



(lack of) Diversity in Wheat Promoters

presented [via Zoom] @ IWGSC Main Workshop, PAGXXIX San Diego CA 🥺, Saturday January 8th 2022

Michael Hammond-Kosack, Department of Biointeractions and Crop Protection, Rothamsted Research, UK

experimental design

- (i) **459 wheat gene IDs (1273 homoeologues) for 10 trait categories**
- (ii) **96 cultivars** chosen: 84 hexaploid (15 landraces), 2 tetraploid & 10 diploid
- (iii) target sequence = **1700bp of promoter and 5'UTR sequences**, all extracted from **IWGSC_CS_refseq1.0** in Spring 2017 (**prepublication access**)
- (iv) decision to use **Daicel Arbor Biosciences** (formerly MYcroarray) as sub-contractor, because of **proprietary RNA MYbaits technology**
- (iv) **high-stringency workflow for MYbaits design** (Jake Enk) to make individual capture of homoeologue promoters more likely
- (v) target sequence capture and amplification by **myReads team** (Arbor Bio) and PE100 sequencing



Trait	Category	Unique genes	Homoeologues
T1	Yield Resilience	28	82
T2	Grain Composition	59	154
T3	Grain Development	19	52
T4	Biotic Stress (fungi & insects)	59	164
T5	Abiotic Stress (drought, temperature)	30	81
T6	Nutrient Use Efficiency	67	199
T7	Canopy Development/Plant Architecture	56	161
T8	Flower Biology	23	66
T9	Root Architecture	72	200
T10	Recombination	46	114
Total		459	1273

Complete workflow:



Summary of Data obtained

total sequences, SNPs & InDels:

cultivar	ploidy	Sequence* (Mbp)	Total polymorphisms		Homozygous polymorphisms	
			SNPs	InDels	Hom	Frequency
CS	hexaploid, reference	3.15	6,022	555	1,894	0.6
Abbot	hexaploid	2.93	9,896	933	5,351	1.8
Bobwhite	hexaploid	2.34	6,917	651	3,941	1.7
Cadenza	hexaploid	3.05	10,107	949	5,558	1.8
Einstein	hexaploid	3.10	10,724	981	5,782	1.9
Hobbit	hexaploid	5.34	23,466	1,744	10,845	2.0
Kronos	tetraploid, durum wheat	2.58	13,783	1,262	8,804	3.4
M031	diploid, T.monococcum	8.63	216,835	13,200	127,922	14.8
M037	diploid, T.monococcum	7.58	191,506	11,687	111,202	14.7
M049	diploid, T.monococcum	9.03	224,696	13,456	129,843	14.4
M308	diploid, T.monococcum	9.81	248,884	15,116	146,610	14.9
Paragon	hexaploid	3.68	12,288	1,104	6,428	1.7
Riband	hexaploid	3.77	13,395	1,149	6,980	1.8
Soisson	hexaploid	3.15	9,029	821	5,295	1.7
W115	hexaploid, Watkins landrace	3.25	10,502	952	5,887	1.8
W141	hexaploid, Watkins landrace	4.83	19,466	1,529	9,593	2.0
W160	hexaploid, Watkins landrace	2.84	8,883	864	5,205	1.8
W199	hexaploid, Watkins landrace	2.87	6,540	689	3,867	1.3
W203	hexaploid, Watkins landrace	2.65	7,394	733	4,688	1.8

average length of target sequences captured:

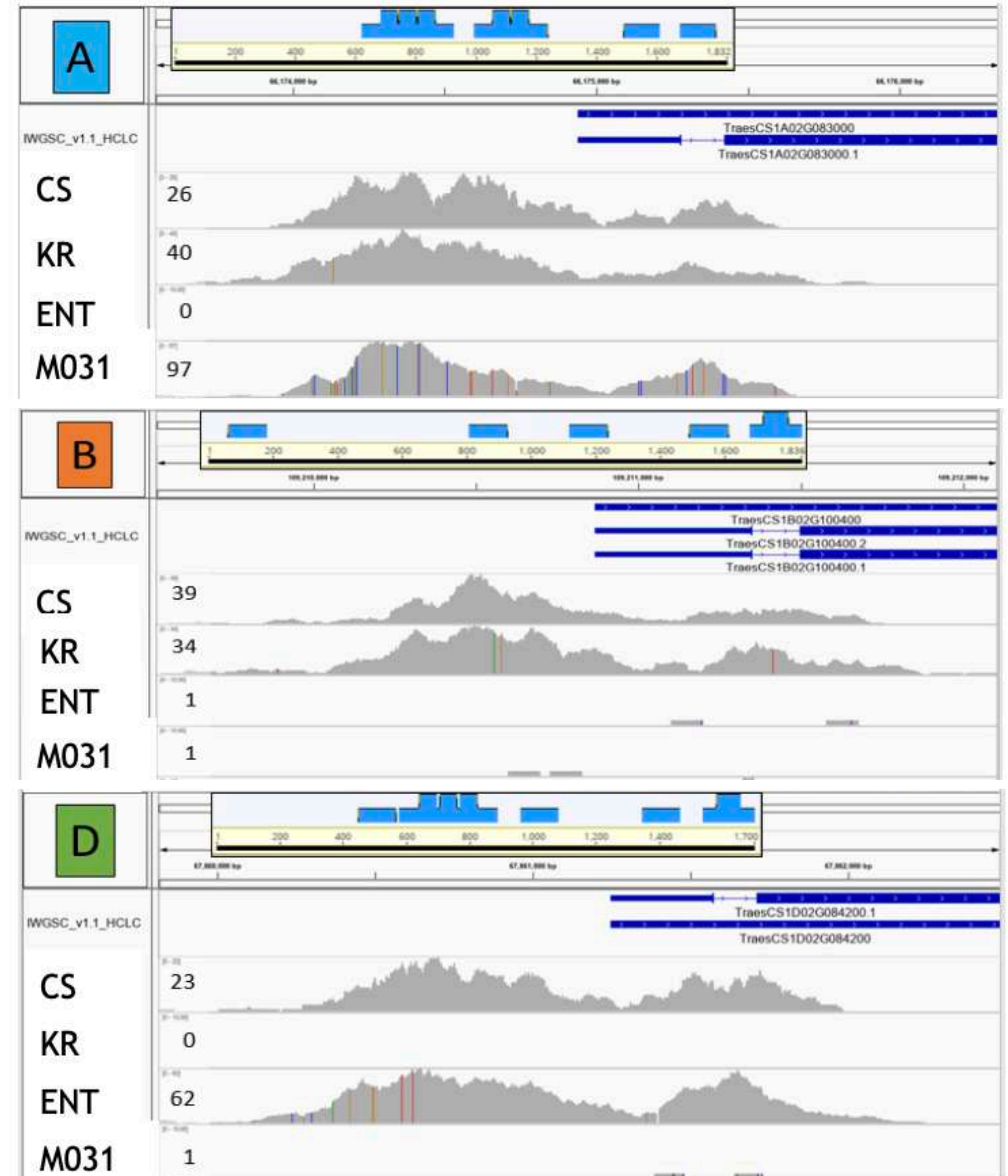
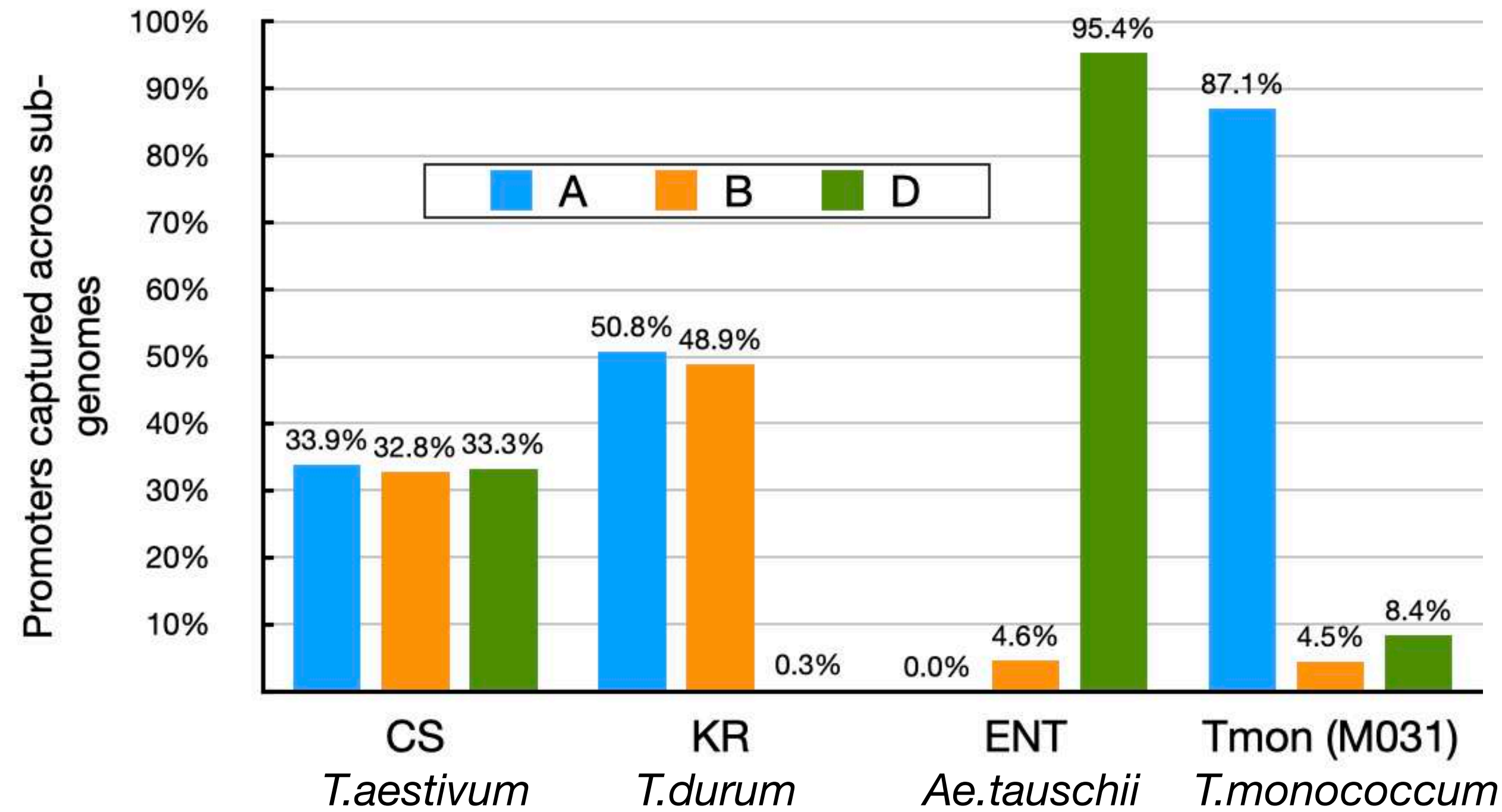
a	Promoter	5'UTR	Target sequence (promoter+5'UTR)	Exons/Introns	Total sequence
Average Length (bp) (n=908)	1416bp	235bp	1650bp	342bp	1993bp
± Stdev (bp)	575	327	536	496	568
± SEM (bp)	19	11	18	16	19

average sequencing depths by ploidy:

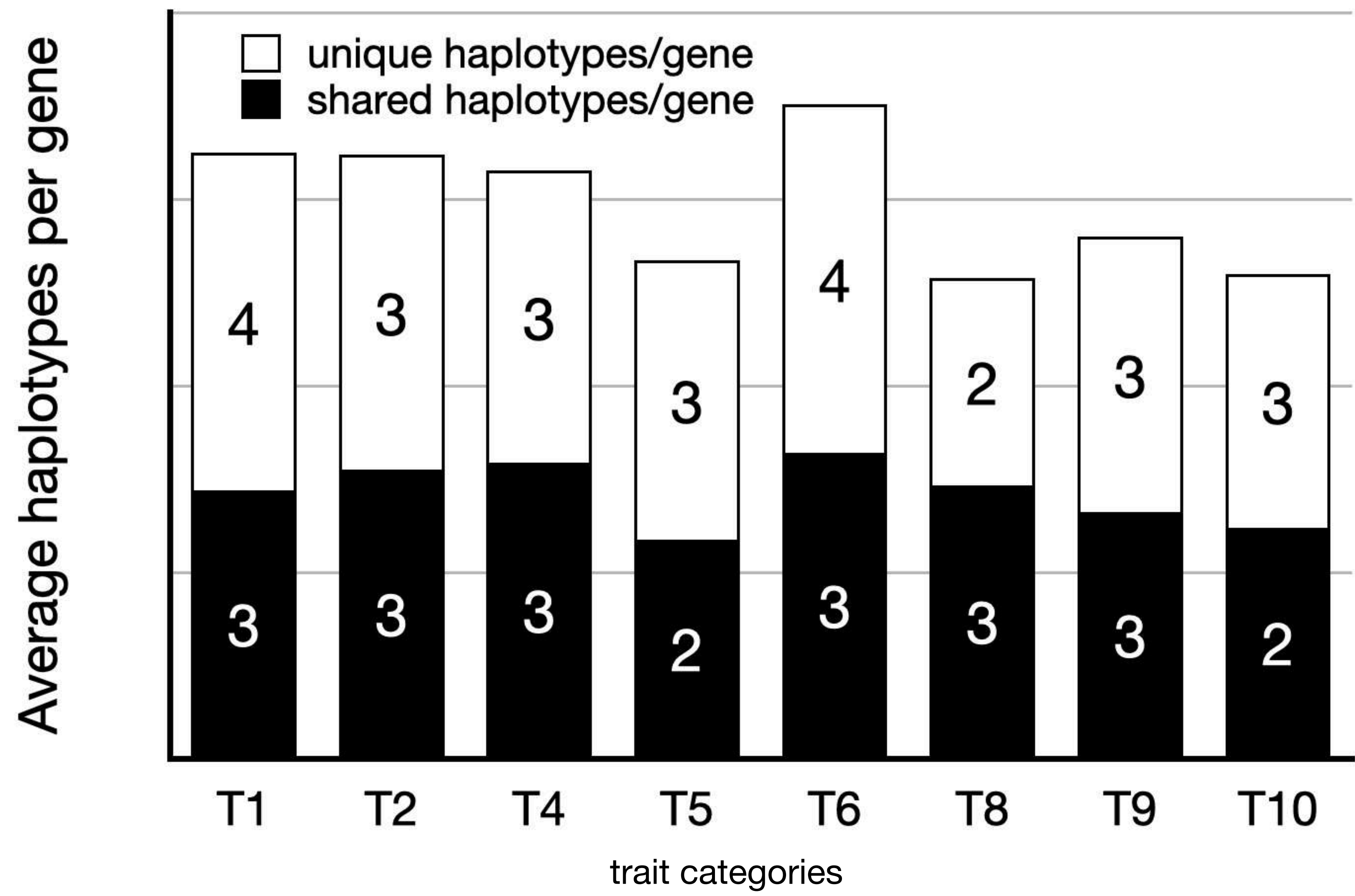
b	hexaploid	tetraploid		diploid		
cultivar	CS	KR	APG	M031	ASP	ENT
n	908	585	386	311	267	306
Average of maximum depth [ratio vs CS]	50	60	65	180	119	130
	[1]	[1.2]	[1.3]	[3.6]	[2.4]	[2.6]
± Stdev	20	25	37	109	88	53
± SEM	0.65	1.03	1.88	6.17	5.39	3.03

Homoeologue-specific Promoter Capture

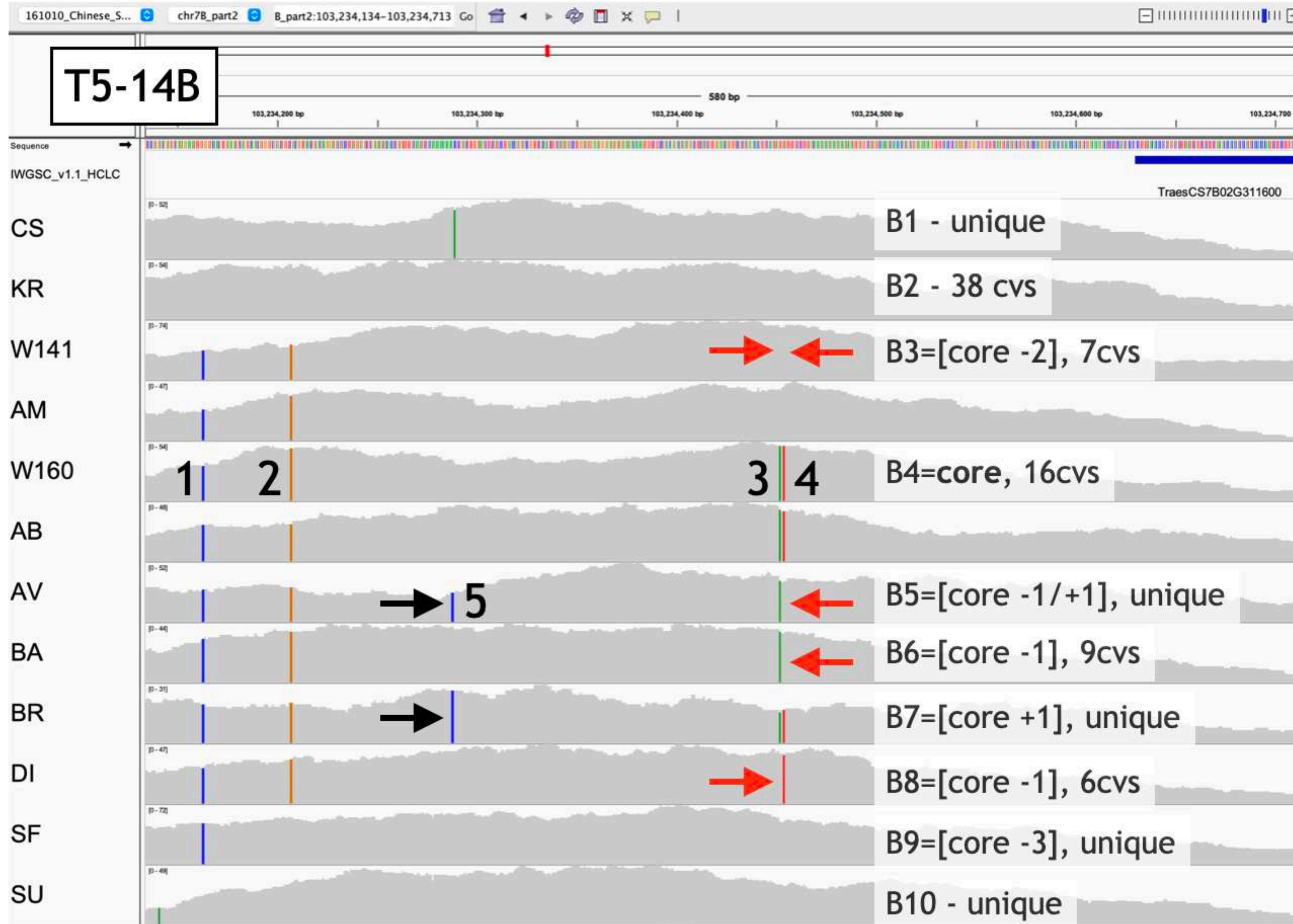
accession:	CS (AABBDD)			KR (AABB)			ENT (DD)			Tmon (AmAm)		
sub-genome:	A	B	D	A	B	D	A	B	D	A	B	D
expected capture:	✓	✓	✓	✓	✓	X	X	X	✓	✓	X	X
ideal capture (%):	33.3	33.3	33.3	50	50	0	0	0	100	100	0	0



hexaploid wheats exhibit few promoter haplotypes across different categories



unique and shared haplotypes are often related



TaOPR2 (Chr7B) -
Jasmonic Acid biosynthesis

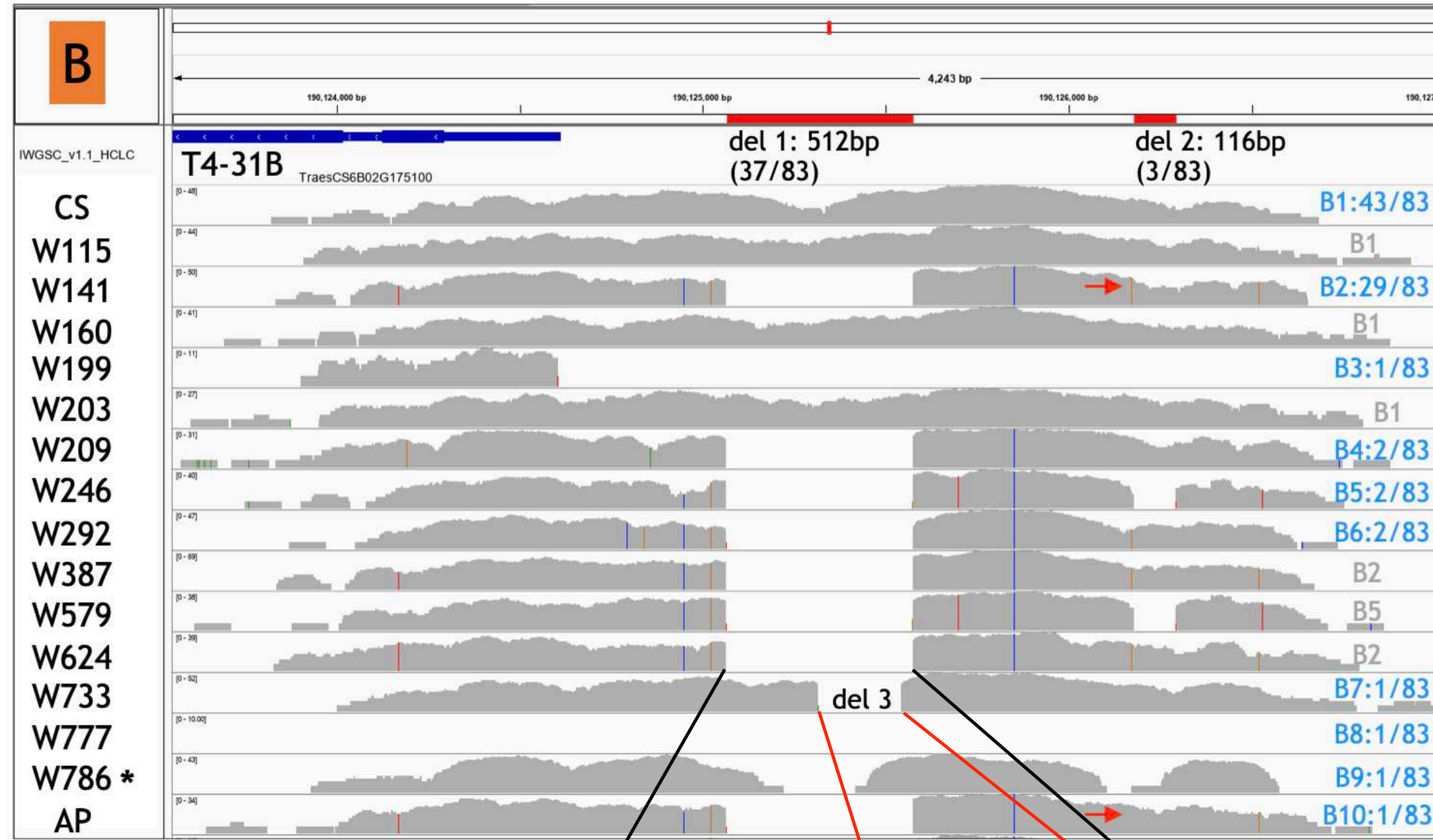
5 unique haplotypes:
B1, B5, B7, B9, B10

5 shared haplotypes:
B2, B3, B4, B6, B8

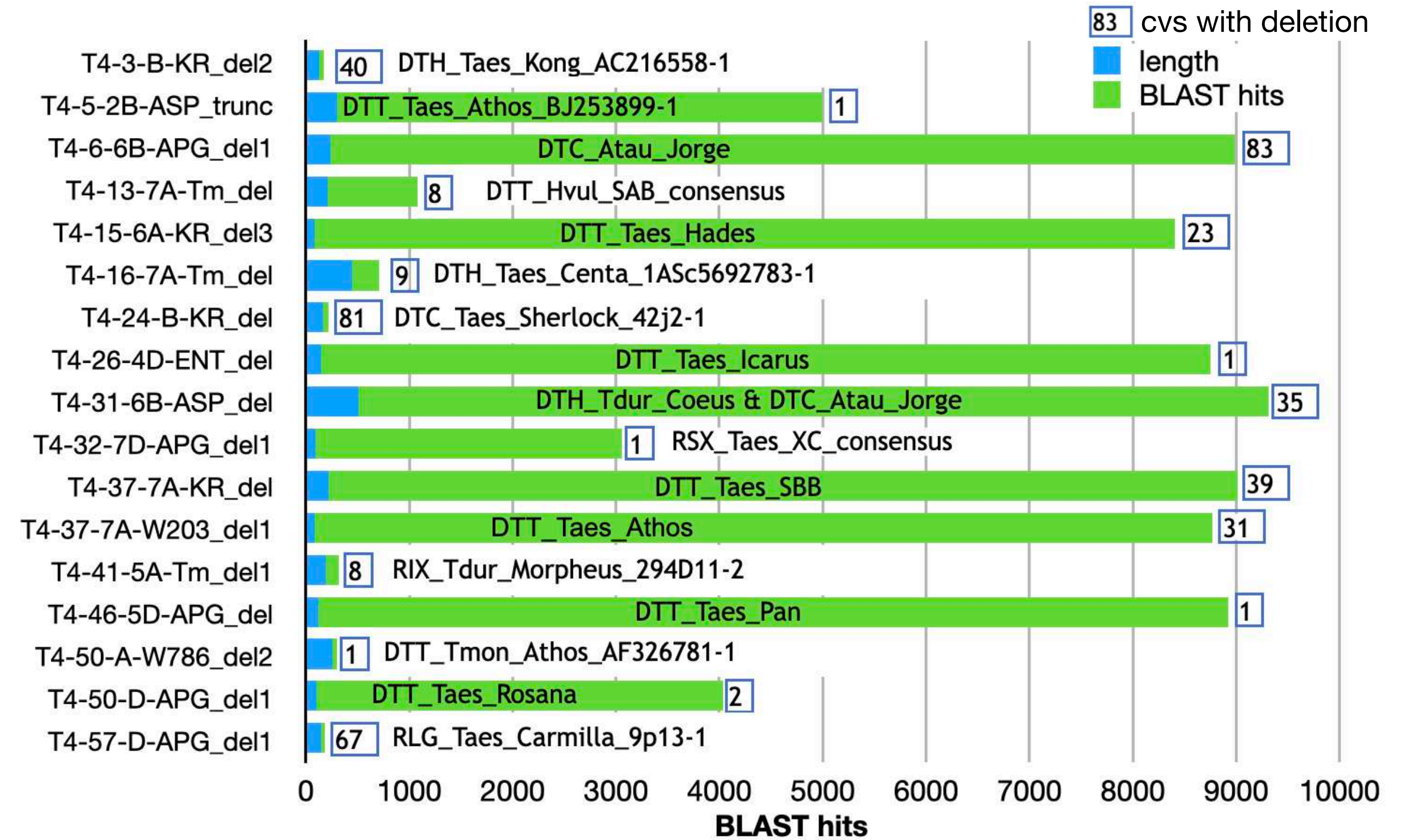
8 of these related with
a core of common SNPs

large deletions map to Transposable Elements

Transcription Factor **TaWRKY B**-homoeologue promoter contains a chimera of 2 transposable elements



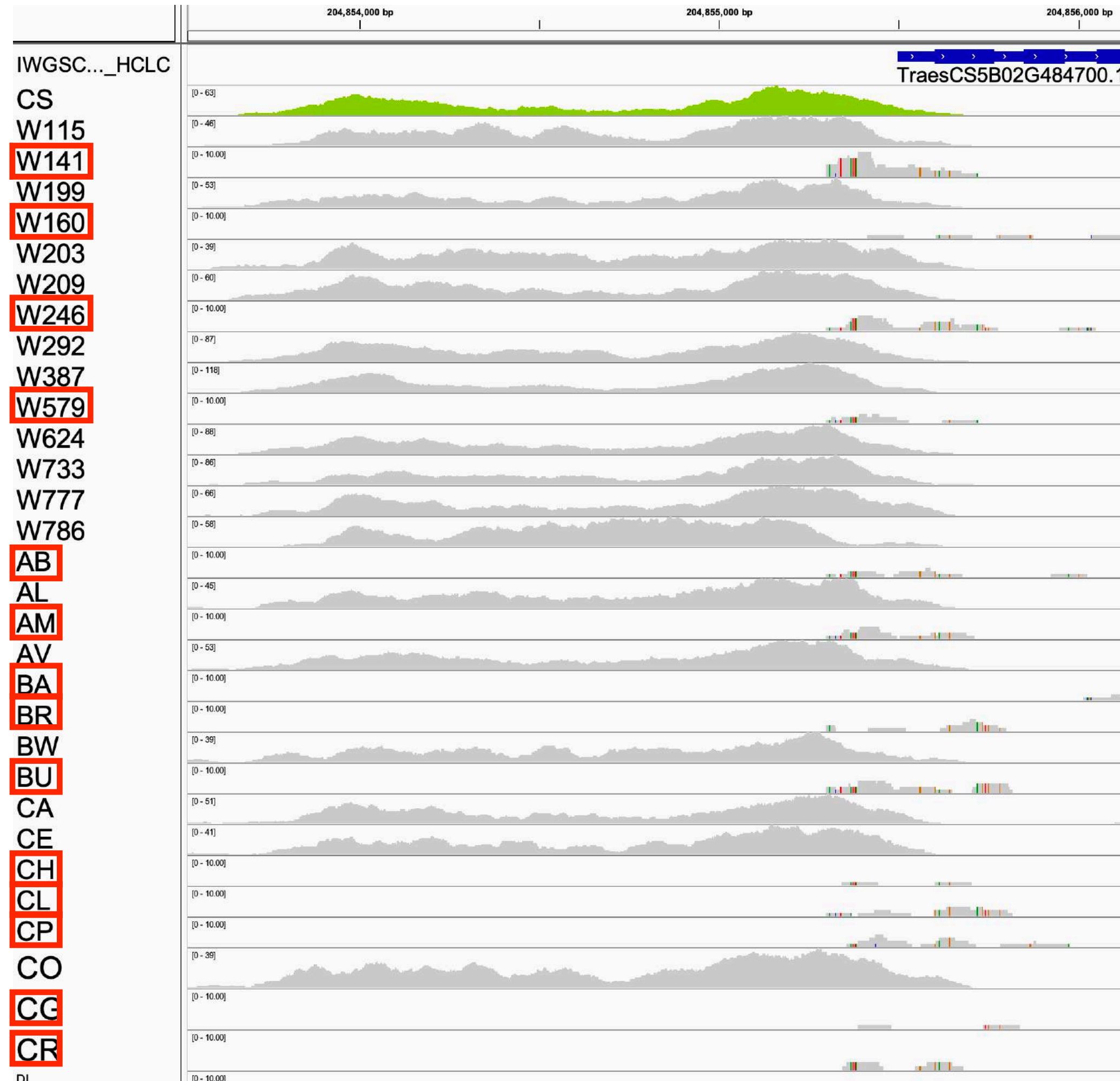
Biotic Stress large promoter deletions



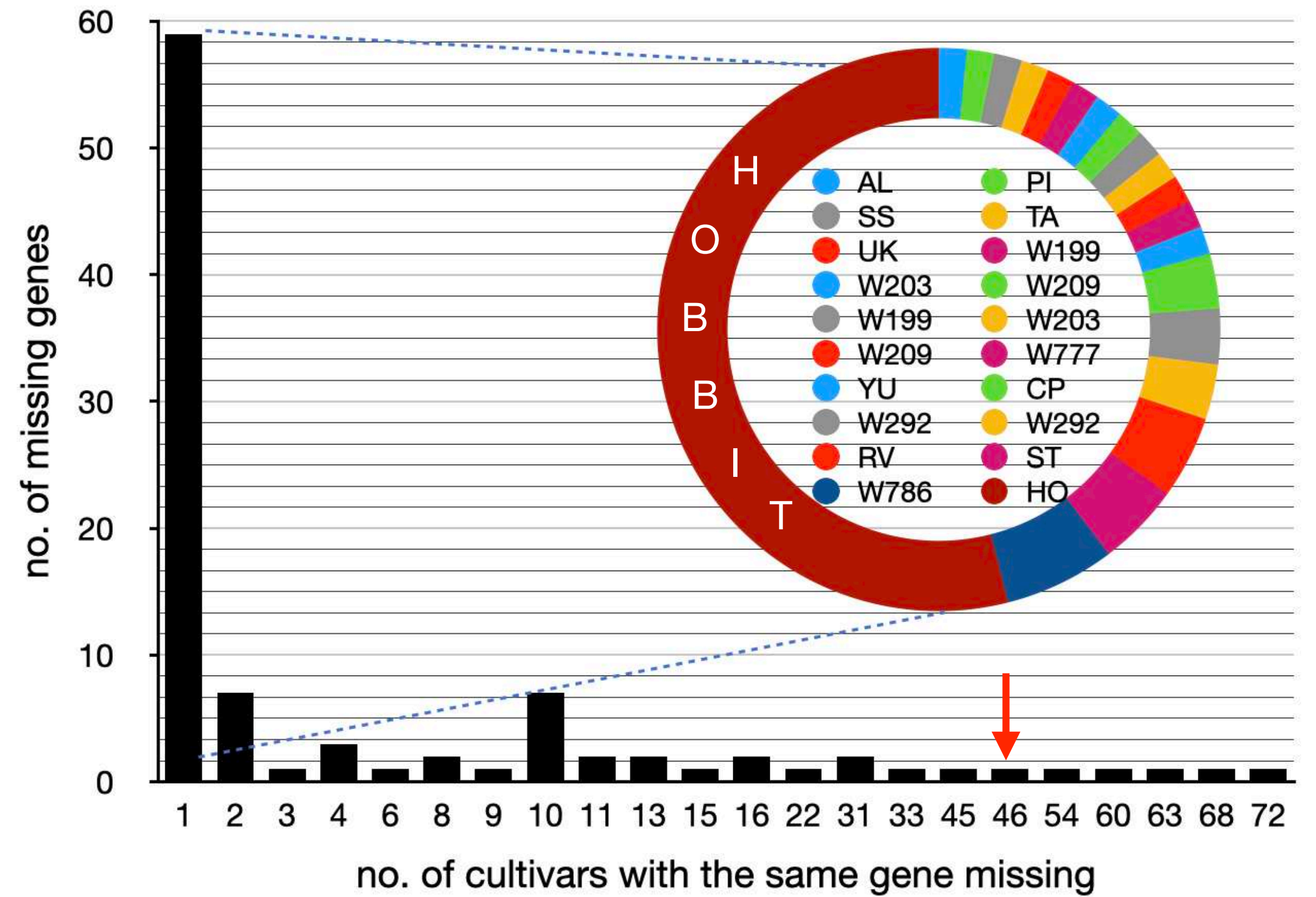
17 deletions: 70% shared amongst up to 83 cultivars

Missing Genes

APL1 B homoeologue (ADP-Glucose Pyrophosphorylase) absent from 46 cultivars

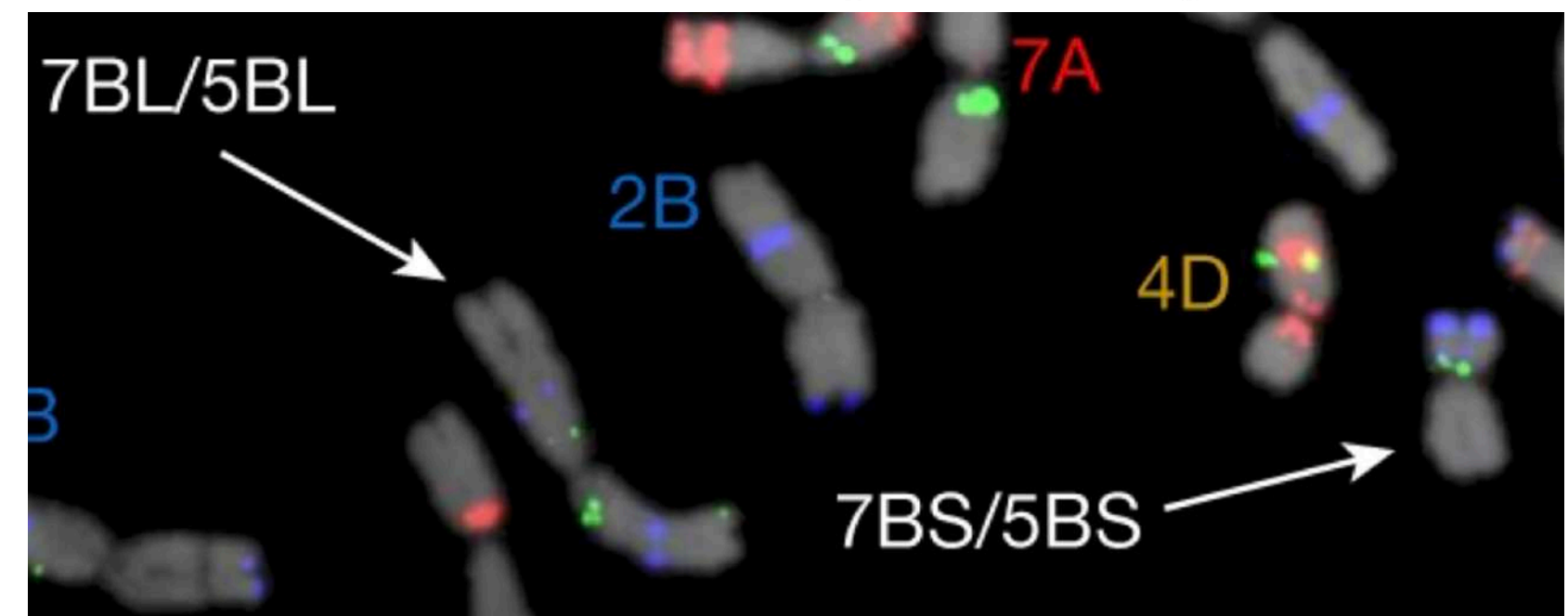
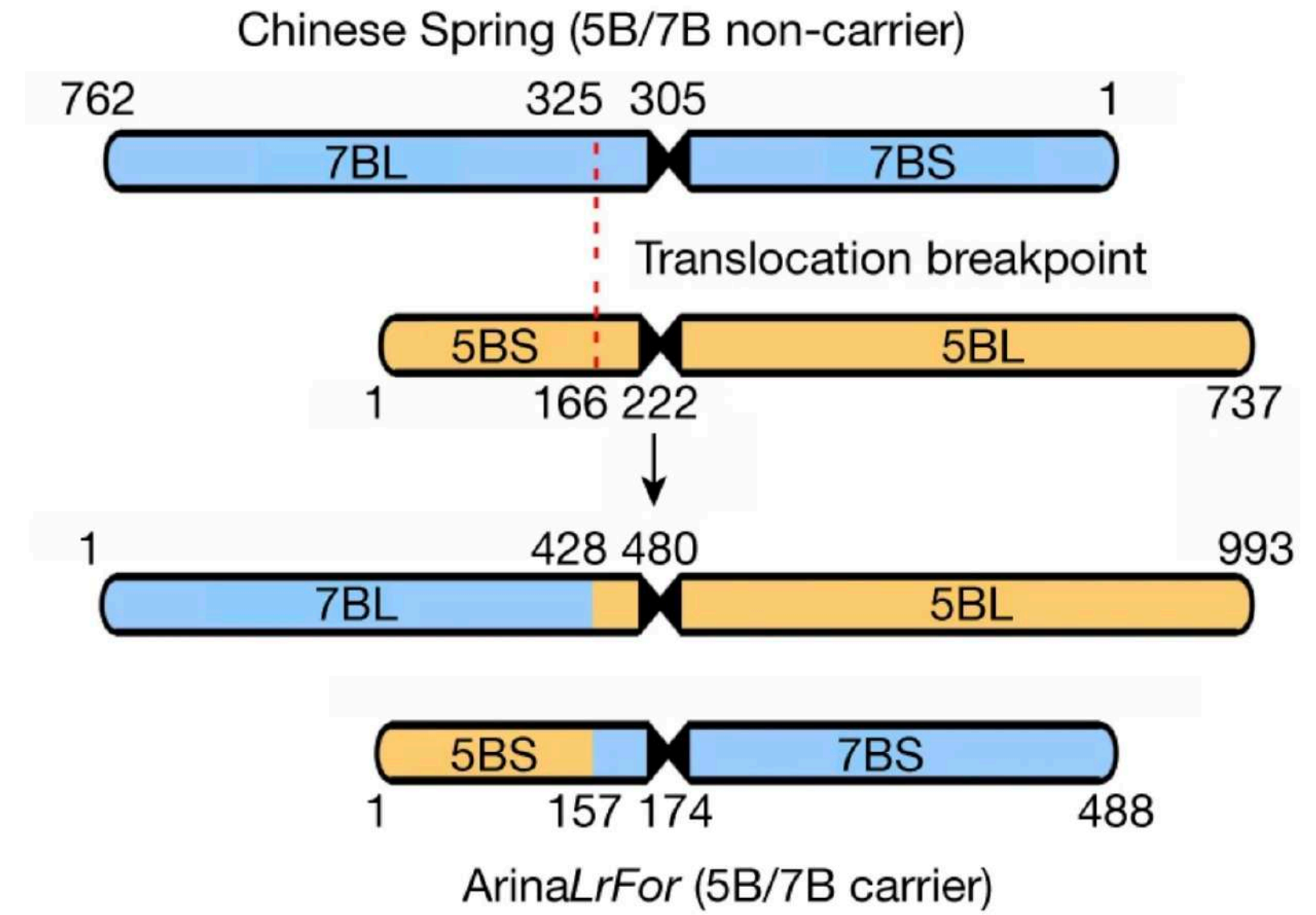
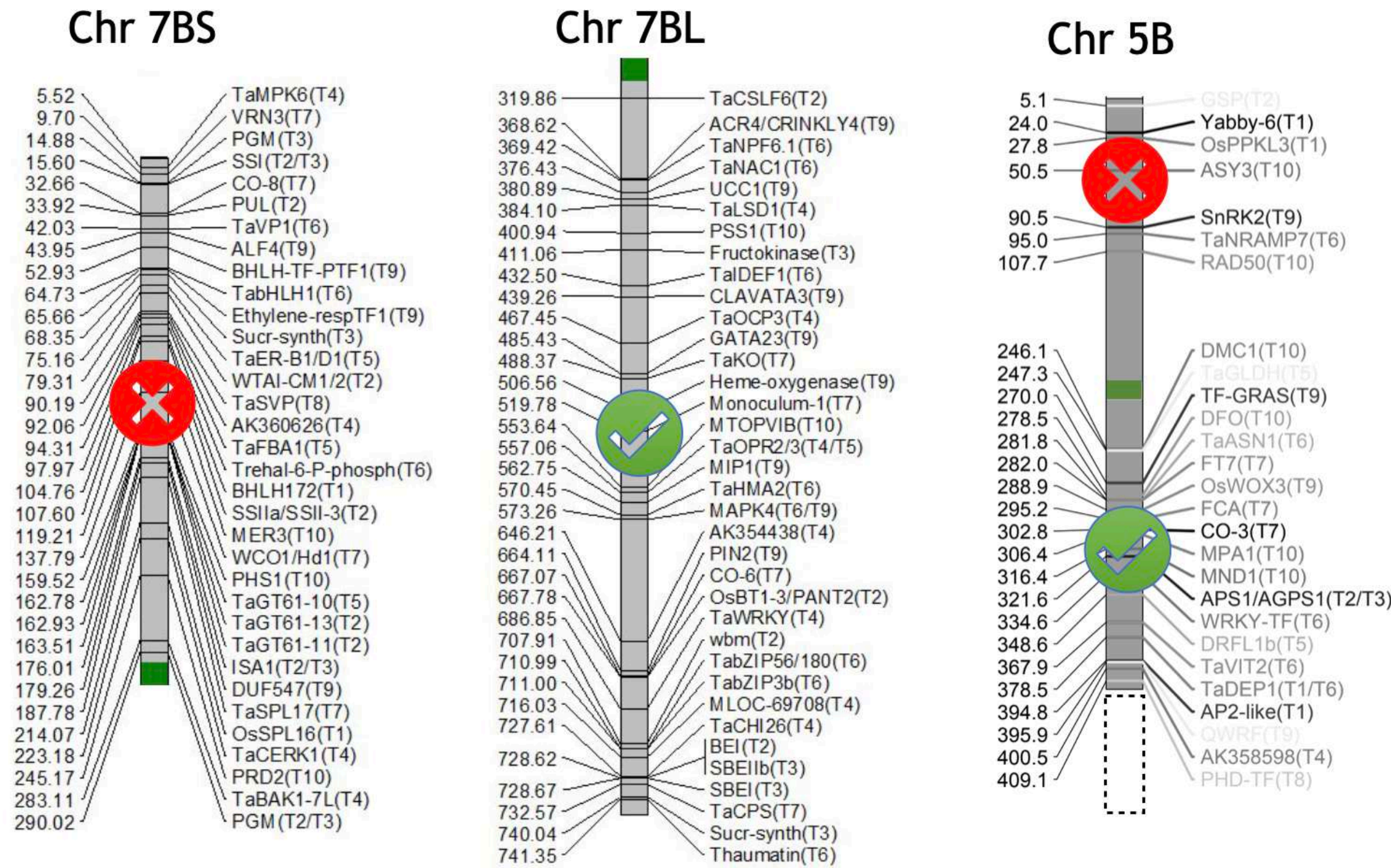


cv. **Hobbit** has by far the most genes missing



our cv. Hobbit is nullisomic for 5BS-7BS translocation

All genes used in this project residing on 7BS & 5BS are missing from cv.Hobbit but all on 7BL & 5BL are present



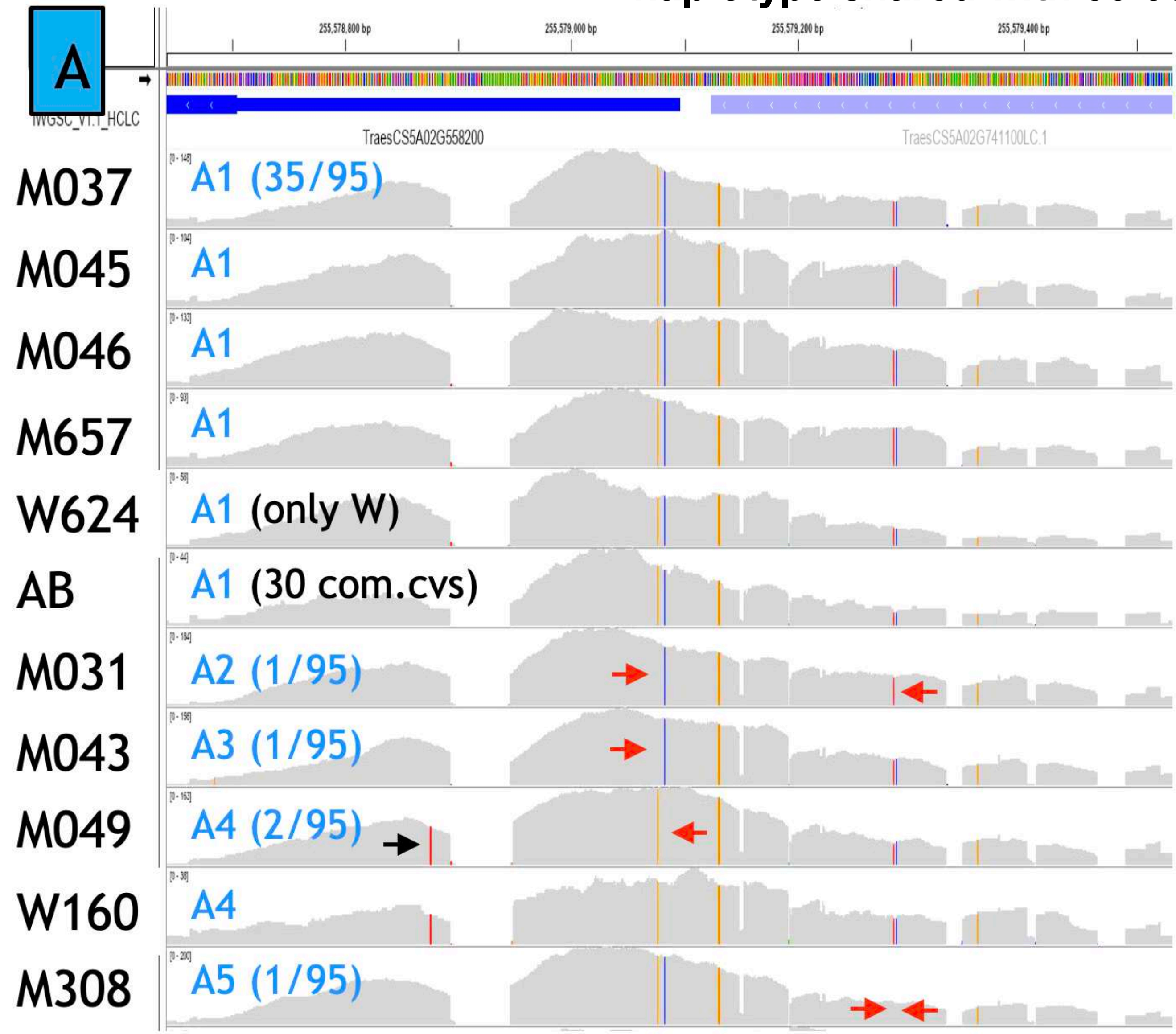
Comparative Study > Nature. 2020 Dec;588(7837):277-283.
doi: 10.1038/s41586-020-2961-x. Epub 2020 Nov 25.

Multiple wheat genomes reveal global variation in modern breeding

Sean Walkowiak # 1 2, Liangliang Gao # 3, Cecile Monat # 4, Georg Haberer # 5, Muluaem T Kassa # 6, Jemima Brinton # 7, Ricardo H Ramirez-Gonzalez # 7, Markus C Kolodziej # 8.

evidence for historic / ancestral introgression

TaGT2L2, an abiotic stress gene homoeologue on Chr5AL has a *Triticum monococcum* (A^mA^m) haplotype shared with 30 commercial cultivars

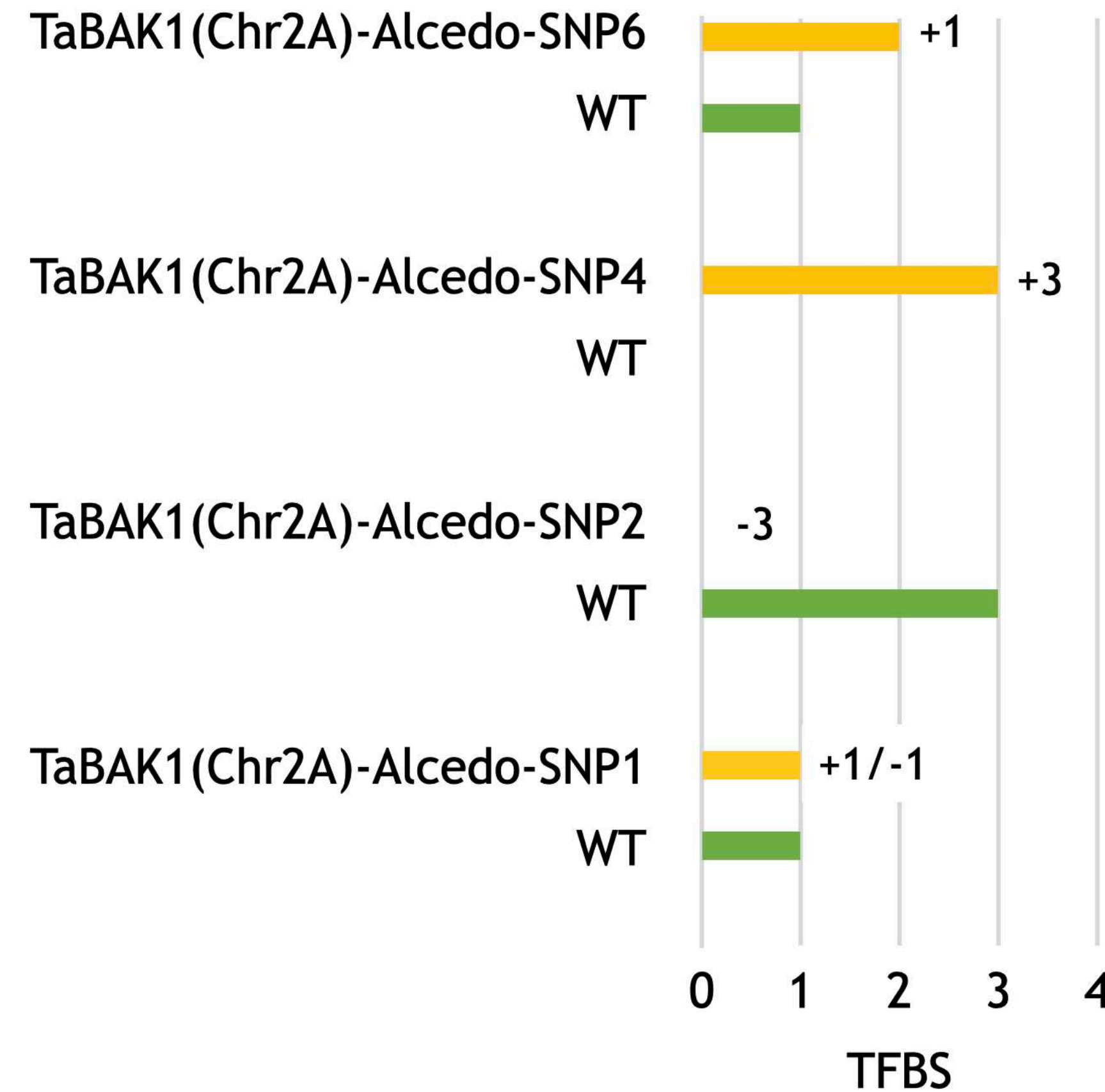
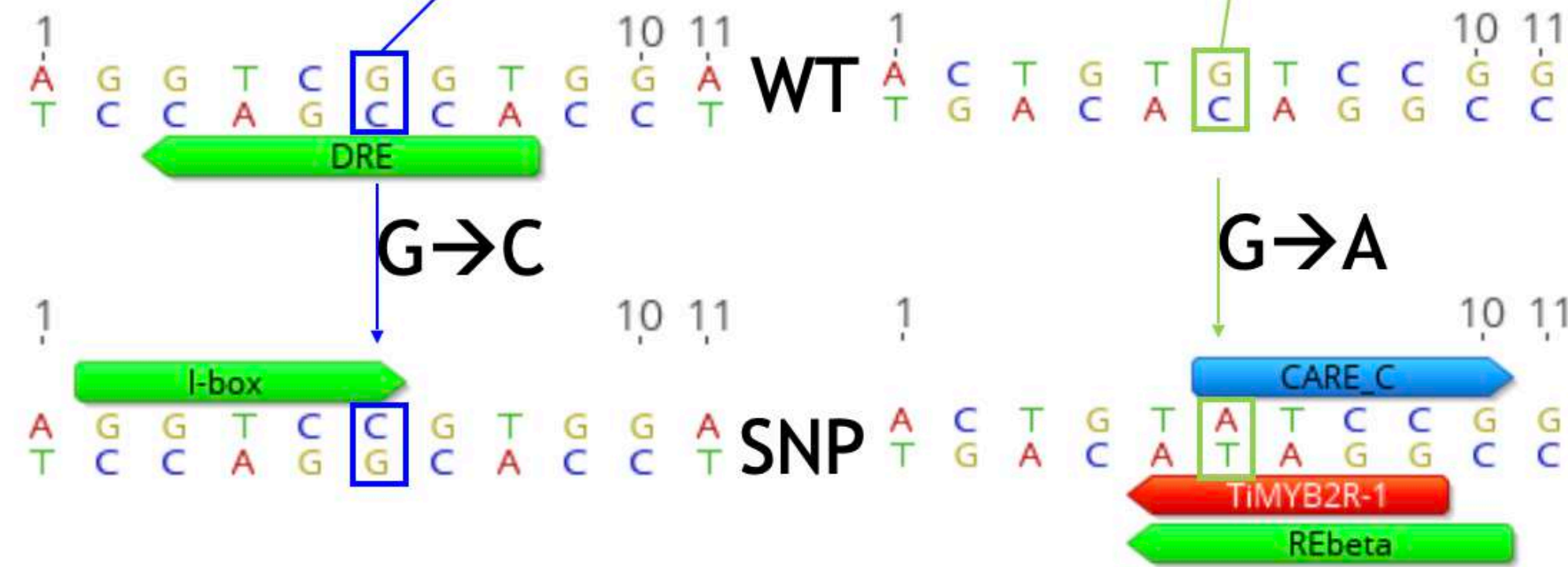
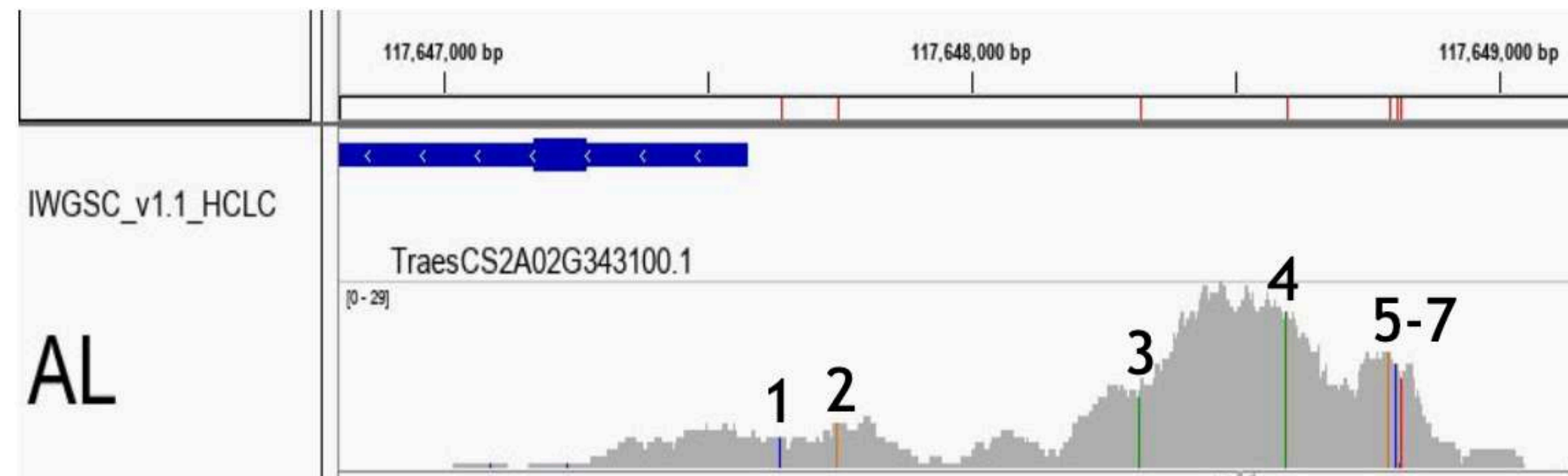


genes with shared T.mon promoter haplotypes reside near telomeres of 5AL, 6AS and 7AS

Many individual SNPs reside within plant Transcription Factor Binding sites (TFBS)

TaBAK1 A homoeologue promoter:

Alcedo haplotype has 7 SNPs of which four reside within TFBS: either changing, creating or abolishing TFBS



NSite-PL (Recognition of PLANT Regulatory motifs with statistics) software online

(<http://www.softberry.com/berry.phtml?topic=nsitep&group=programs&subgroup=promoter>)



Summary

A - technology

1. **MYbaits technology extremely efficient** at capturing wheat gene promoters at **high sequencing depths**
2. **high stringency MYbaits filtration allowed individual capture of homoeologue promoters**
3. Not necessary to completely tile the target sequence with overlapping baits - **25% evenly spaced baits sufficient.**

B - key results

1. Across the diverse set of **83 hexaploid cultivars, ≤ 7 promoter haplotypes**
 2. Most haplotypes contain **≤ 5 or fewer SNPs**
 3. Haplotypes are related: a **core of identical SNPs** \pm a few SNPs
- } evidence for wheat promoter conservation
4. Large promoter deletions map to Transposable Elements
 5. evidence for **ancestral introgression** from *T. monococcum* (more likely indirectly via *T. timopheevii*) to the A sub-genome in many hexaploid wheats
 6. Many of the **SNPs** identified **map to** potential plant **TFBS** and many **small deletions contain TFBS**
 7. No difference in trait categories regarding number of shared and unique haplotypes and SNP diversity.





acknowledgements

(i) **the trait-co-ordinators** who provided the original TGAC gene IDs for individual traits: Cristobal Uauy (JIC, traits 1 & 3), Peter Shewry, Rowan Mitchell (Rothamsted Research (RRes), trait 2), Kay Trafford (NIAB, Cambridge, trait 2), Matthew Moscou (The Sainsbury Laboratory, Norwich, trait 4), Kim Hammond-Kosack (RRes, trait 4), John Foulkes (University of Nottingham (UoN), trait 5), Malcolm Hawkesford (RRes, trait 6), Clare Lister & Simon Griffiths (John Innes Centre (JIC), trait 7), Zoe Wilson (UoN), Jose Fernandez (UoN), Scott Bowden (JIC), trait 8), Malcolm Bennett (UoN, traits 6 and 9) and Peter Buchner (RRes, trait 6 & 9), James Higgins (University of Leicester, trait 10)

(ii) **the suppliers of seed for the 96 chosen cultivars:** Mike Ambrose (Germplasm Resource Unit (GRU) at JIC), Simon Orford (JIC), Jacob Lage (KWS commercial UK based breeder), Lesley Smart (RRes), Clare Lister (JIC), Nick Balaam (Senova Ltd commercial UK based breeder), Kay Trafford (NIAB), Simon Berry (Limagrains commercial UK based breeder), Ruth Bryant (RAGT). Special thanks to Mike Ambrose, the head of the GermPlasm Resource Unit (GRU, Norwich, UK) (since retired), who provided the majority of cultivars within 24 h after selection on their excellent website.

(iii) **Promoter Capture and sequencing:** Alison Devault & Jacob (Jake) Enk at Daicel Arbor Biosciences (formerly MYcroArray, <https://DaicelArborbiosci.com>).

(iv) **The IWGSC** for allowing pre-publication access to the complete IWGSC_refseq_v1.0,

(v) **the entire WGIN 3 Management Team** (Wheat Genetic Improvement Network, <http://www.wgin.org.uk>) which during design of this project consisted of Andrew Riche (RRes), Clare Lister (JIC), David Feuerhelm (Syngenta), Dhan Bhandari (AHDB), Edward Flatman (Limagrains), Gia Aradottir (RRes), Jacob Lage (KWS), Kim Hammond-Kosack (RRes), Kostya Kanyuka (RRes), Lesley Smart (RRes), Malcolm Hawkesford (RRes), Martin Cannell (Defra), Matthew Kerton (DSV UK), Michael Hammond-Kosack (RRes), Peter Shewry (RRes), Richard Jennaway (Saaten-Union), Ruth Bryant (RAGT), Sarah Holdgate (NIAB), Simon Berry (Limagrains), Simon Griffiths (JIC), Simon Penson (Campden BRI), Stephen Smith (Elsoms UK), Vanessa McMillan (RRes)

(vi) **glasshouse staff at Rothamsted Research:** Jill Maple, Jack Turner, Tom Yaxley

(vii) **laboratory expertise:** Carlos Bayon (RRes) and Martin Urban (RRes) for reminding MHK how to do things in the laboratory

(viii) Keywan Hassani-Pak (RRes) for initial advice on Bioinformatics, Dan Smith (RRes), Keith Edwards (University of Bristol), Kay Trafford (NIAB), Chris Burt (RAGT) and Simon Berry (Limagrains) for **helpful discussions** and for sharing knowledge gained through their own research.

(ix) **Defra** (Department for Environment, Food & Rural Affairs) for funding this project as part of WGIN





Exploring the diversity of promoter and 5'UTR sequences in ancestral, historic and modern wheat

Michael C.U. Hammond-Kosack^{1,†}, Robert King², Kostya Kanyuka¹  and Kim E. Hammond-Kosack^{1,*,†} 

¹Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, UK

²Department of Computational and Analytical Sciences, Rothamsted Research, Harpenden, UK

Received 7 March 2021;

revised 15 June 2021;

accepted 8 July 2021.

*Correspondence (Tel +44 (0)1582 763133;
email kim.hammond-kosack@rothamsted.

ac.uk, michael.hammond-

kosack@rothamsted.ac.uk)

[†]Co-corresponding authors.

Keywords: promoter capture,

Triticum aestivum, *Triticum*

monococcum, Watkins landraces,

agronomic traits, sequence variation,

haplotypes, transposable elements

(TE), repetitive elements (RE),

transcription factor binding sites

(TFBS).

Abstract

A data set of promoter and 5'UTR sequences of homoeo-alleles of 459 wheat genes that contribute to agriculturally important traits in 95 ancestral and commercial wheat cultivars is presented here. The high-stringency myBaits technology used made individual capture of homoeo-allele promoters possible, which is reported here for the first time. Promoters of most genes are remarkably conserved across the 83 hexaploid cultivars used with <7 haplotypes per promoter and 21% being identical to the reference Chinese Spring. InDels and many high-confidence SNPs are located within predicted plant transcription factor binding sites, potentially changing gene expression. Most haplotypes found in the Watkins landraces and a few haplotypes found in *Triticum monococcum*, germplasms hitherto not thought to have been used in modern wheat breeding, are already found in many commercial hexaploid wheats. The full data set which is useful for genomic and gene function studies and wheat breeding is available at <https://rrescloud.rothamsted.ac.uk/index.php/s/DMCFDu5iAGTI50u/authenticate>.

1) The full post bioinformatics data set used here is available at <https://rrescloud.rothamsted.ac.uk/index.php/s/DMCFDu5iAGTI50u/authenticate>.

2) Raw sequencing reads have been deposited in the ENA database under **BioProject PRJEB45647**.

contact for further information and access to Owncloud:
wgin.defra@rothamsted.ac.uk

Plant Biotechnology Journal

Open Access

Volume 19 • Issue 12 • December 2021

