Kernza Genomics

Kevin Dorn
Postdoctoral Fellow
Poland Lab | Kansas State University
dorn@ksu.edu | www.kdorn.com

This talk is Twitter friendly!

@Dornomics

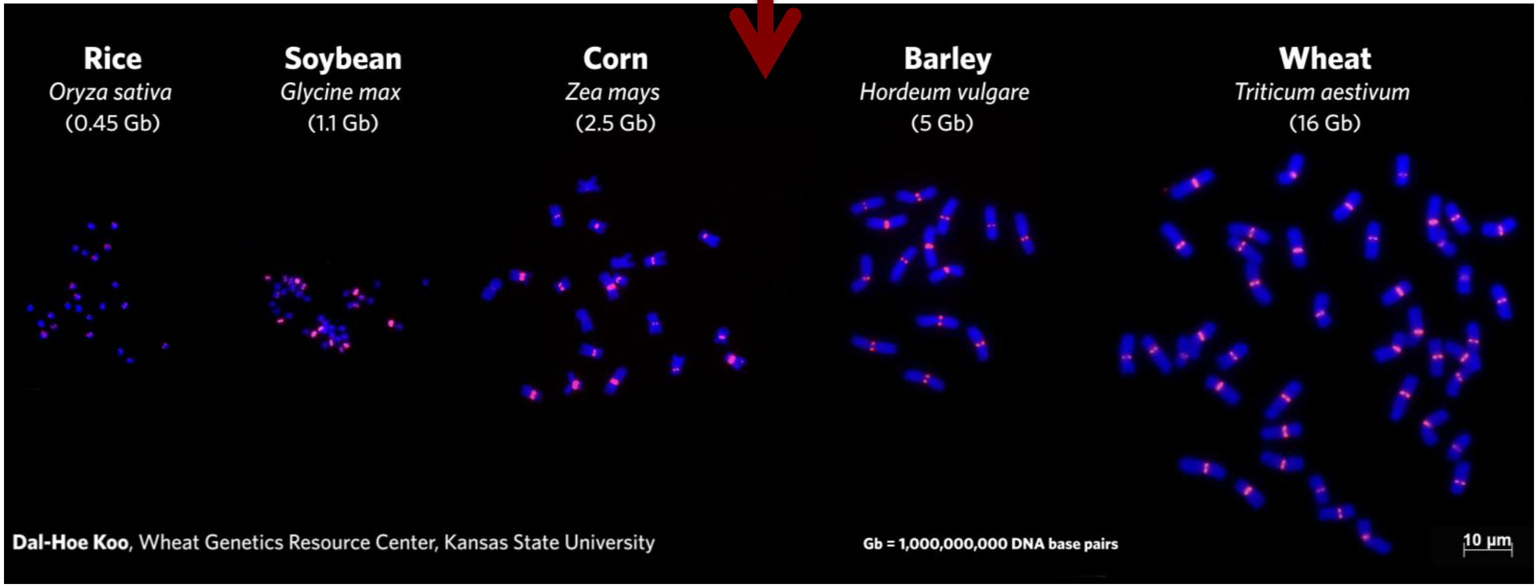
KANSAS STATE
UNIVERSITY

Kernza Genomics

The problem: Plant genomes are huge and complex

Human
3 Gb

Kernza
25 Gb



Dal-Hoe Koo, Wheat Genetics Resource Center, Kansas State University

Gb = 1,000,000,000 DNA base pairs

New DNA sequencing tools have made utilizing genomics in species like Kernza possible

The (genomic) challenge with intermediate wheatgrass…

- Segmental Allohexaploid (unclear progenitors)
- 2n=6x=42
- 1C=12.6 Gb
- Obligate outcrosser

"…but that genome will be impossible!"

- Enable functional studies of key traits
- Increase marker quality/quantity for breeding

KANSAS STATE UNIVERSITY

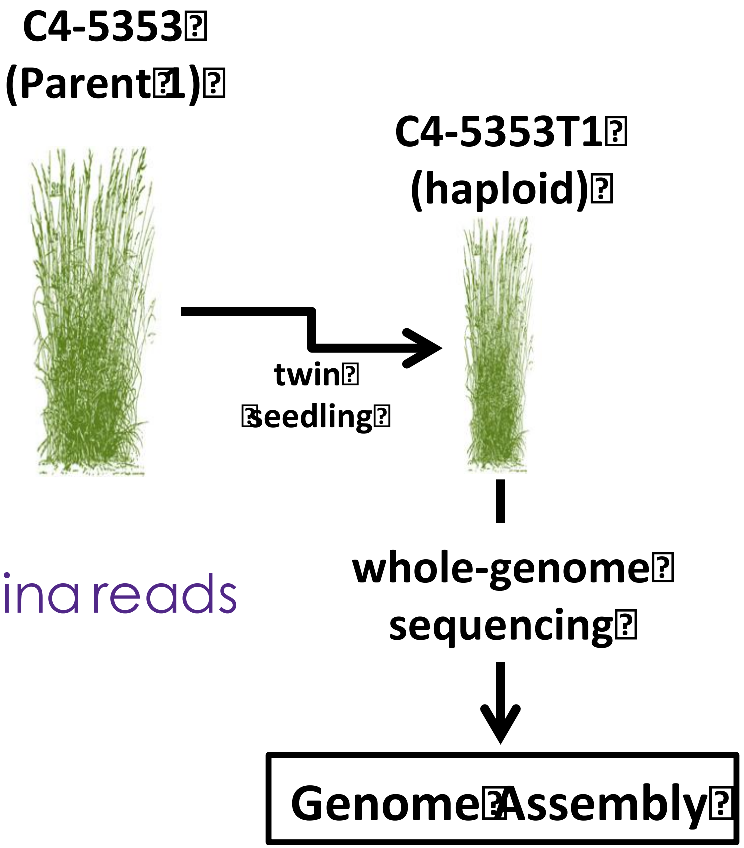
The genome of *Thinopyrum intermedium*

1.) Sequencing, assembly, anchoring/ordering of genome

2.) Parsing subgenomes & deciphering diploid progenitor history

3.) Biology of perenniality in the Triticeae

Genome sequencing



C4-5353
(Parent 1)

C4-5353T1
(haploid)

twin
seedling

whole-genome
sequencing

Genome Assembly

31 ***Billion*** Illumina reads

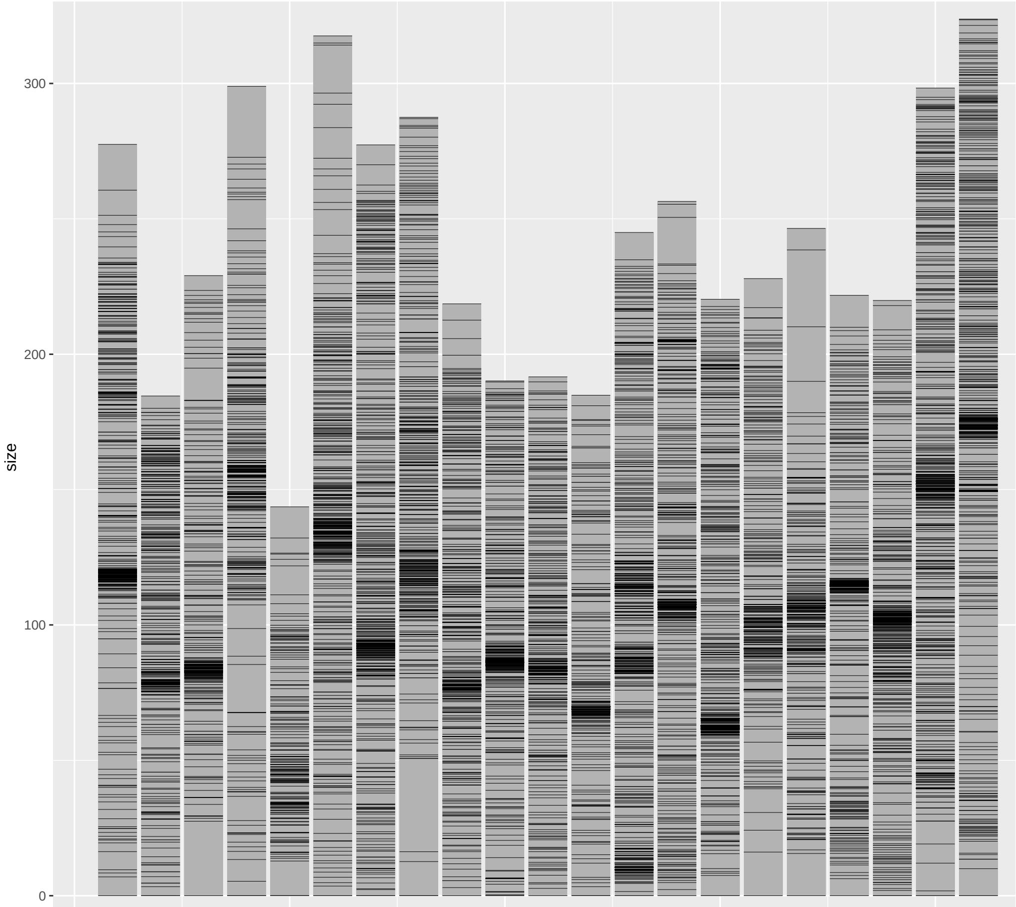
Haploid plant derived from twin seedling by The Land Institute

Genome assembly

NRGene DeNovoMAGIC™ 2.0

	Scaffold	Contig
Number	**237,622**	1,059,461
Total Assembly Length (bp)	11,605,214,228 (**92%**)	11,193,042,853
N50 (bp)	2,794,550 bp	26,238 bp
L50	1,072 scaffolds	118,697 contigs

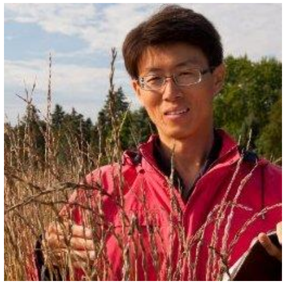
Anchoring/Ordering to Linkage Map



21 linkage groups
10,029 total markers
5,061 cM

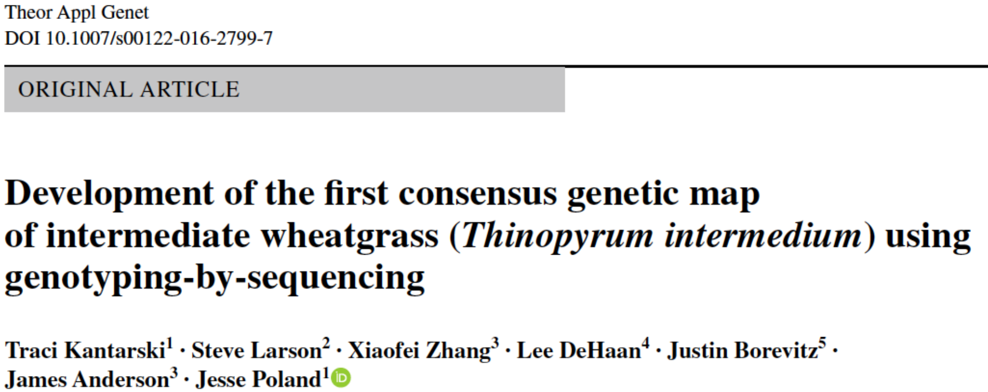
3 breaks made in NRgene assembly based on consensus map

Traci Kantarski
Columbia Univ.



Xiaofei Zhang
Univ. of MN

Steve Larson
USDA-ARS



Theor Appl Genet
DOI 10.1007/s00122-016-2799-7

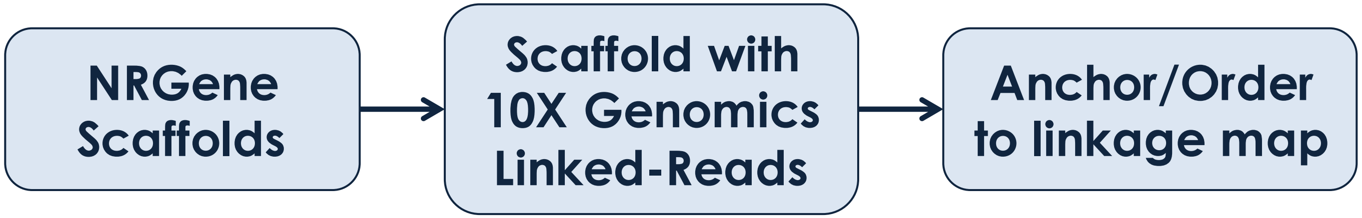
ORIGINAL ARTICLE

Development of the first consensus genetic map of intermediate wheatgrass (*Thinopyrum intermedium*) using genotyping-by-sequencing

Traci Kantarski[1] · Steve Larson[2] · Xiaofei Zhang[3] · Lee DeHaan[4] · Justin Borevitz[5] · James Anderson[3] · Jesse Poland[1]

Scaffolding with 10X Genomics Linked-Reads

10X coverage with 2x150 bp linked-reads


NRGene Scaffolds → Scaffold with 10X Genomics Linked-Reads → Anchor/Order to linkage map

	Pre-10X Scaffolding	Post-10X Scaffolding
Total Bases (Gb)	11.194	11.177
Scaffold N50 (Mb)	2.8	4.2
Contig N50 (Kb)	26.3	26.2
Anchored Bases (Gb)	6.623	7.539

8 breaks made after scaffolding & anchoring to map
All breaks made to correct mis-scaffolding

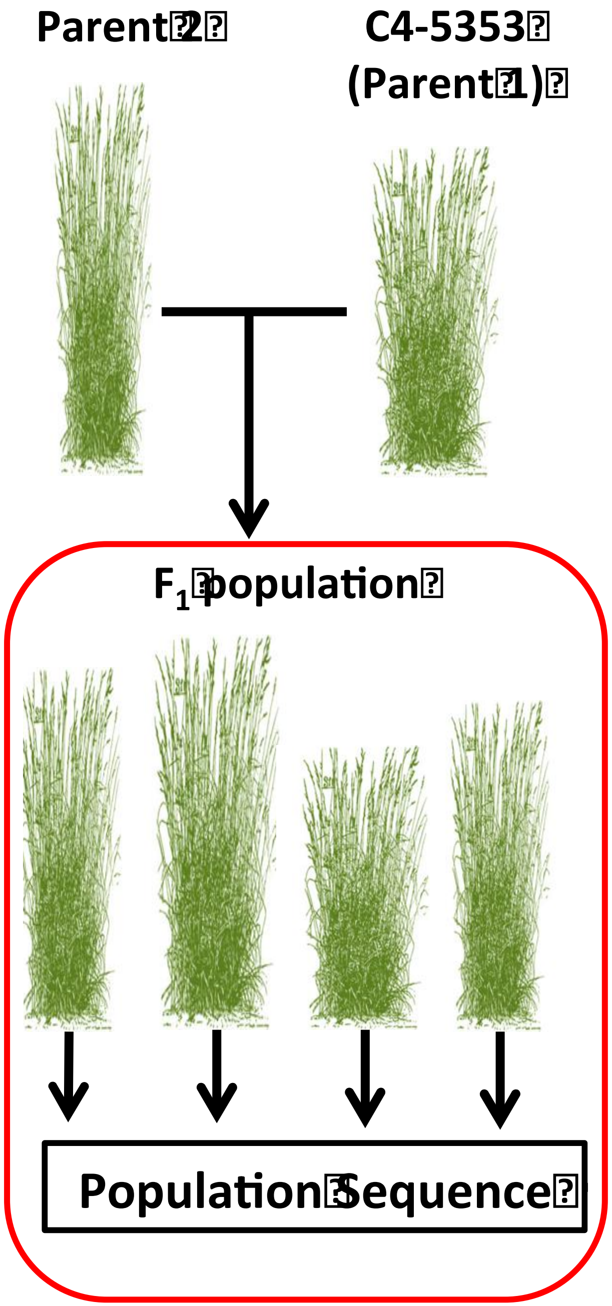
HUDSONALPHA
INSTITUTE FOR BIOTECHNOLOGY


Jeremy Schmutz


Jerry Jenkins

KANSAS STATE
UNIVERSITY

Population Sequencing (POPSEQ)



Parent 2

C4-5353
(Parent 1)

F_1 population

Population Sequence

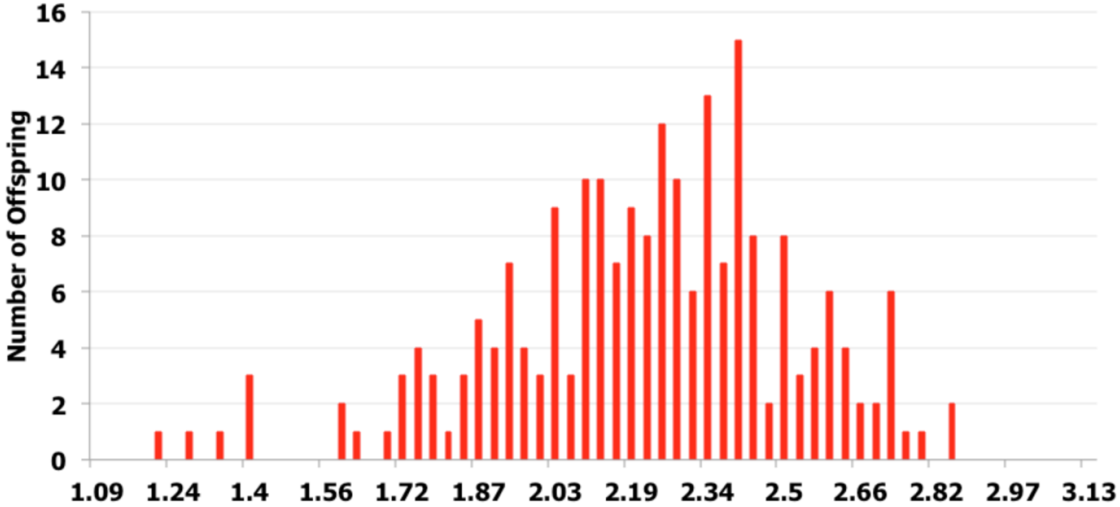
A total of 350,885 75mer markers were generated

793 breaks made in assembly, or a break every 16Mb

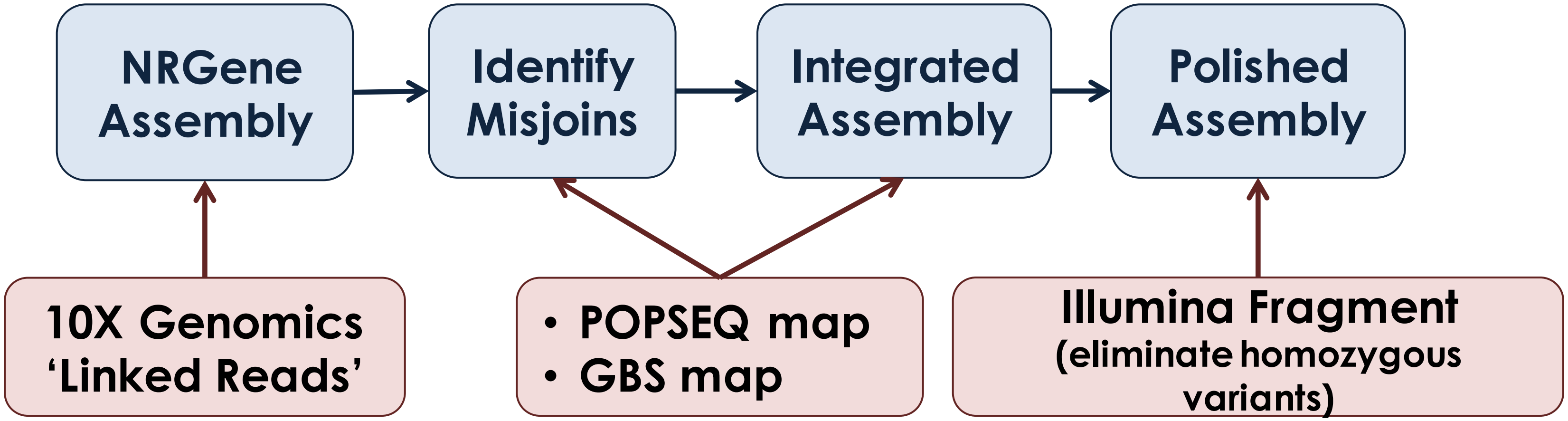
<u>Two F1 populations</u>

C4-5353 x C4-8134 (92 F1 individuals)
C4-5353 x C4-2856 (92 F1 individuals)

Depth Distribution of 184 Offspring



v1 *Thinopyrum intermedium* chromosomes

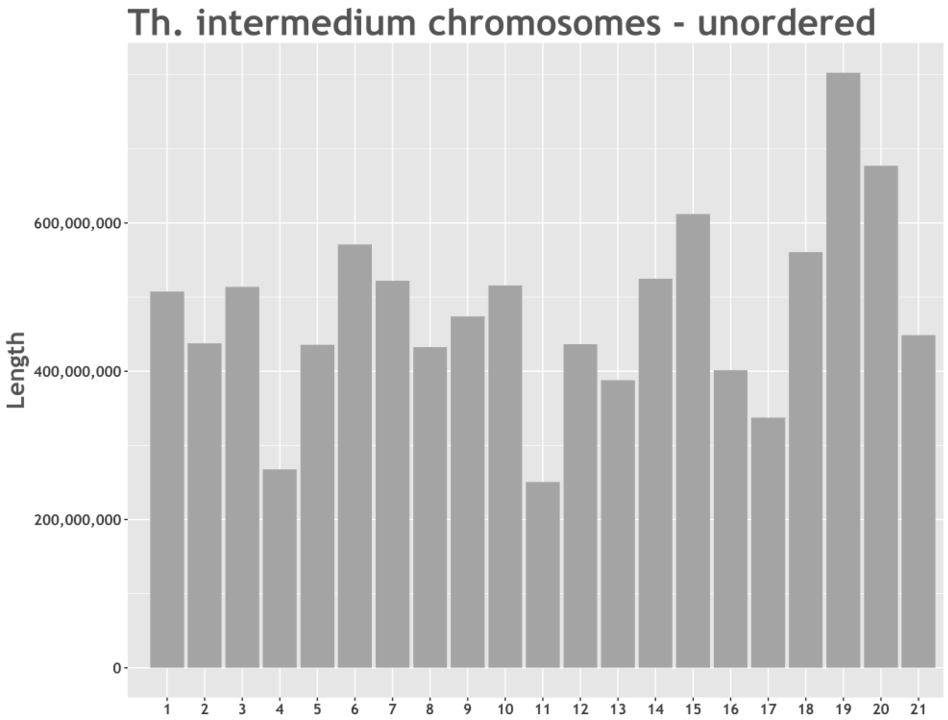


NRGene Assembly → Identify Misjoins → Integrated Assembly → Polished Assembly

10X Genomics 'Linked Reads'

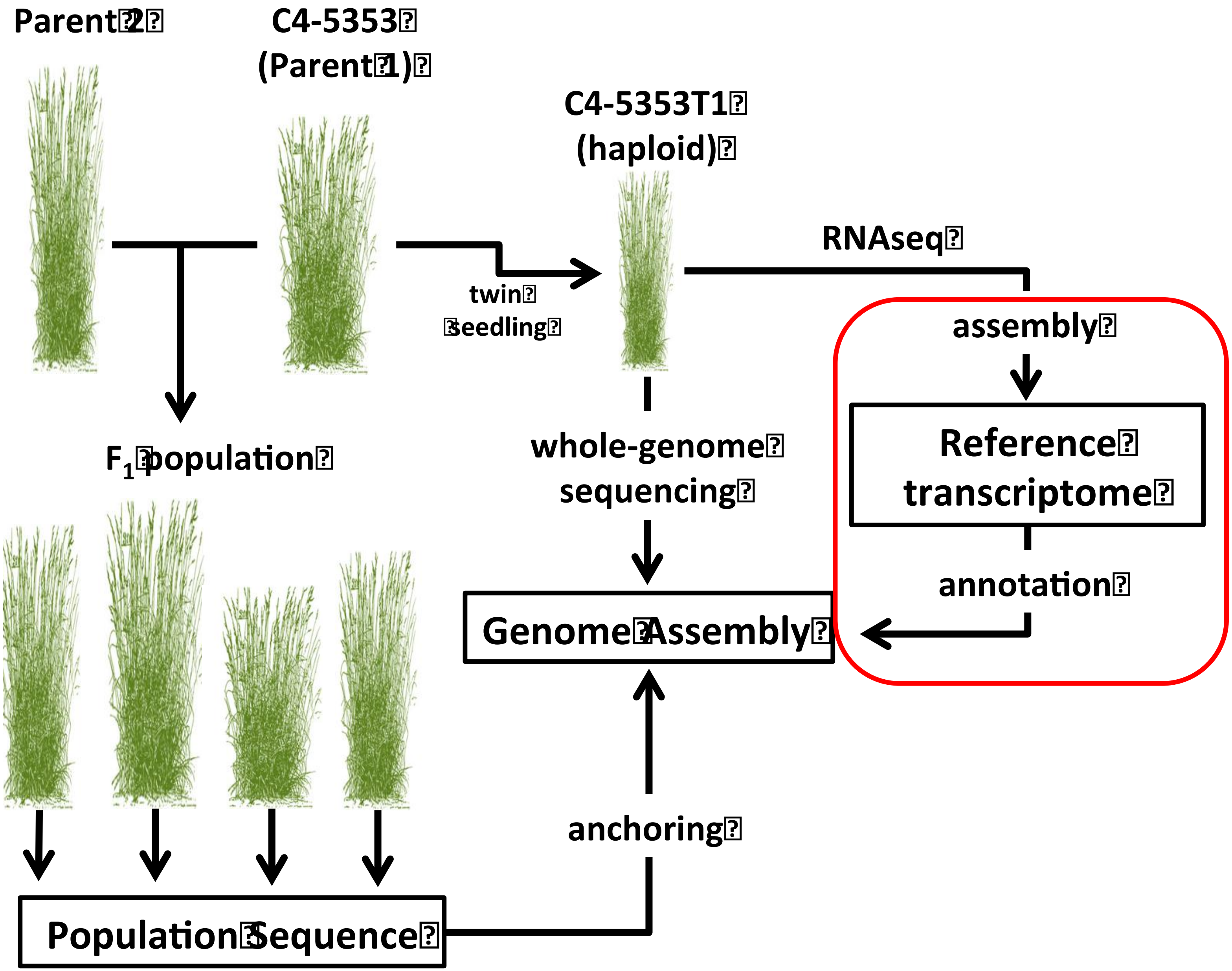
- POPSEQ map
- GBS map

Illumina Fragment (eliminate homozygous variants)



	Anchored Scaffolds
Number	**59,064**
Assembly Length	11,347,191,088
Gap Size	3.8%

+ 76,548 unanchored >1kb scaffolds (53 Mb)
+ 94,839 unanchored 'repeat' scaffolds (260 Mb)



Parent 2

C4-5353 (Parent 1)

C4-5353T1 (haploid)

twin seedling

RNAseq

assembly

Reference transcriptome

annotation

whole-genome sequencing

Genome Assembly

F_1 population

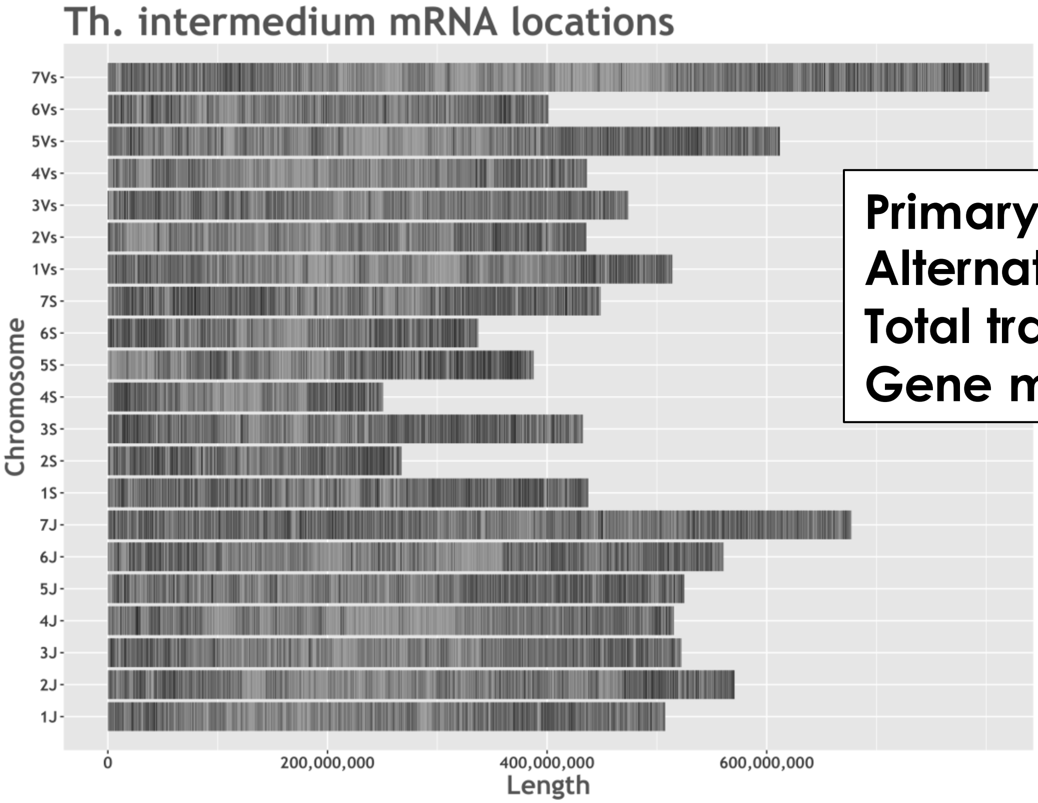
anchoring

Population Sequence

Reference Transcriptome and Preliminary Annotation (JGI v2.1)

16 tissue types from the C4-5353T1 haploid plant
HiSeq2500 2x150bp @ JGI - 2 Billion Illumina reads

Trinity assembly	**Contig**
Number	1,263,377
Total Assembly Length (bp)	782,859,424



Primary transcripts (loci) = 159,905
Alternative transcripts = 54,238
Total transcripts = 214,143
Gene models w/ RNAseq support = 122,028

Kansas State University

The genome of *Thinopyrum intermedium*

1.) Sequencing, assembly, anchoring/ordering of genome

2.) Parsing subgenomes & deciphering diploid progenitor history

Deciphering the diploid progenitors of *Th. intermedium*

- Many conflicting studies on evolution of IWG genome and progenitors
  - Most based on minimal markers / GISH studies
  - Best current thought: J J^b S

Thinopyrum elongatum (J)

Pseudoroegneria spp. (S)

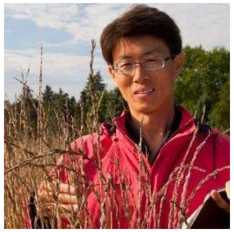
Thinopyrum bessarabicum (J^b)

<u>Other diploid species suggested as progenitors:</u>
Dasypyrum villosum (V)
Taeniatherum spp. (Ta)
Aegilops tauschii (D)

Deciphering the diploid progenitors of *Th. intermedium*

Genotype-By-Sequencing of 47 potential progenitor species*

Dataset courtesy of Xiaofei Zhang



↓

Extract 64-mers shared within each potential diploid progenitors (TASSEL & custom scripts)

↓

Align non-polymorphic, full length tags to IWG v1 chromosomes

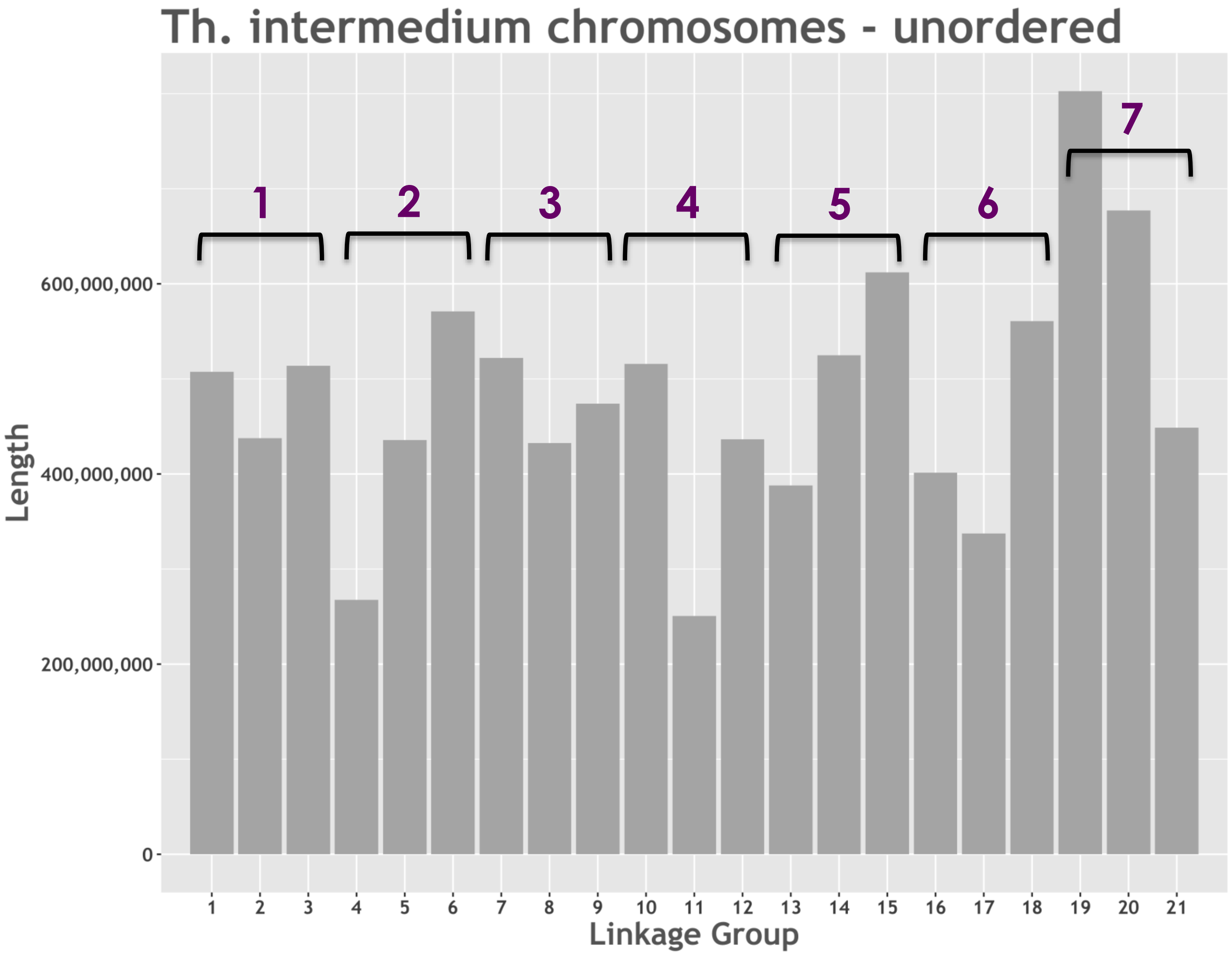
*<u>*Including…</u>*

P. gracillima
P. libanotica
P. spicata
P. strigosa
P. tauri
T. bessarabicum
T. curvifolium
T. elongatum
T. caepitosum
T. curvifolium
T. junceiforme

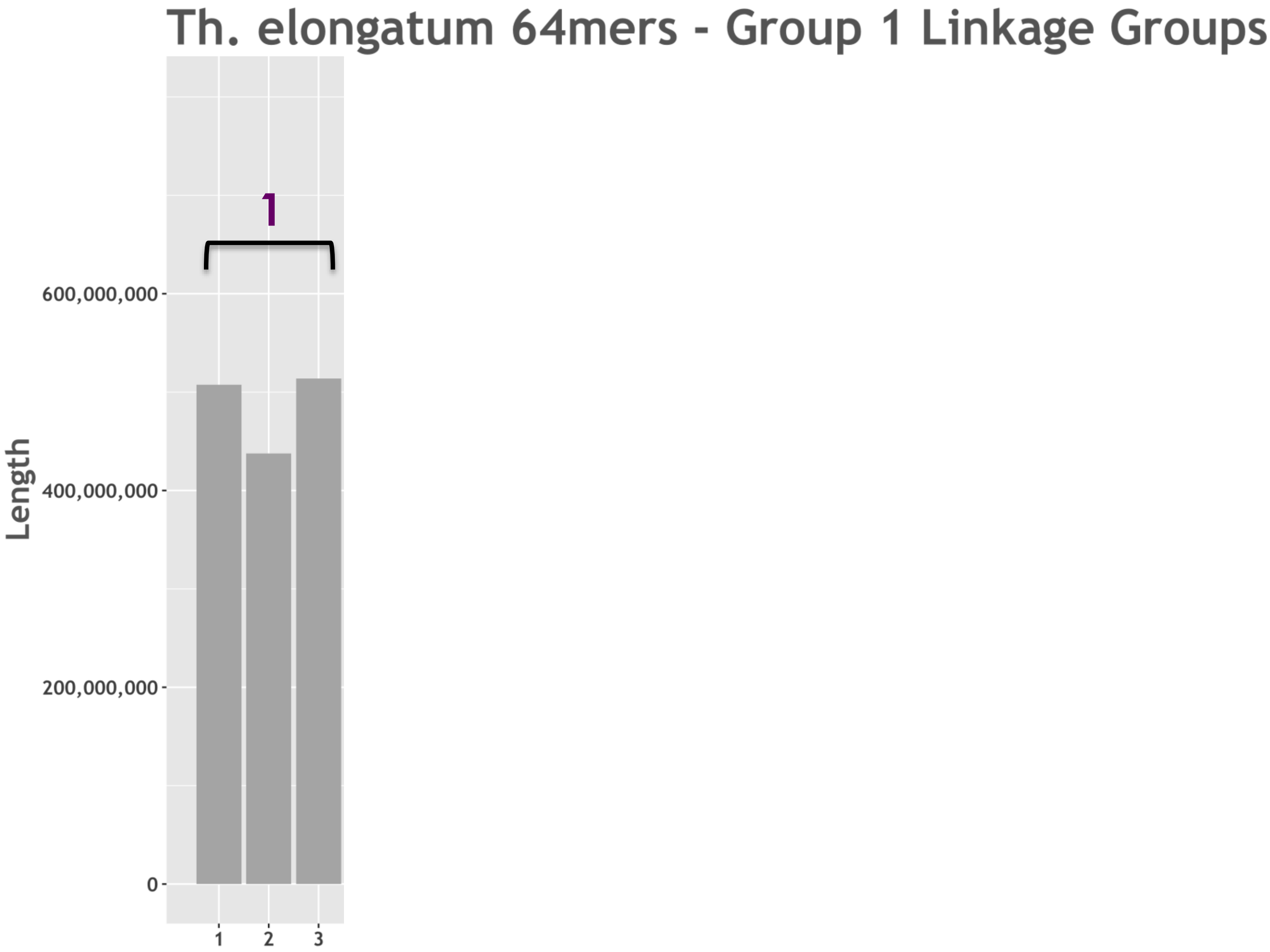
A. bicornis
A. tauschii
A. umbellulata
T. urartu
T. monococcum
Ag. cristatum
Ag. glaucum

KANSAS STATE
UNIVERSITY

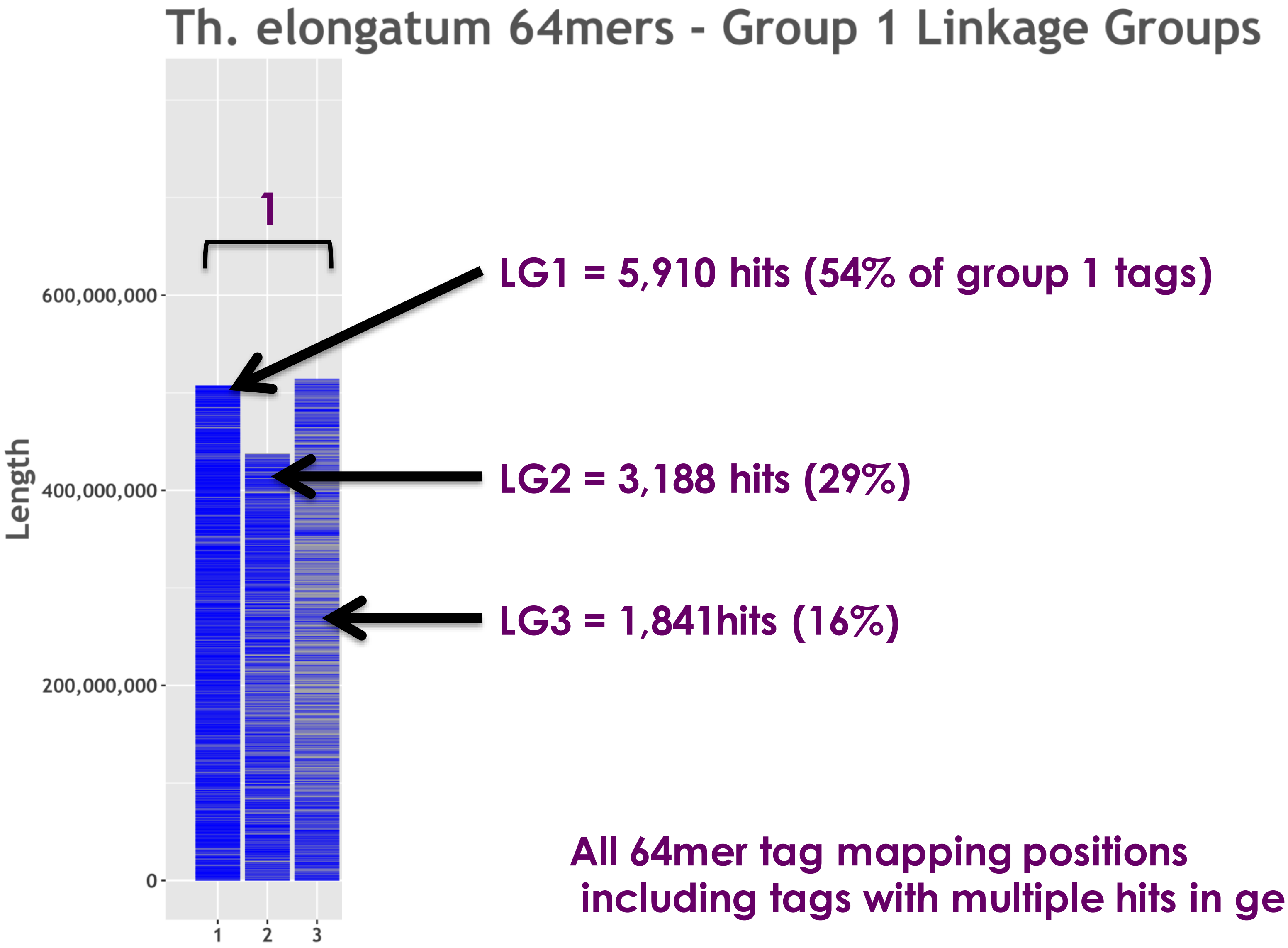
Parsing homoeologous groups into subgenomes



Parsing homoeologous groups into subgenomes



Parsing subgenomes of IWG



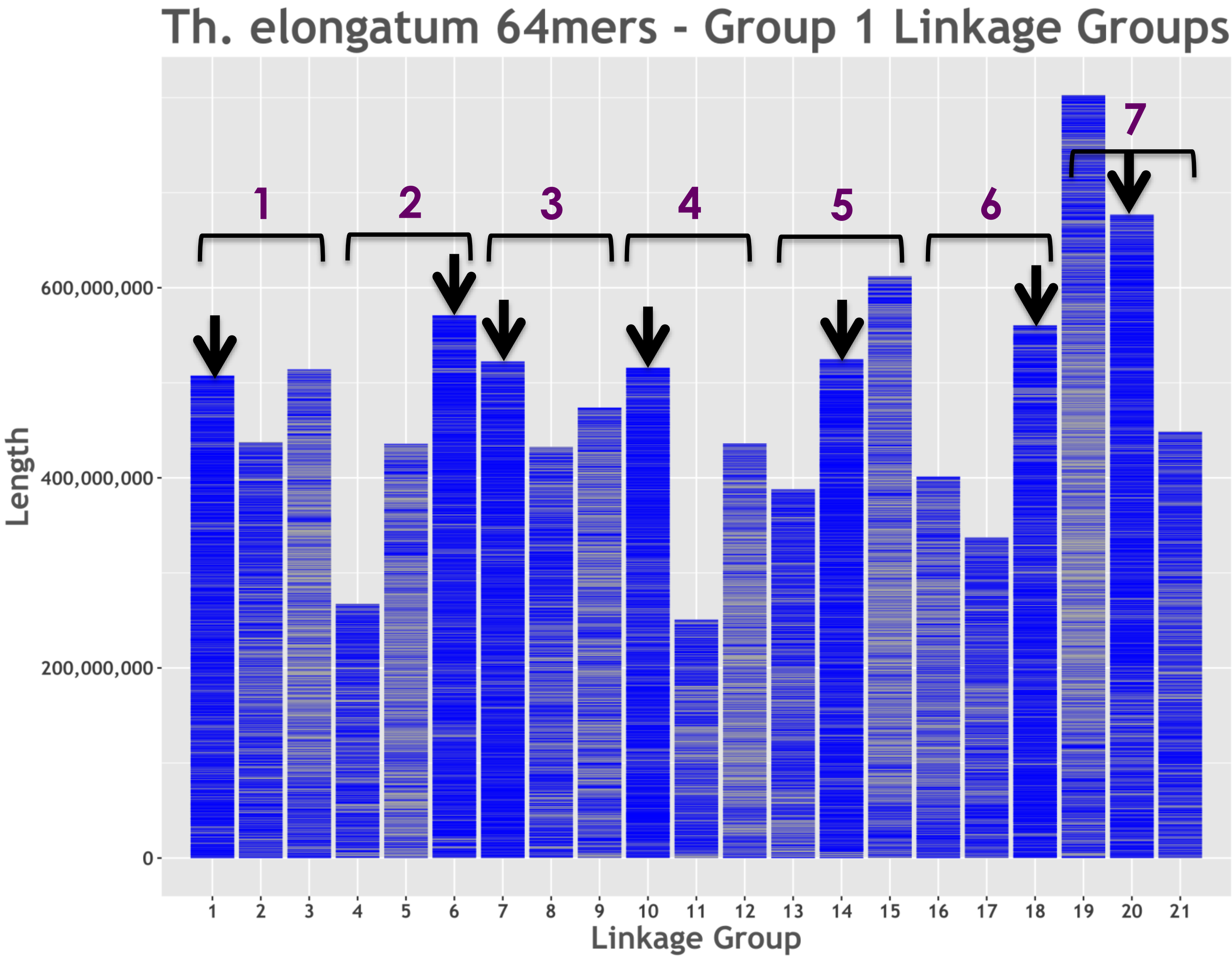
LG1 = 5,910 hits (54% of group 1 tags)

LG2 = 3,188 hits (29%)

LG3 = 1,841hits (16%)

All 64mer tag mapping positions including tags with multiple hits in genome

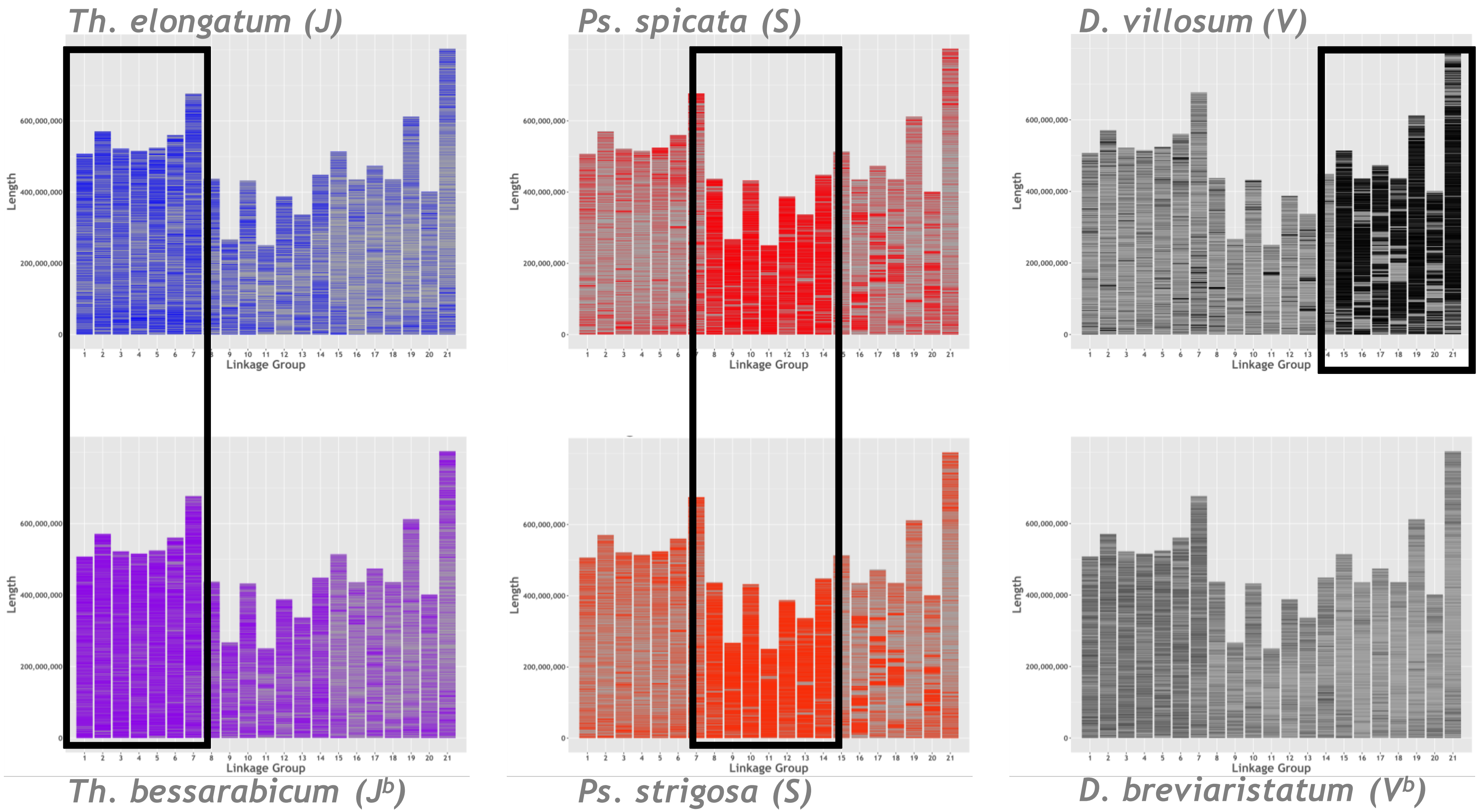
Parsing subgenomes of IWG



All 64mer tag mapping positions
including tags with multiple hits in genome

Kansas State University

Re-ordered linkage groups into subgenomes



Th. elongatum (J)

Ps. spicata (S)

D. villosum (V)

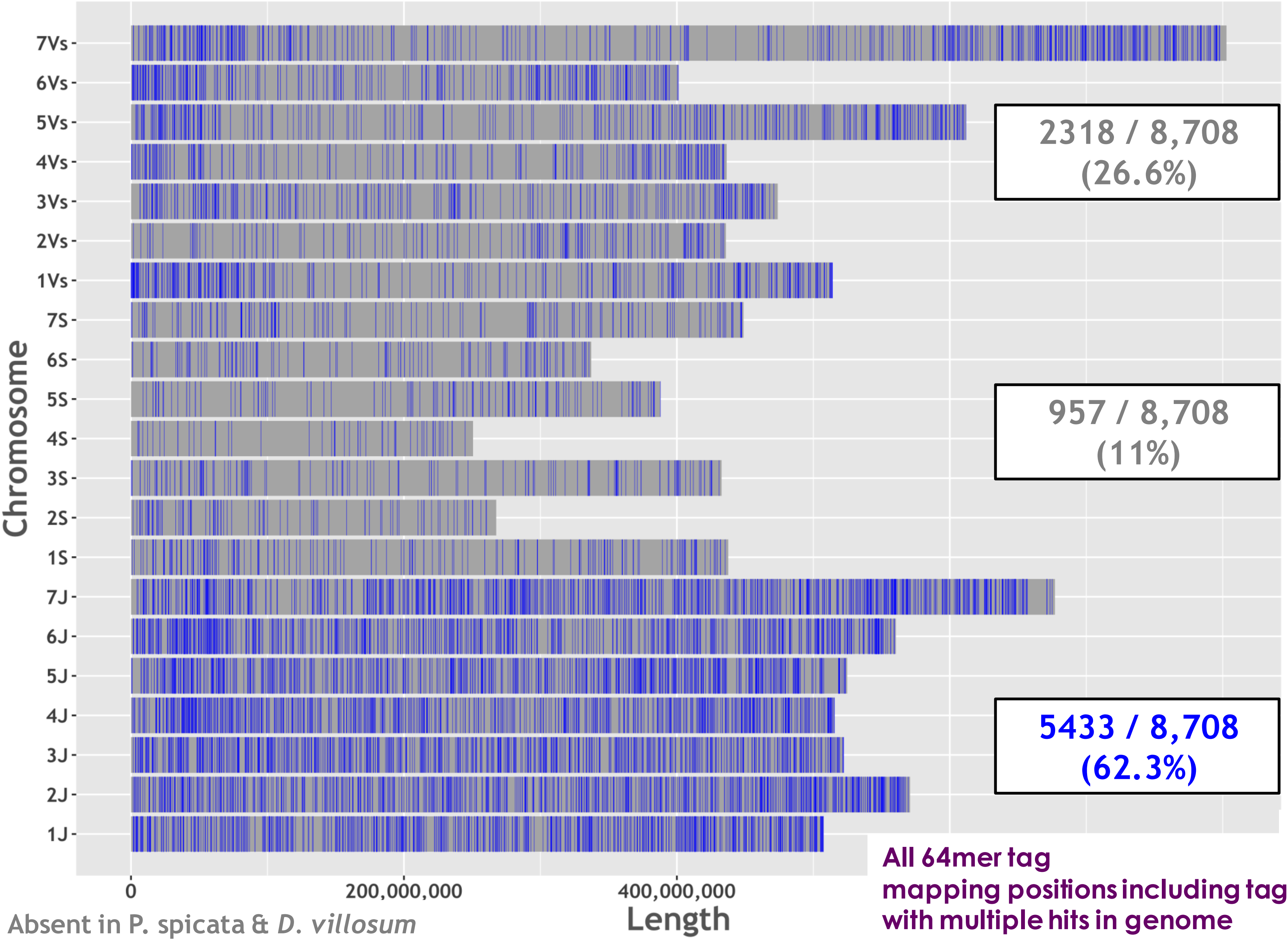
Th. bessarabicum (J^b)

Ps. strigosa (S)

D. breviaristatum (V^b)

All 64mer tag mapping positions including tags with multiple hits in genome

Th. elongatum - species unique 64mers *



2318 / 8,708
(26.6%)

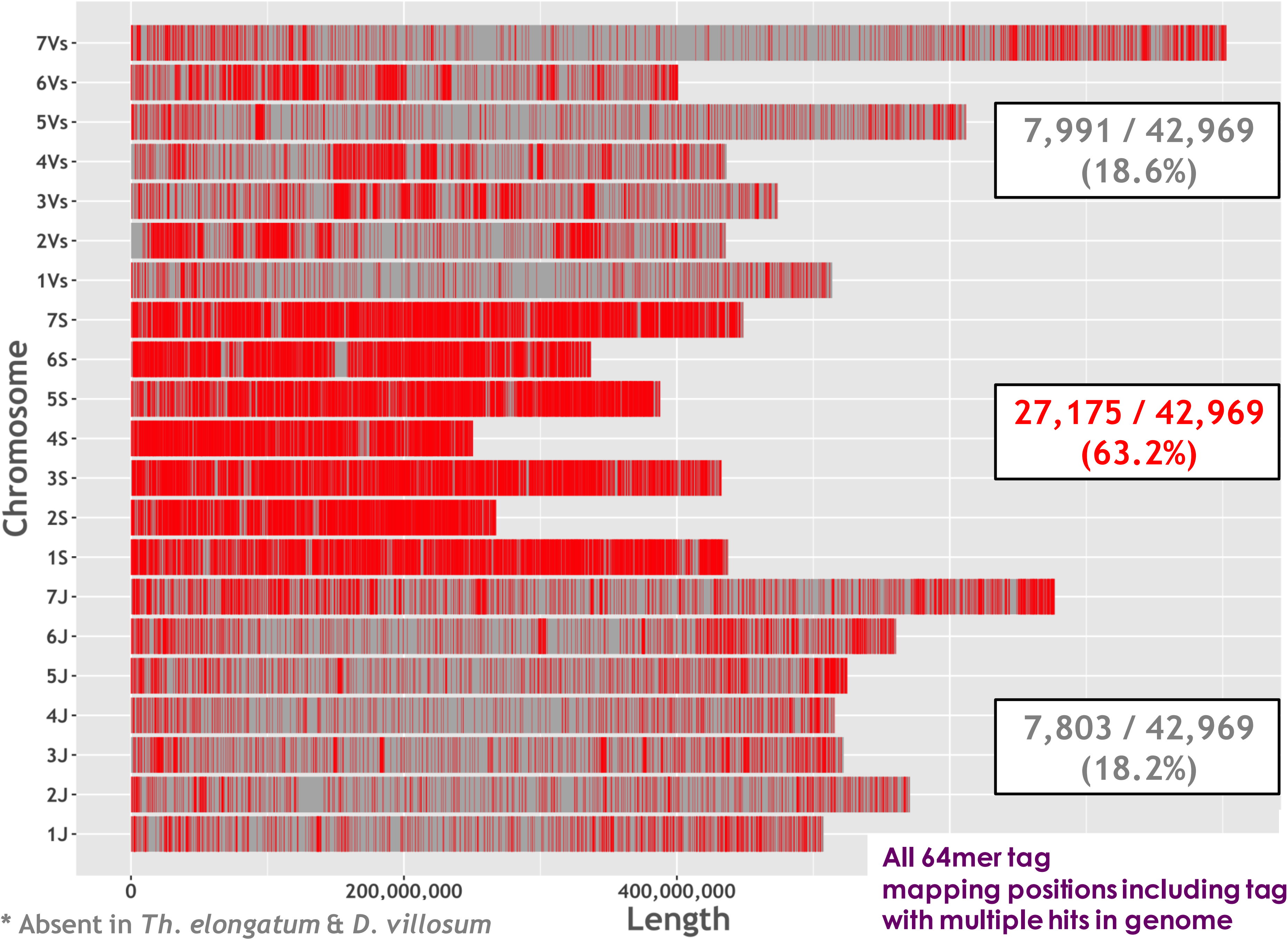
957 / 8,708
(11%)

5433 / 8,708
(62.3%)

All 64mer tag mapping positions including tags with multiple hits in genome

* Absent in P. spicata & *D. villosum*

Ps. spicata - species unique 64mers **



7,991 / 42,969
(18.6%)

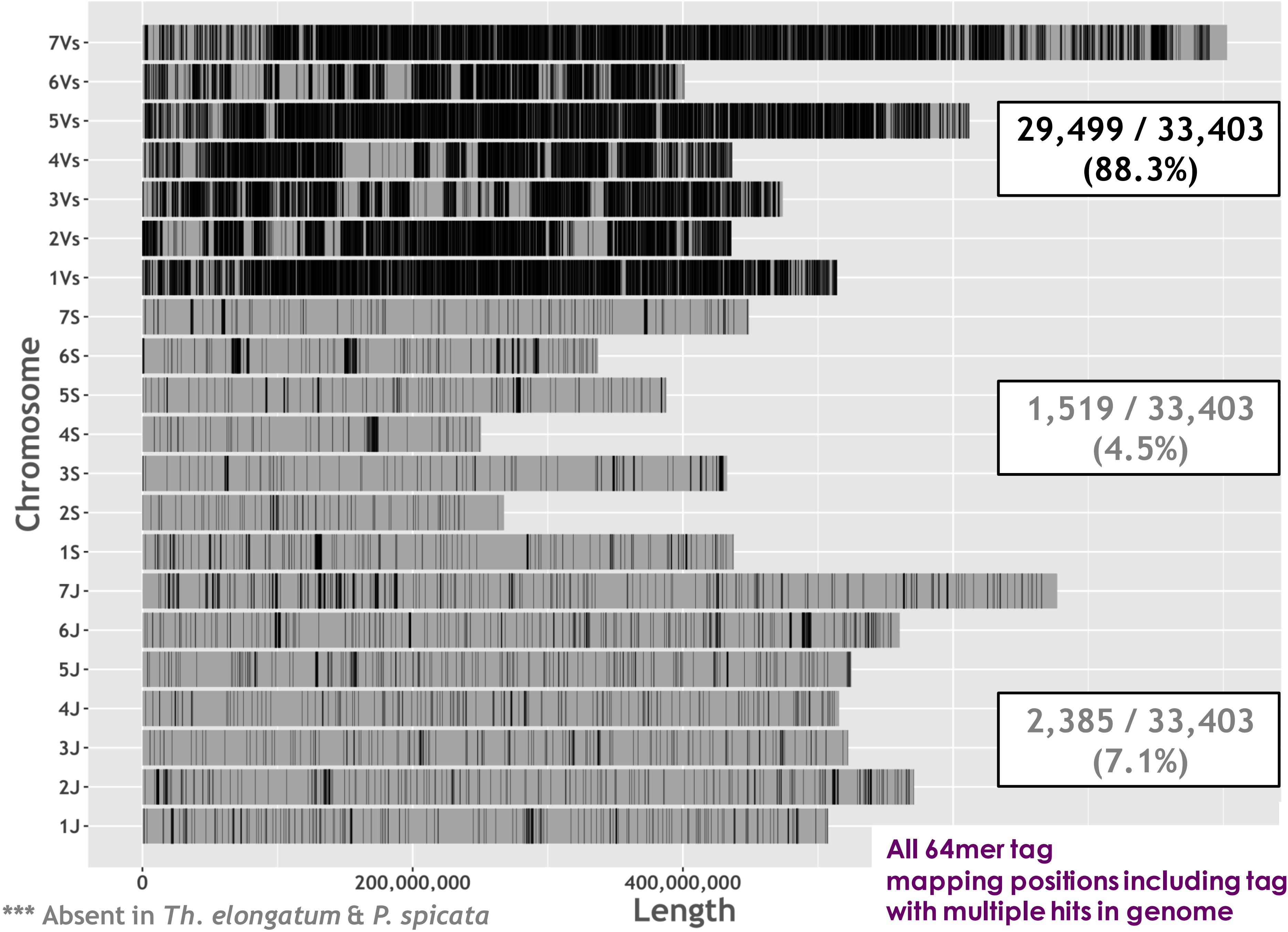
27,175 / 42,969
(63.2%)

7,803 / 42,969
(18.2%)

** Absent in *Th. elongatum* & *D. villosum*

All 64mer tag mapping positions including tags with multiple hits in genome

D. villosum - species unique 64mers ***



29,499 / 33,403
(88.3%)

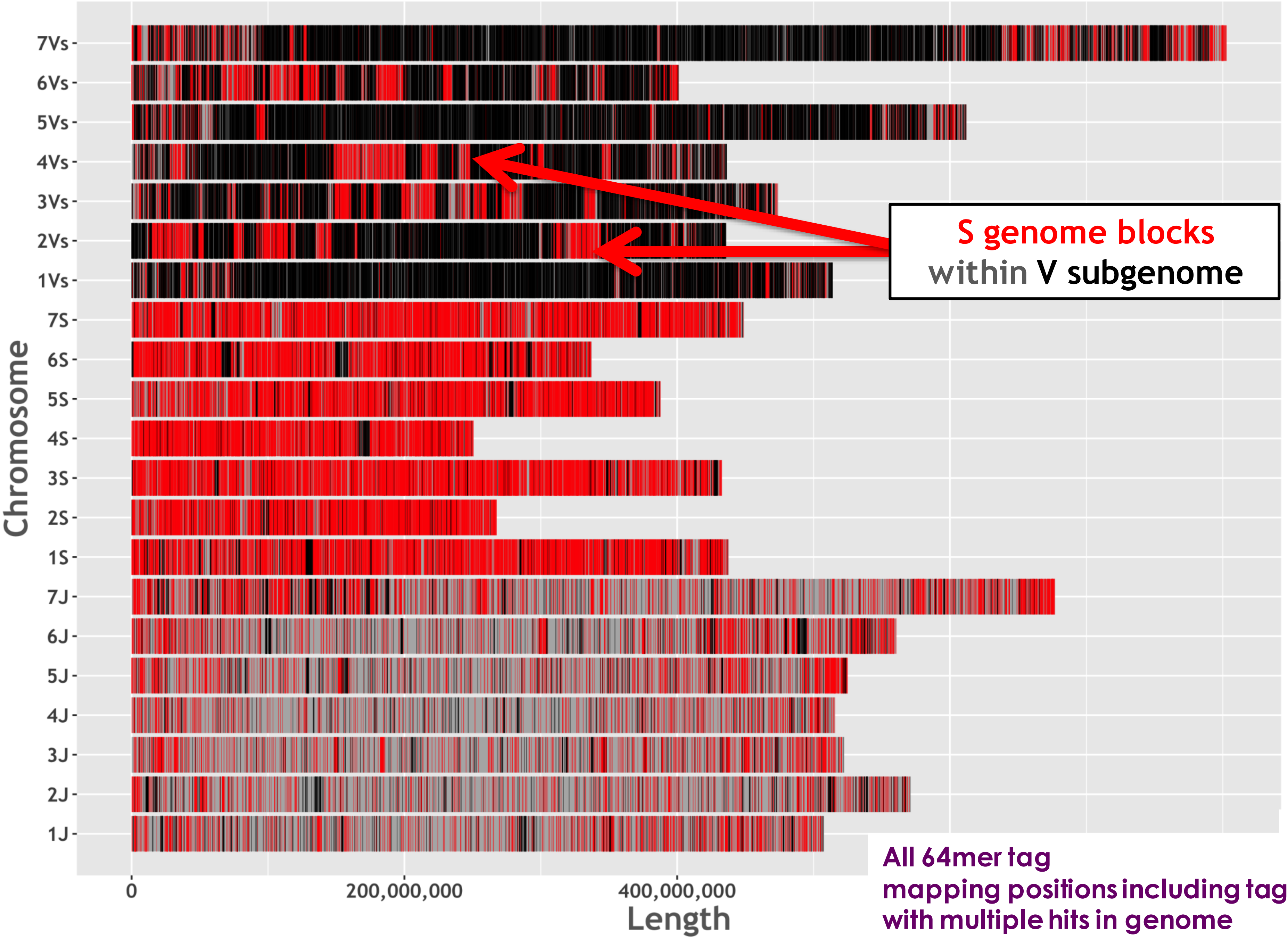
1,519 / 33,403
(4.5%)

2,385 / 33,403
(7.1%)

All 64mer tag mapping positions including tags with multiple hits in genome

*** Absent in *Th. elongatum* & *P. spicata*

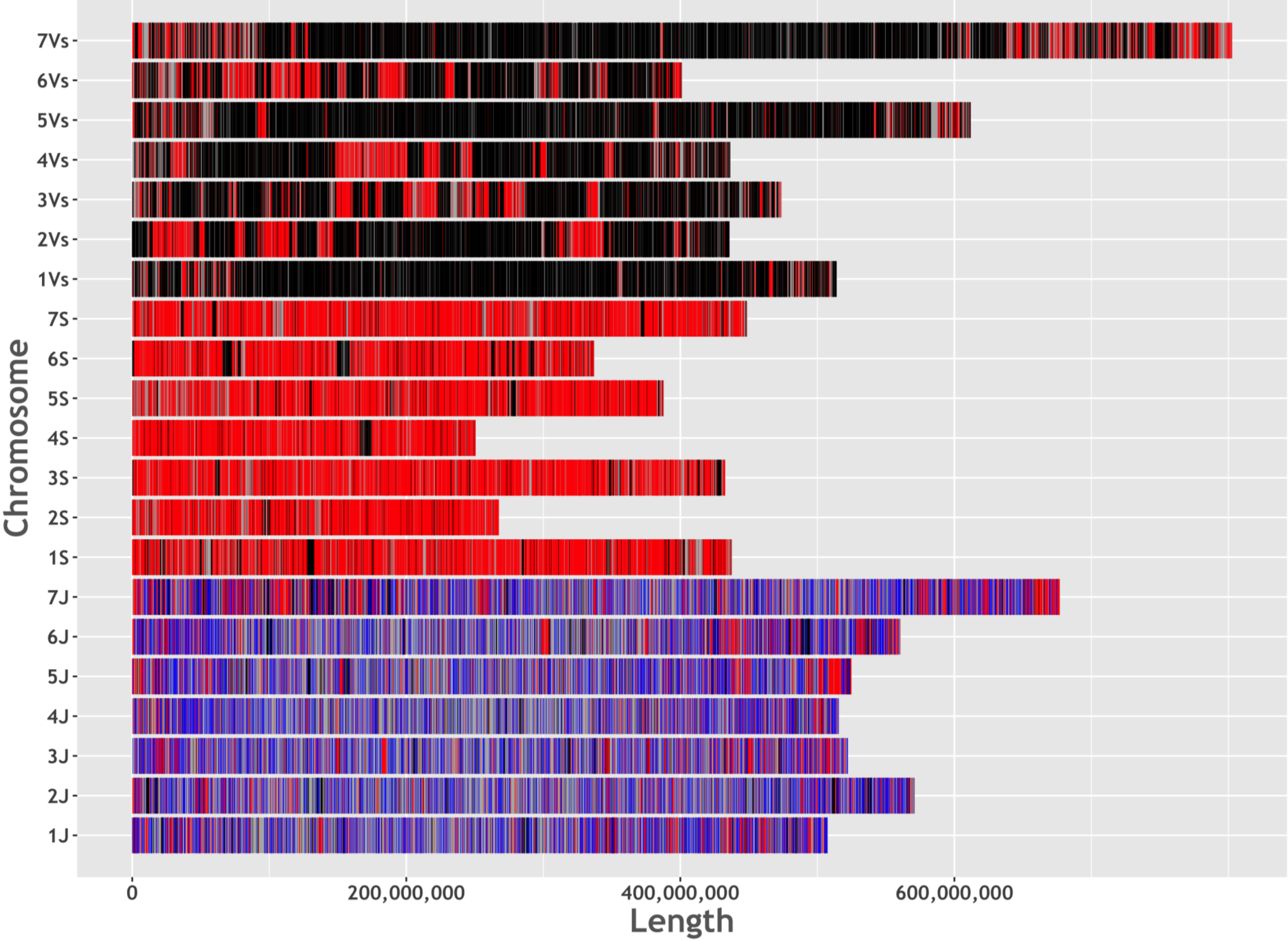
Ps. spicata + D. villosum species specific tags



S genome blocks within V subgenome

All 64mer tag mapping positions including tags with multiple hits in genome

The Thinopyrum intermedium genome: J S V^s



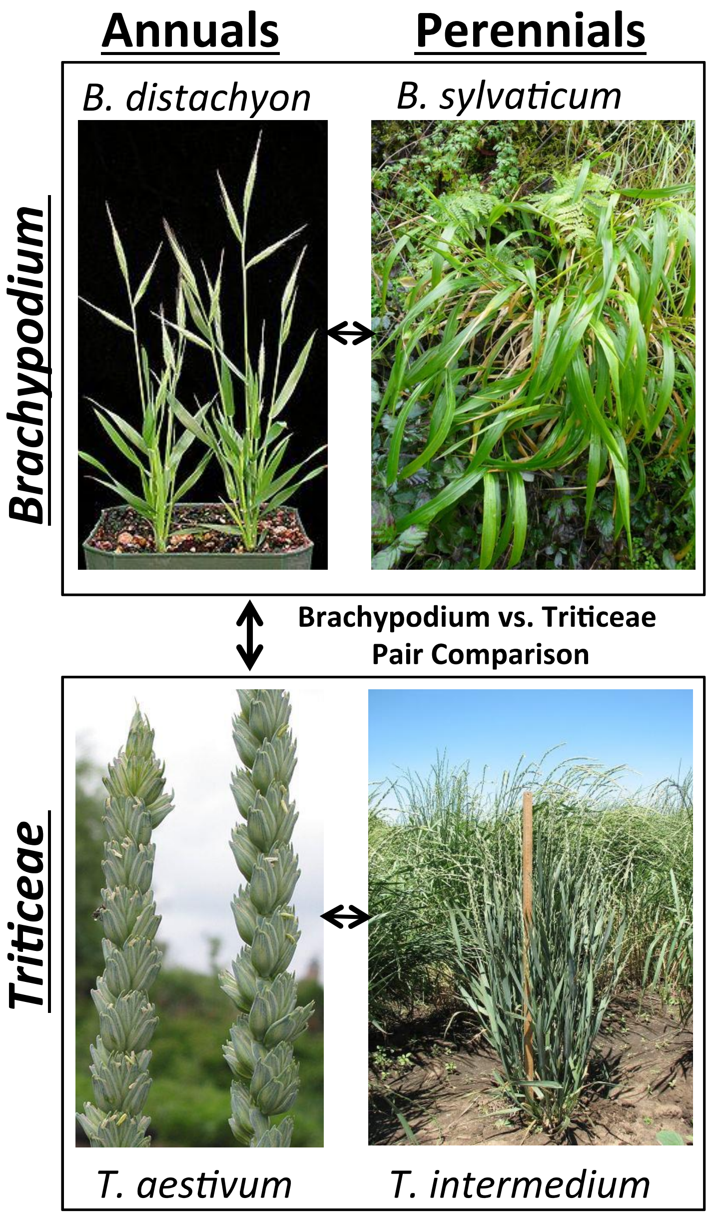
The genome of *Thinopyrum intermedium*

1.) Sequencing, assembly, anchoring/ordering of genome

2.) Parsing subgenomes & deciphering diploid progenitor history

3.) Biology of perenniality in the Triticeae

Examining mechanisms of perenniality across the Pooideae



Annuals **Perennials**

Brachypodium

B. distachyon ↔ *B. sylvaticum*

Brachypodium vs. Triticeae Pair Comparison

Triticeae

T. aestivum ↔ *T. intermedium*

Annual vs. Perennial Pair Comparison

RNAseq comparison of gene expression in perenniality relevant tissues, including:

- -roots
- -rhizomes
- -leaves and meristems pre/post vernalization
- -crown buds throughout two sexual cycles (of perennials)
- -developing seeds

Comparative genomics/Systems approach to identify novel genes and expression patterns in each perennial vs. annual pair

KANSAS STATE UNIVERSITY

Sequencing *Thinopyrum elongatum* (J/E subgenome progenitor)

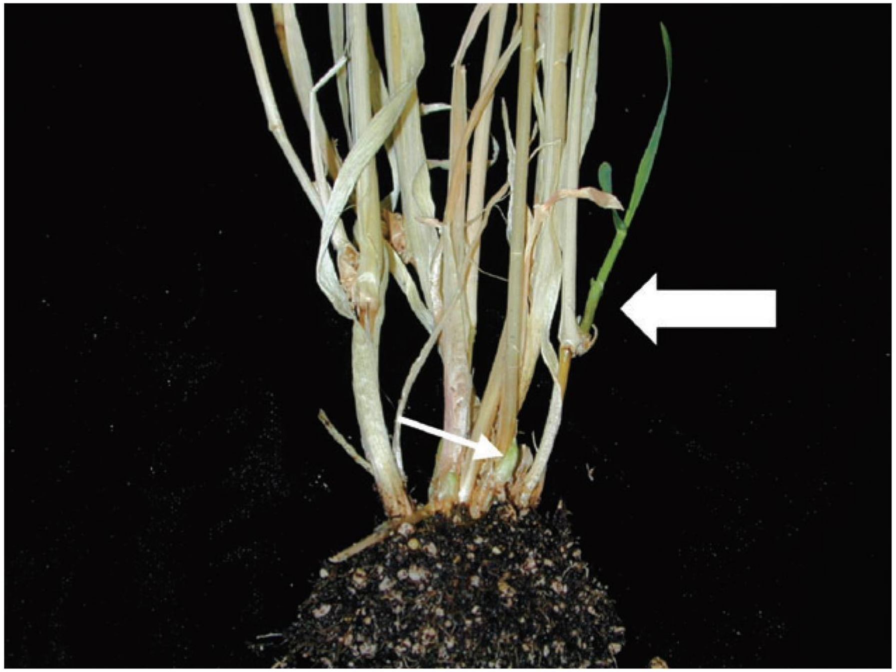
1C = 5.6 Gb
at least 5 generations of selfing completed

Why work on *Th. elongatum*?
1 of the 2 perennial progenitors of Kernza
Great system for functional genomics
Hybrids w/ annual wheat -> perennial wheat

wheat

wheat
+ all 7 chr from
Th. elongatum

wheat
+ chr 4E from
Th. elongatum



KANSAS STATE
UNIVERSITY

Sequencing *Thinopyrum elongatum* (J/E subgenome progenitor)

Genome Sequencing Approach for *Th. elongatum*

PacBio + 10X Genomics + HiC + GBS map


wheat

wheat
+ all 7 chr from
Th. elongatum

wheat
+ chr 4E from
Th. elongatum

Sequencing the 4E chromosome via flow sorting
(Jaroslav Dolezel – Czech Republic)

Genetic systems for unraveling the perennial regrowth trait

-Radiation Hybrids to knock out perenniality
(Shuwen Wang – TLI)

-Recombination System to add perenniality
(Dal-Hoe Koo & Bernd Friebe – Kansas State)

KANSAS STATE
UNIVERSITY

Summary

1.) Refining & improving the IWG genome

- NRgene assembly
    - scaffolded with 10X Linked-Reads
    - anchored with GBS consensus map + PopSeq markers
    - annotation in progress
- version 2
    - incorporate additional 10X reads and HiC scaffolding

2.) Parsing subgenomes & deciphering diploid progenitors

- Genotype-by-Sequencing enabled k-mer approach in candidate diploid species

- Transcriptome sequencing of diploids
    - on-going genome-independent reconstruction using transcriptome assemblies

3.) Biology of perenniality

- Cross-tribe RNAseq
- Genomic characterization of addition lines

KANSAS STATE UNIVERSITY

Acknowledgements

Kansas State University
Jesse Poland
Traci Kantarski
Jared Crain
Mark Lucas

The Land Institute
Lee DeHaan
Shuwen Wang
Kathryn Turner

HudsonAlpha
Jeremy Schmutz
Jane Grimwood
Chris Plott
Jerry Jenkins

USDA – Logan, Utah
Steve Larson

University of Minnesota
Jim Anderson
Xiaofei Zhang
Kayla Altendorf
Don Wyse
David Marks

Funding Sources

Department of Energy –
Joint Genomes Institute Community Science Program

The Land Institute
Perennial Agriculture Project &
Malone Foundation

Minnesota Dept. of Agriculture

UMN Forever Green Initiative

USDA Postdoctoral Fellowship

Kevin Dorn @Dornomics www.kdorn.com