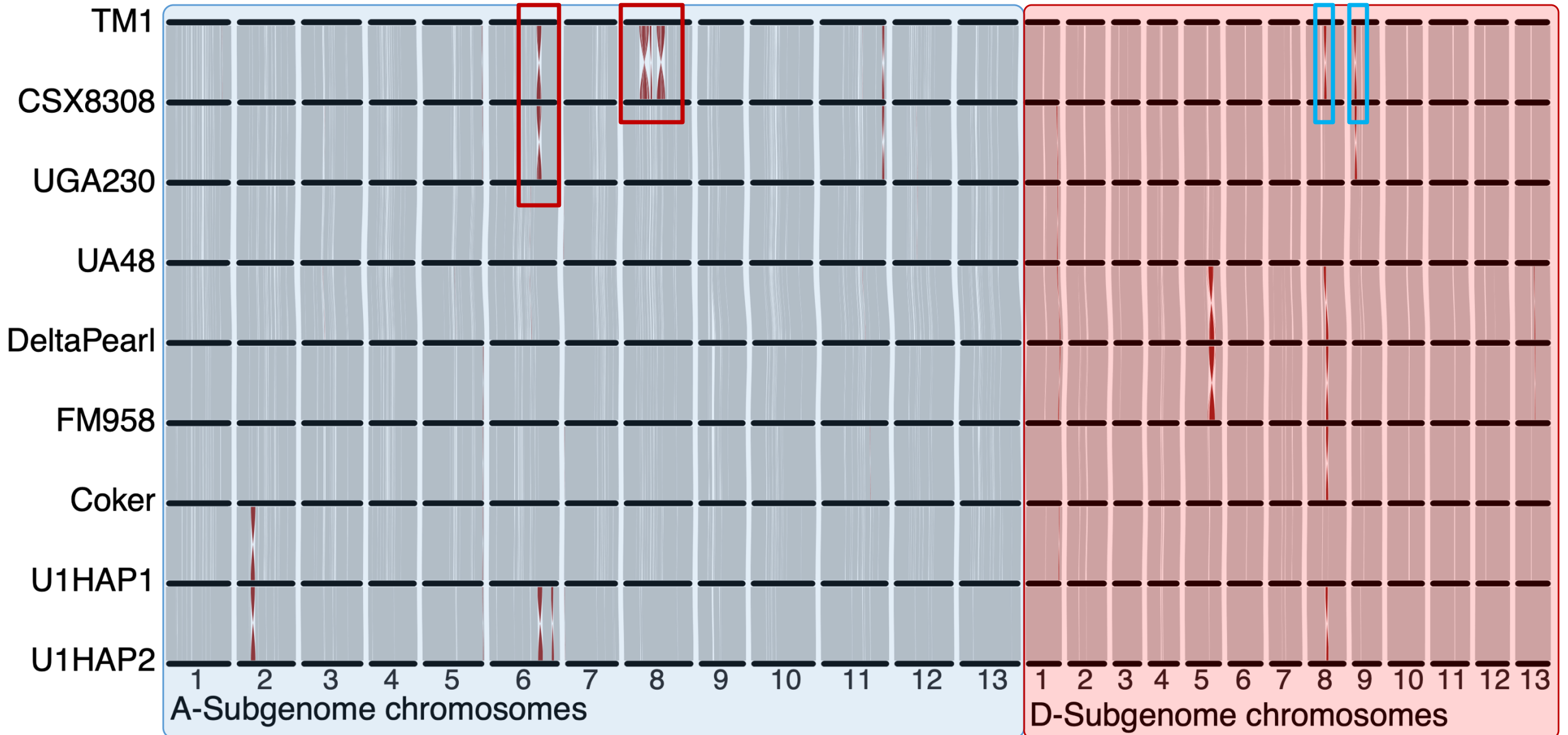
YOU MUST CHOOSE



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'?



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'?

Upland Cottons Pangenome Home Gene expression Genomic variations Genic space Help Contact



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.

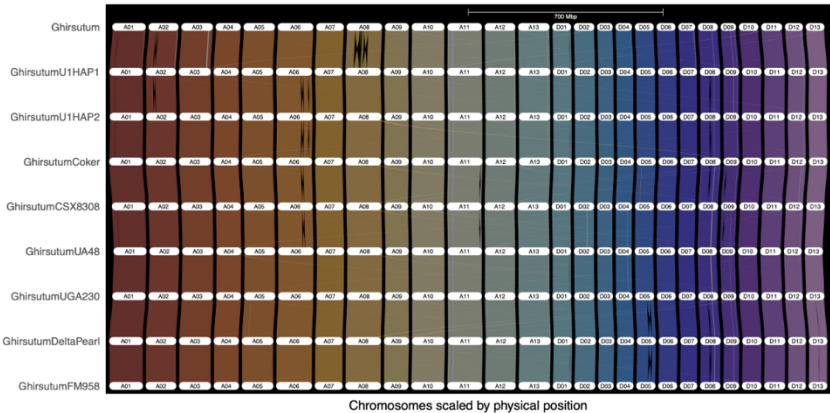


<https://tinyurl.com/uplandcottonspangenome>

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							
Copy number	Genome	chr	start	end	Cultivars												
					UGA230	UA48	U1HAP2	U1HAP1	FM958	DeltaPearl	Coker	CSX8308					
MULTI																	
						</											

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

Gossypium hirsutum var. U1HAP1 v1.2

Gossypium hirsutum var. U1HAP2 v1.2

Gossypium hirsutum var. CSX8308 v1.2

Gossypium hirsutum var. Coker v1.2

Gossypium hirsutum var. UA48 v1.2

Gossypium hirsutum var. DeltaPearl v1.2

Gossypium hirsutum var. TM1 v3.2

Gohir.1Z049314

Location: scaffold_246:62136..64180 reverse

Defline:

DataVisualizationOrthologs

Show10entries

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

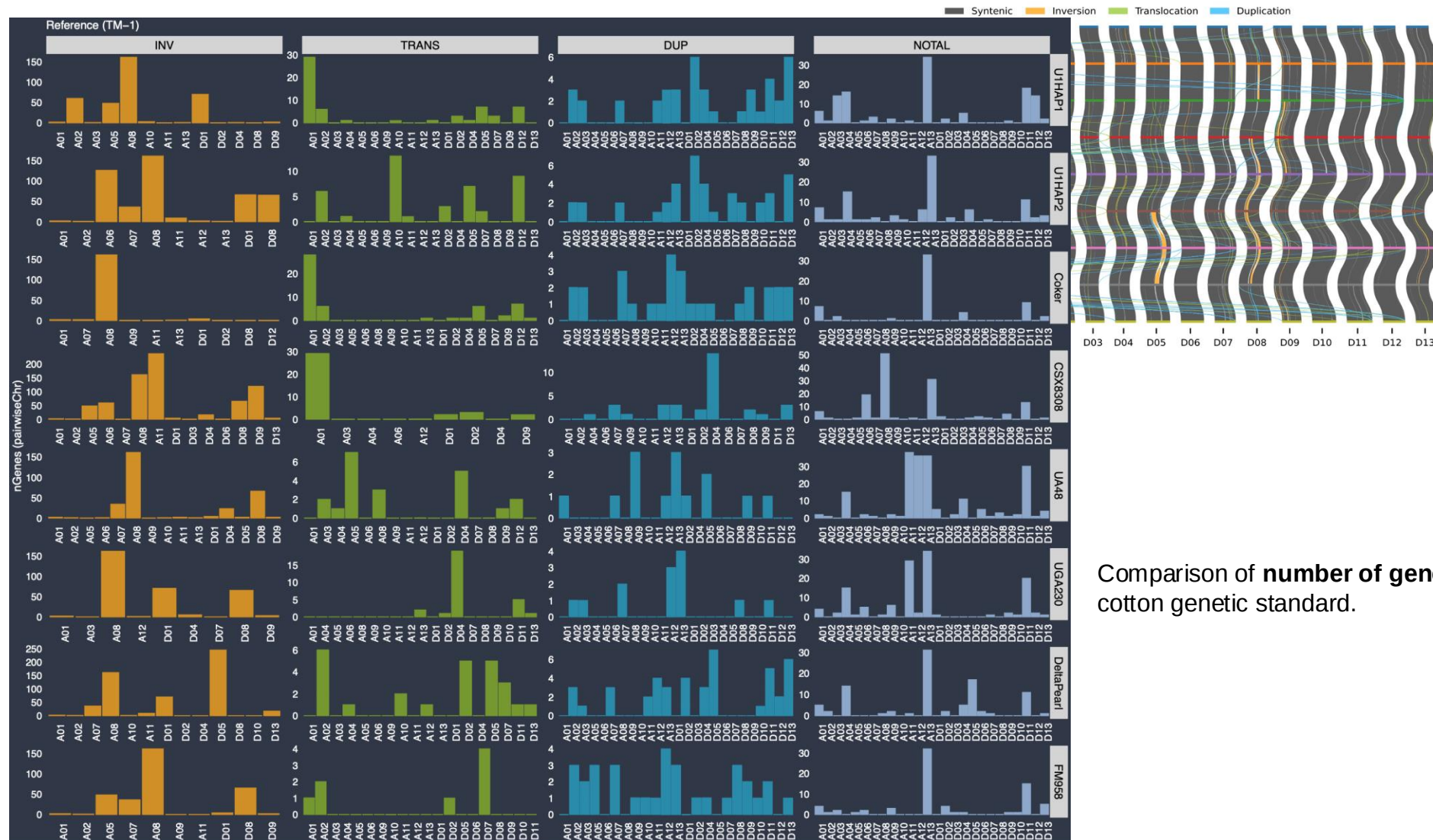
Showing 1 to 10 of 71,247 entries

Previous12345...7125Next

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

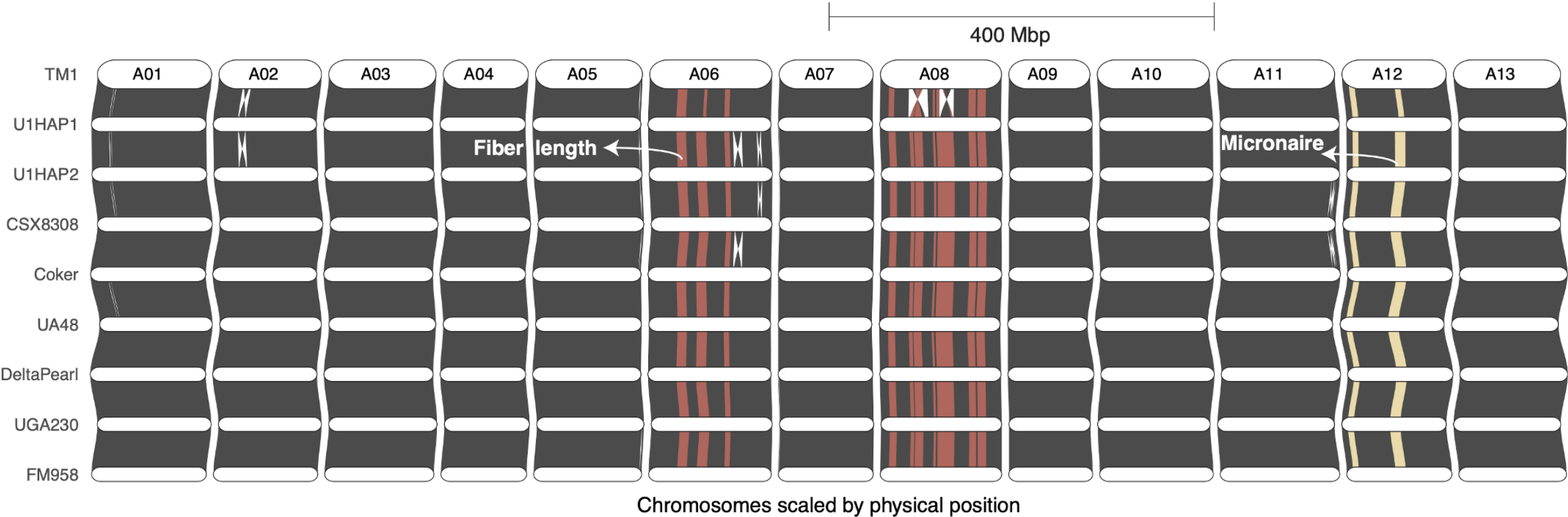


Comparison of **number of genes** affected by LSVs relative to TM-1, cotton genetic standard.

What can we learn from a ‘pangenome’?

- 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)

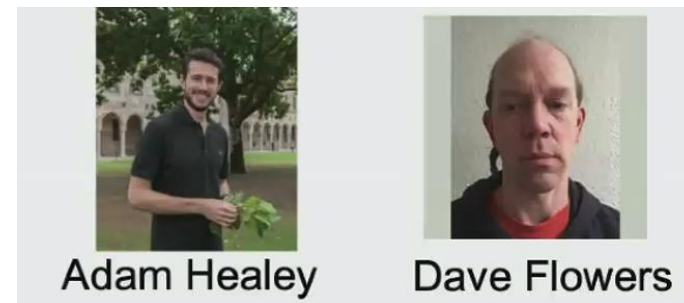
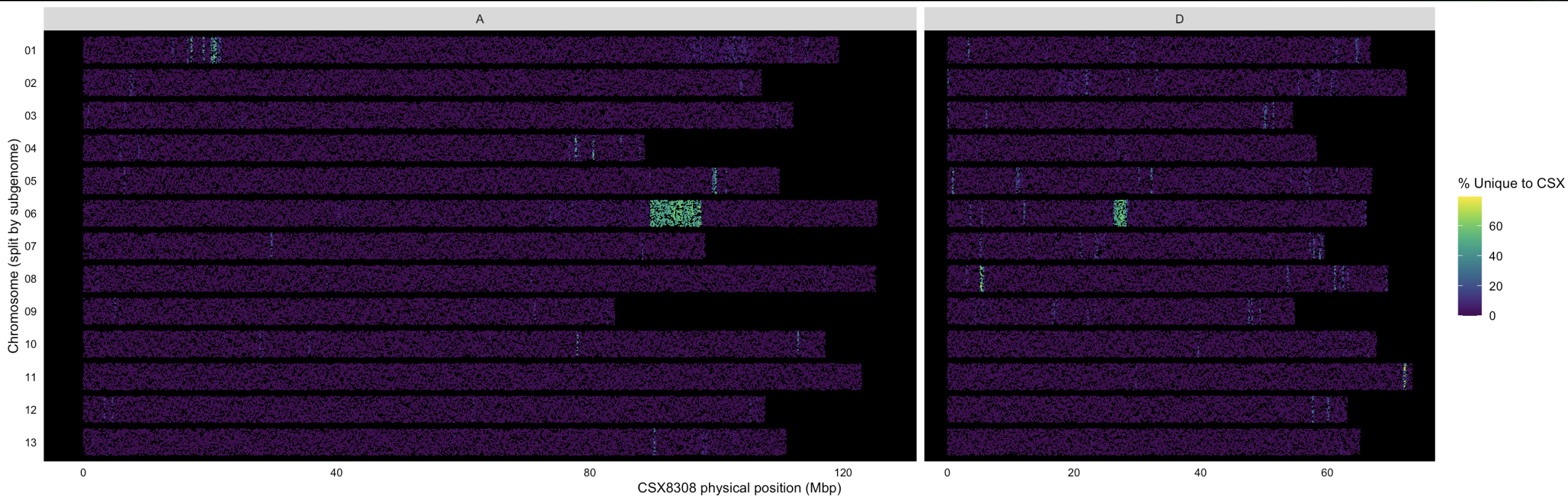
Tracking QTL across multiple genomes



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

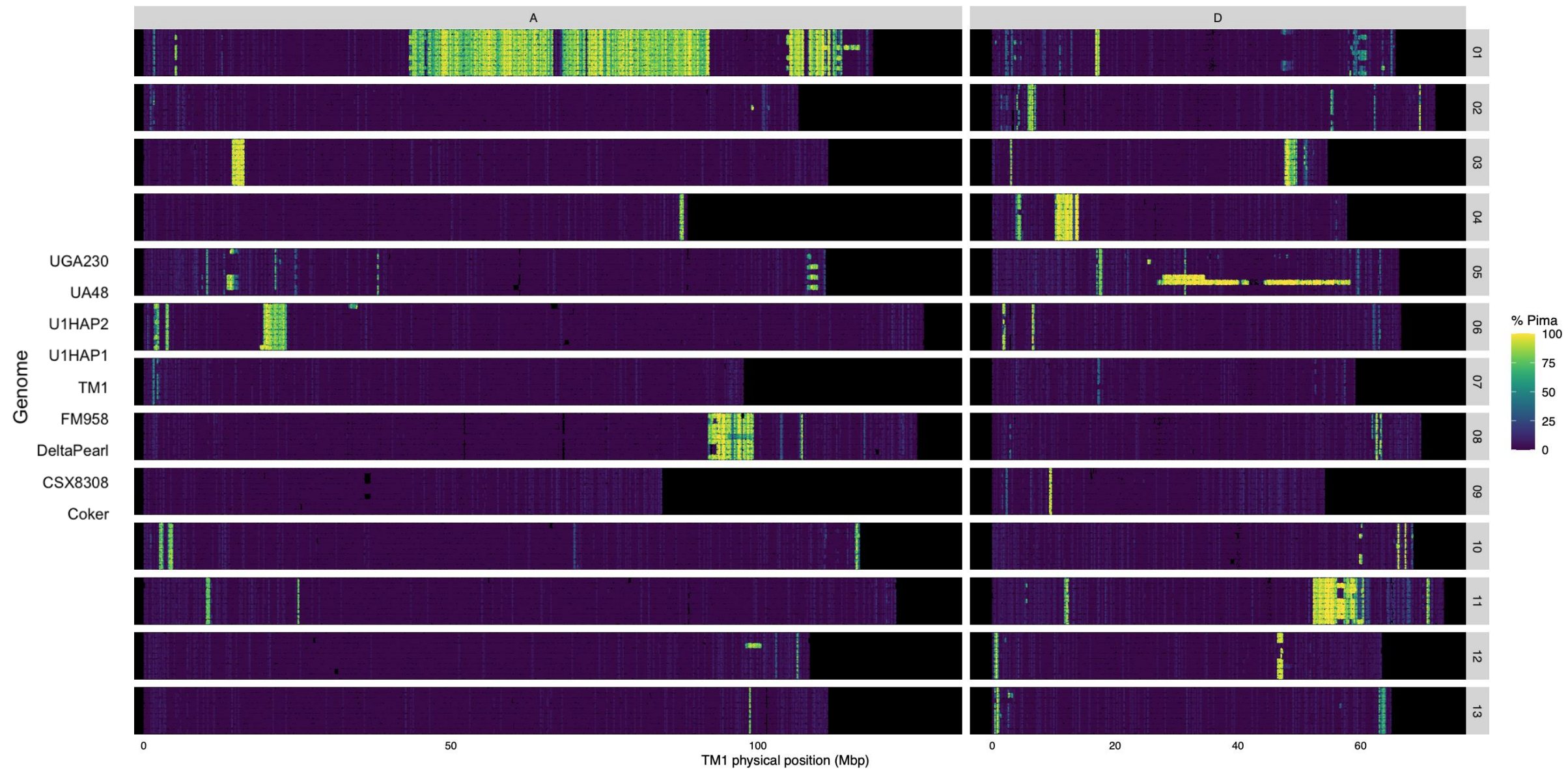


What can we learn from a ‘pangenome’?

- 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
- Test for **regions of co-ancestry** among genomes: example Pima cotton introgressions

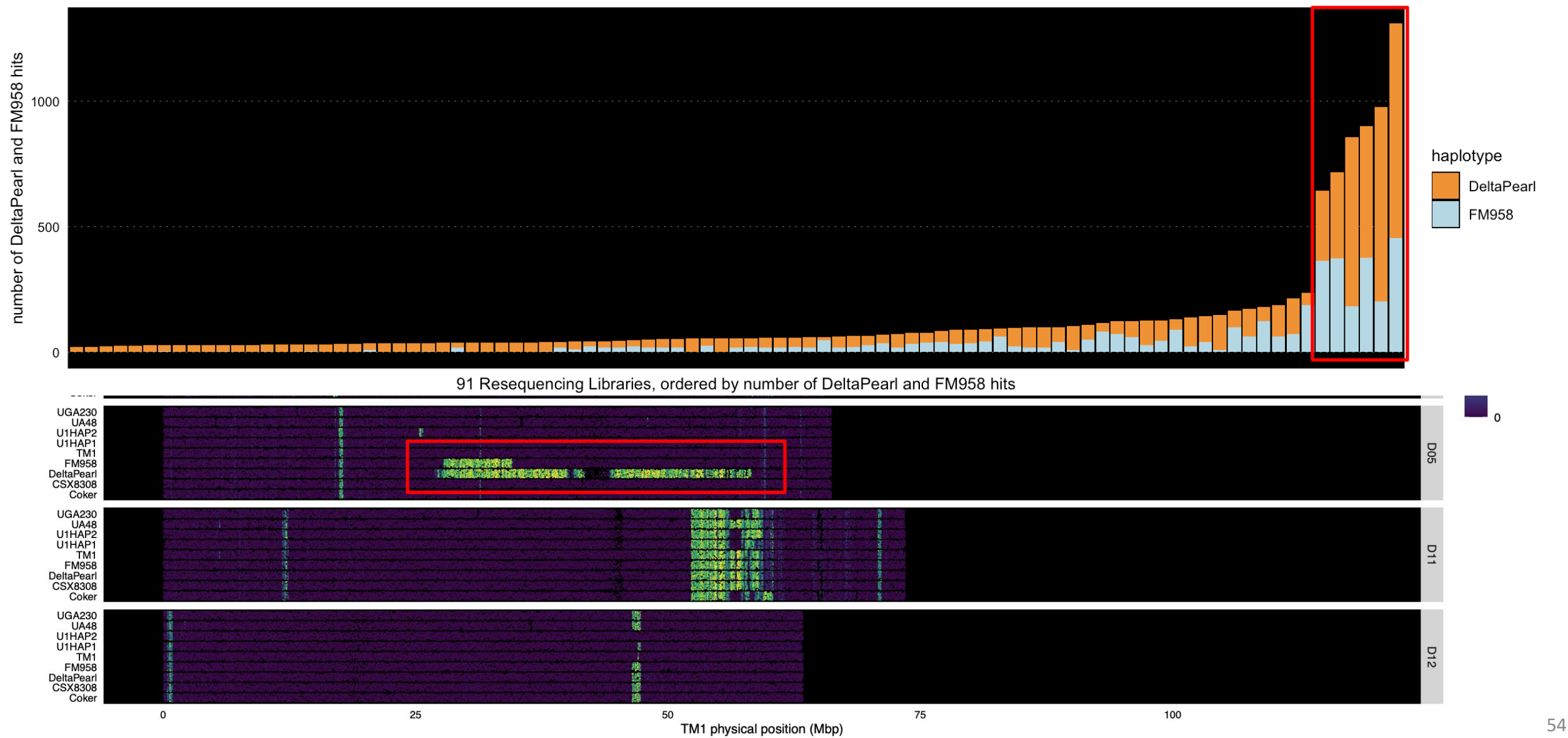
Tracking Pima introgressions in upland cotton

UGA230



Tracking Pima introgressions in upland cotton

UGA230



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.)
 - ...