



Department  
for Environment  
Food & Rural Affairs



# **WGIN Wheat Promotome Capture**

**A sub-contractor project of WGIN 3**

**Michael Hammond-Kosack (WGIN PA)**

**15<sup>th</sup> WGIN Stakeholders' Meeting 30<sup>th</sup> November 2017 @ RRes**

## What is Promotome Capture & what are the **Main Objectives** for WGIN ?

### What Is It?

#### ~~Exome~~ Promotome Capture

“~~Exome~~ Promotome capture is a method used to extract and sequence the ~~exome~~ **promotome** (collection of all ~~exons~~ **promoters**) in a genome and compare this variation across a sample of individual organisms (**wheat cultivars**). This allows studies to quickly focus in on the small percent of the genome that is most likely to contain variation that strongly affects phenotypes of interest and/or to identify rates of ~~codon~~ **promoter** evolution between a set of species to infer the effects of mutation and selection among genes.” ([http://hawaiireedlab.com/gwiki/index.php?title=Exome\\_Capture](http://hawaiireedlab.com/gwiki/index.php?title=Exome_Capture))

### Main Objectives

- Comparison of promoters of genes of interest in a large number of wheat cultivars used in breeding, farming & scientific research
- Identification of cis-acting elements important for gene regulation
- Linking this comparison to phenotypic and transcriptomics data to increase our understanding of gene regulation to generate these phenotypes

## Complete workflow:



info@mycroarray.com | 1-734-998-0751



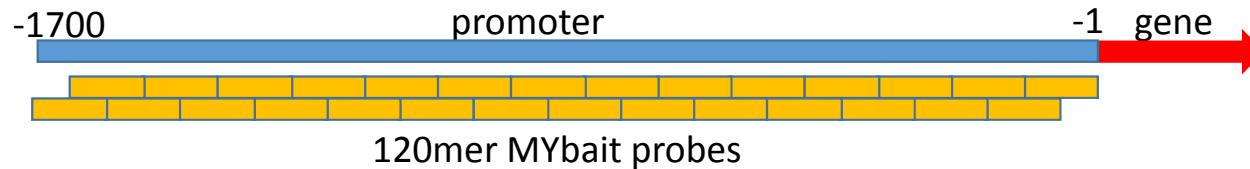
PLEASE NOTE: MYcroarray is now called **Arbor Biosciences**

## Time Line

- 1 • FASTA with 1,402 promoter sequences submitted to MYcroarray **May 10<sup>th</sup>**
- 2 • MYbaits synthesis (@ Mycroarray) completed **August 24<sup>th</sup>**
- 3 • 96 DNA samples (80%, up to 8µg) posted **September 11<sup>th</sup>**
- 4 • MYbaits Capture: Library preps started **September 14<sup>th</sup>**
- 5 • Next Generation Sequencing @ MYcroarray: **October**
- 6 • Data received @ RRes **November 10<sup>th</sup> 2017**
- 7 • Bio-Informatics **November 2017 onwards**

## WGIN Funding & Design - Traits, Trait-Coordinator & Wheat Cultivars

- Defra's WGIN 3 funding for this sub-contractor project allowed for the use of up to **40,000 MYbait probes** and **96 "Reactions"** (which equates to **96 Wheat Cultivars**)
- **MYbaits set-up: 1700bp promoter** uses **28 MYbaits (120bp)** with **2 fold** coverage :



- 3 wheat homoeologues (A, B, D)    **3 x 28 MYbait probes / gene**
- Thus Defra's funding will allow for the capture of **1428 promoter sequences**.
- The WGIN Management Team decided on **10 Trait Categories**,  
and thus **≤143 promoter sequences per trait**

## Traits & Trait-Coordinators

**1. Yield Resilience** – Cristobal Uauy (JIC)

**2. Grain Composition** – Peter Shewry, Rowan Mitchell (RRes), Kay Trafford (NIAB)

**3. Grain Development** – Cristobal Uauy (JIC)

**4. Biotic Stress (fungi and insects)** – Matthew Moscou (TSL), Kim Hammond-Kosack (RRes)

**5. Abiotic Stress (drought, high temp)** – John Foulkes (UoN)

**6. Nutrient Use Efficiency** – Malcolm Hawkesford (RRes)

**7. Canopy Development / Whole Plant Architecture** – Clare Lister & Simon Griffiths (JIC)

**8. Flower biology** – Zoe Wilson (UoN), Scott Bowden (JIC)

**9. Root architecture** – Malcolm Bennett (UoN), Peter Buchner (RRes)

**10. Recombination** – James Higgins (Leicester Uni)

# The 96 Wheat Cultivars

**Chosen by WGIN Management Team Members and Trait Coordinators**

ID	cultivar	nominators
1	A. speltooides 2140022	LS/GA
2	Abbot	SB
3	Aegilops peregrina (variabilis)	KT
4	Alcedo	RAGT
5	Ambrosia	SB
6	Avalon	SG/MH/VM/SB/CL/EO+JF
7	Badger	CL
8	Bobwhite	KK
9	Brompton	LG
10	Buster	SB
11	Cadenza	SG/MH/VM/SB/CL
12	Cellule	RAGT
13	Charger	SB
14	Chinese Spring	SG
15	Claire	RAGT/SG/PS/MH/CL
16	Coppadra	KK
17	Cordiale	PS/MH
18	Cougar	RAGT
19	Crusoe	RAGT/MH
20	Dickens	LG
21	Einstein	SB
22	ENT-228	LS/GA
23	Fielder	KK
24	Flanders	RAGT
25	Gallant	MH
26	Garcia	SG/CL
27	Gatsby	ECS
28	Gladiator	SB
29	Graham	LG/syn
30	Hereford	KHK/VM
31	Hereward	RAGT/PS/MH/SB/CL
32	Hobbit	SB

ID	cultivar	nominators
33	Hustler	SB
34	Isengrain	PS
35	Istabraq	PS/MH
36	JB Diego	RAGT
37	Kronos	WGIN MM team
38	KWS Santiago	RAGT
39	KWS Silverstone	LG
40	KWS Siskin	RAGT
41	KWS Trinity	LG
42	Malacca	PS/MH/CL
43	Maris Huntsman	SB
44	Maris Widgeon	MH
45	Marksman	PS
46	Mercia	MH
47	Napier	CL
48	Oakley	RAGT
49	Paragon	SG/MH/AR/VM
50	Piko	RAGT
51	Reflection	LG
52	Relay	RAGT
53	Revelation	LG
54	Rialto	RAGT/SB/PS
55	Riband	MH/KK
56	Robigus	RAGT/SG/MH/AR
57	Savannah	CL
58	Scout	RAGT
59	Sear Synthetic	SG
60	Skyfall	RAGT
61	Soisson	RAGT/PS/MH
62	Solstice	RAGT/GA_LS/MH
63	Spark	PS/CL
64	Stigg	MH

ID	cultivar	nominators
65	Sumai 3	RAGT
66	<i>T. monoccocum</i> MDR031	KHK/VM
67	<i>T. monoccocum</i> MDR037	LS/GA/VM
68	<i>T. monoccocum</i> MDR043	KHK/VM
69	<i>T. monoccocum</i> MDR045	LS/GA
70	<i>T. monoccocum</i> MDR046	RRES/VM
71	<i>T. monoccocum</i> MDR049	LS/GA
72	<i>T. monoccocum</i> MDR308	KHK
73	<i>T. monoccocum</i> MDR657	LS/GA
74	Taichung 29	KK/JRudd
75	Ukrainka	PS
76	USU-Apogee	KK
77	Valoris	PS
78	Veranopolis	KK
79	Watkins 115	LS/GA
80	Watkins 141	SG
81	Watkins 160	SG
82	Watkins 199	LS/GA/SG
83	Watkins 203	VM
84	Watkins 239	SG
85	Watkins 209	
86	Watkins 246	SG
87	Watkins 292	SG
88	Watkins 387	SG
89	Watkins 579	LS/GA
90	Watkins 624	LS/GA
91	Watkins 733	VM
92	Watkins 777	VM/SG
93	Watkins 786	VM
94	Xi19	RAGT/PS/MH
95	Yumai 34	PS
96	Zebedee	EO+JF

# Some of the Hexaploid (AABBDD) Wheat Cultivars





A genome relatives (A<sup>m</sup>A<sup>m</sup>)



B genome relative (SS)



Tetraploid (AABB) – Durum wheat

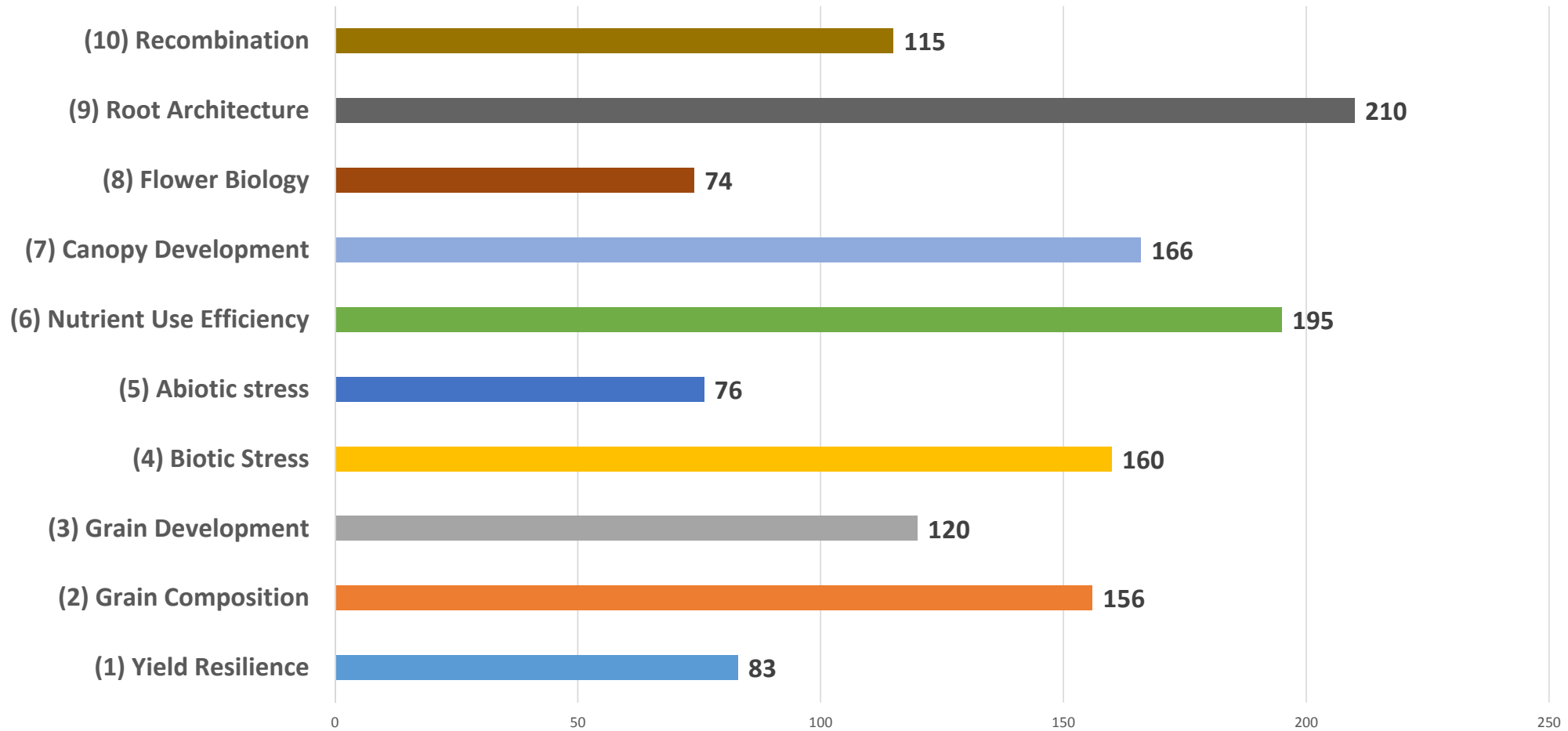


D genome – Aegilops tauschii (DD)



## Gene IDs received

A total of **1355 Gene IDs** distributed as:





## Changes in Chromosome locations between Ensembl (TGAC) & IWGSC RefSeq v1.0

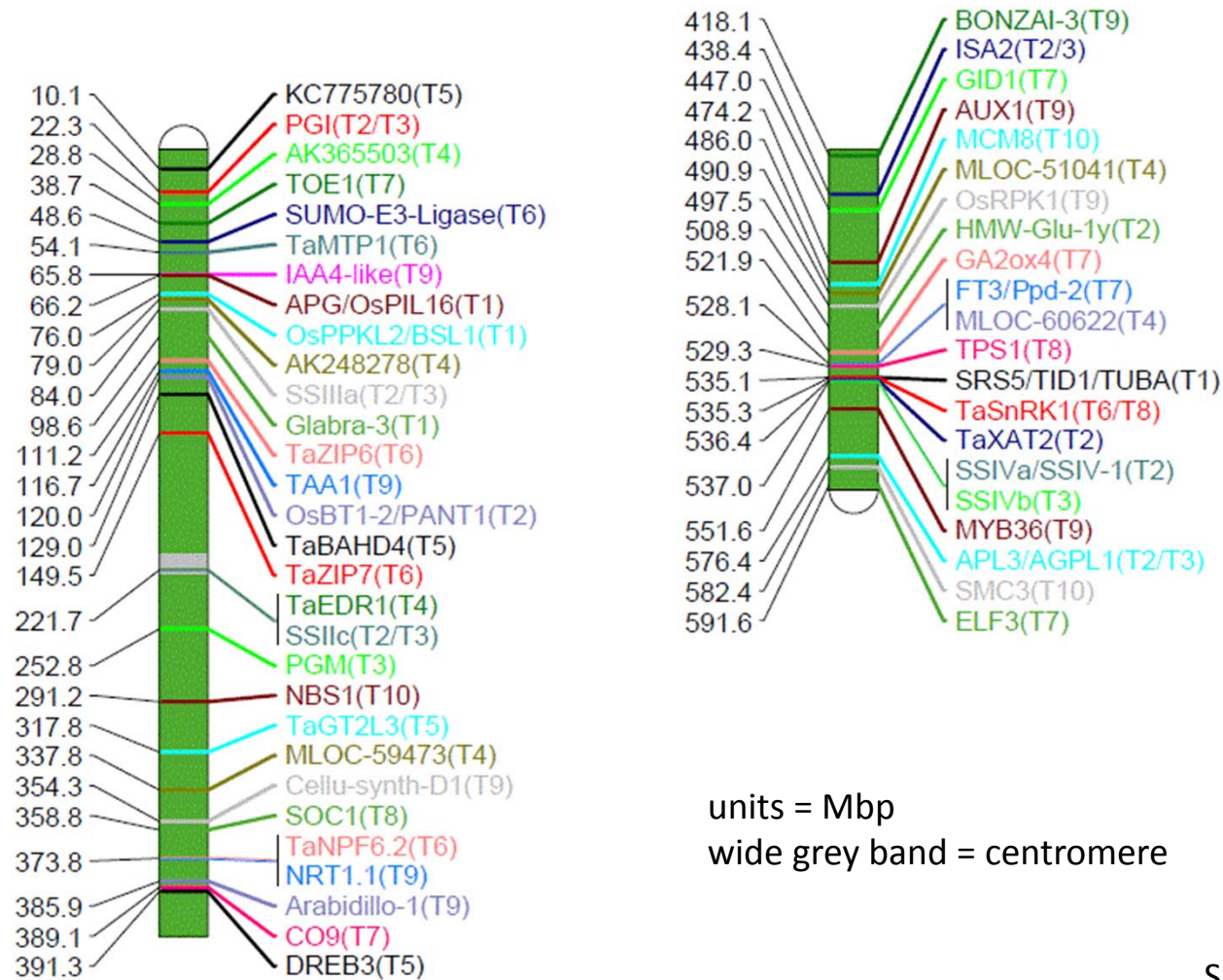
trait&gene	homoeologue	Ensembl	IWGSC
T1-18	A	5A	U
T1-5	A	2A	U
T2-1	D	U	4D
T2-24	D	5A	7D
T2-40	A	U	1A
T2-53	B	1B	4A
T2-69	B	6B	4D
T2-72	B	6B	4B
T2-74	B	4B	U
T3-1	B	6A	6B
T3-28	B	1B	4A
T3-5	A	U	7A
T4-17	B	1B	1B & 1D
T4-24	D	U	4D
T4-25	D	U	4D
T4-3	B	U	4B
T4-32	B	U	7B
T4-43	D	U	3D
T4-45	A	3A	7A
T4-5	B	U	2B
T5-10	D	4D	U
T6-23	D	U	3D
T6-31	D	U	4A
T6-36	D	2D	U & 6B
T6-42	B	3B	7A
T6-49	A	U	7A
T6-68	A	2A	U
T6-9	B	1D	1B

trait&gene	homoeologue	Ensembl	IWGSC
T7-31	D	U	7D
T7-35	B	U	3B
T7-36	B	2B	U
T7-45	B	U	5B
T7-46	D	U	7D
T7-47	A	U	3A
T7-48	B	U	3B
T7-5	D	4D	U
T8-17	A	2A	2D & U
T8-18	B	U	5B
T8-3	D	U	5D
T8-7	D	U	6D
T9-5	D	5D	2D
T9-72	A	4A	U
T9-77	A	U	7A
T10-11	A	4A	4D
T10-13	D	U	5D
T10-38	D	U	5D
T10-39	B	4B	U
T10-43	A	U	3A
T10-5	A	U	7A
T10-9	B	U	1B

for 32 genes (64%) change completes ABD

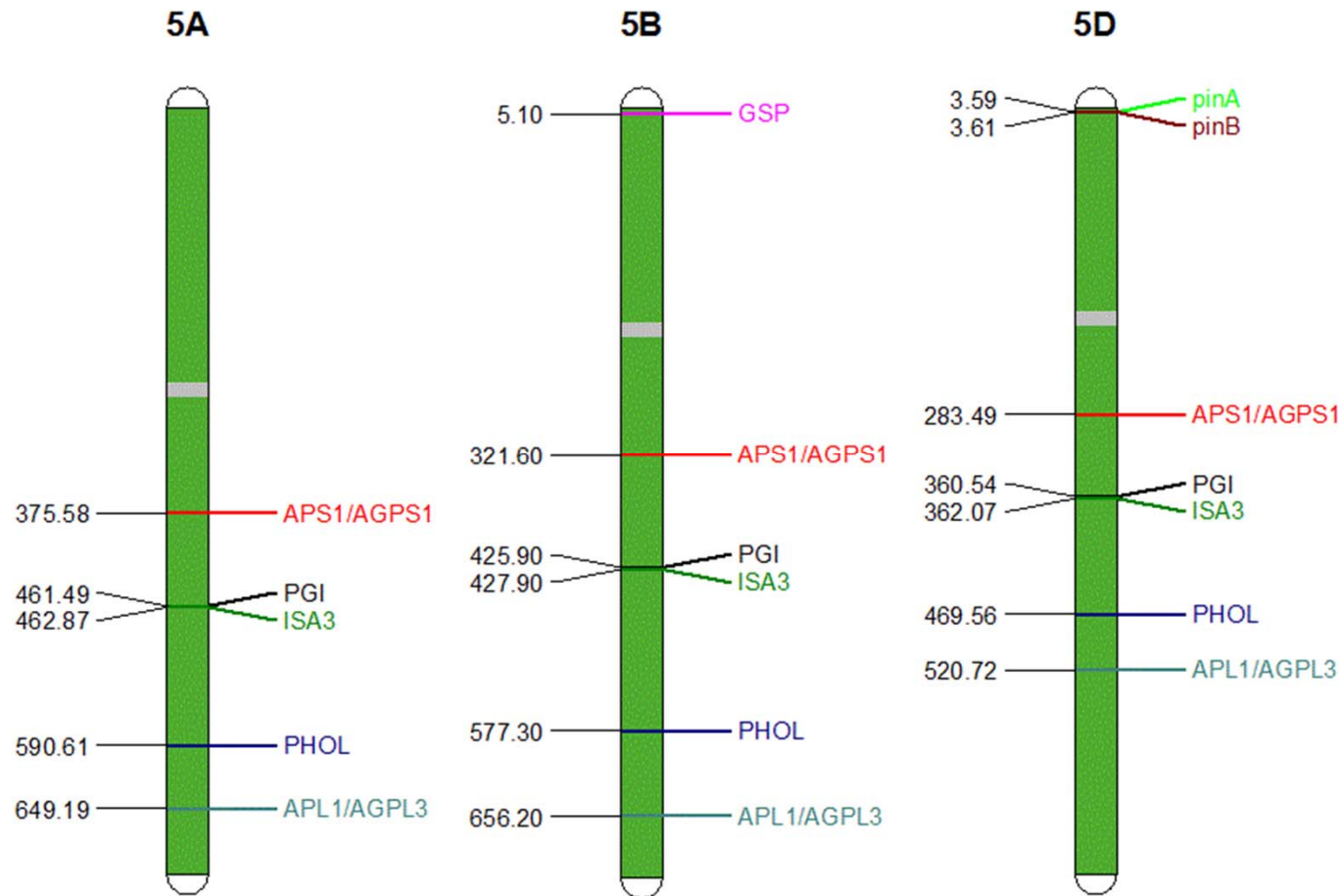
for 18 genes (36%) change disrupts ABD and/or adds allele

## Physical Location of All 10 Trait Genes on Chromosome 1A

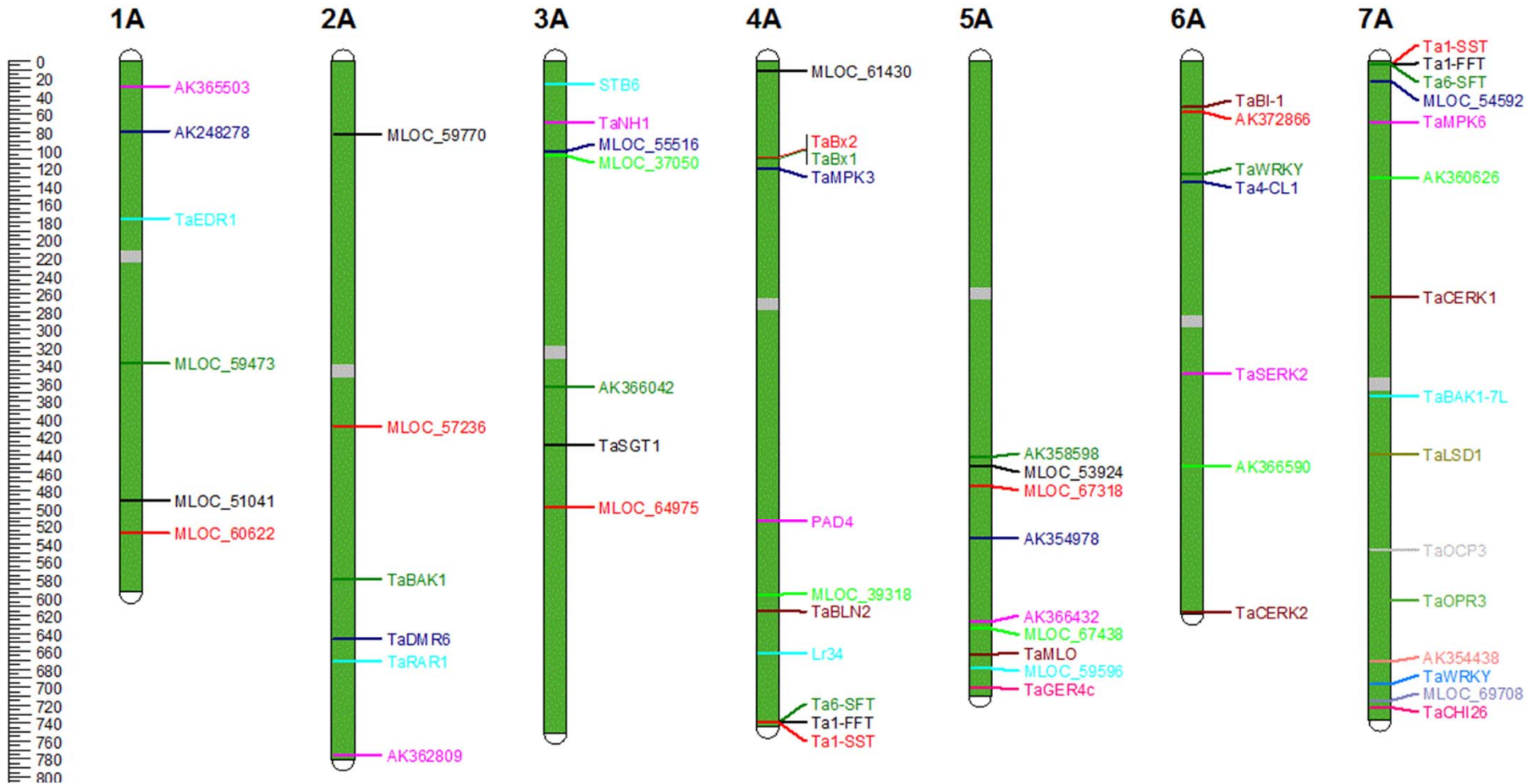


Software used: MapChart

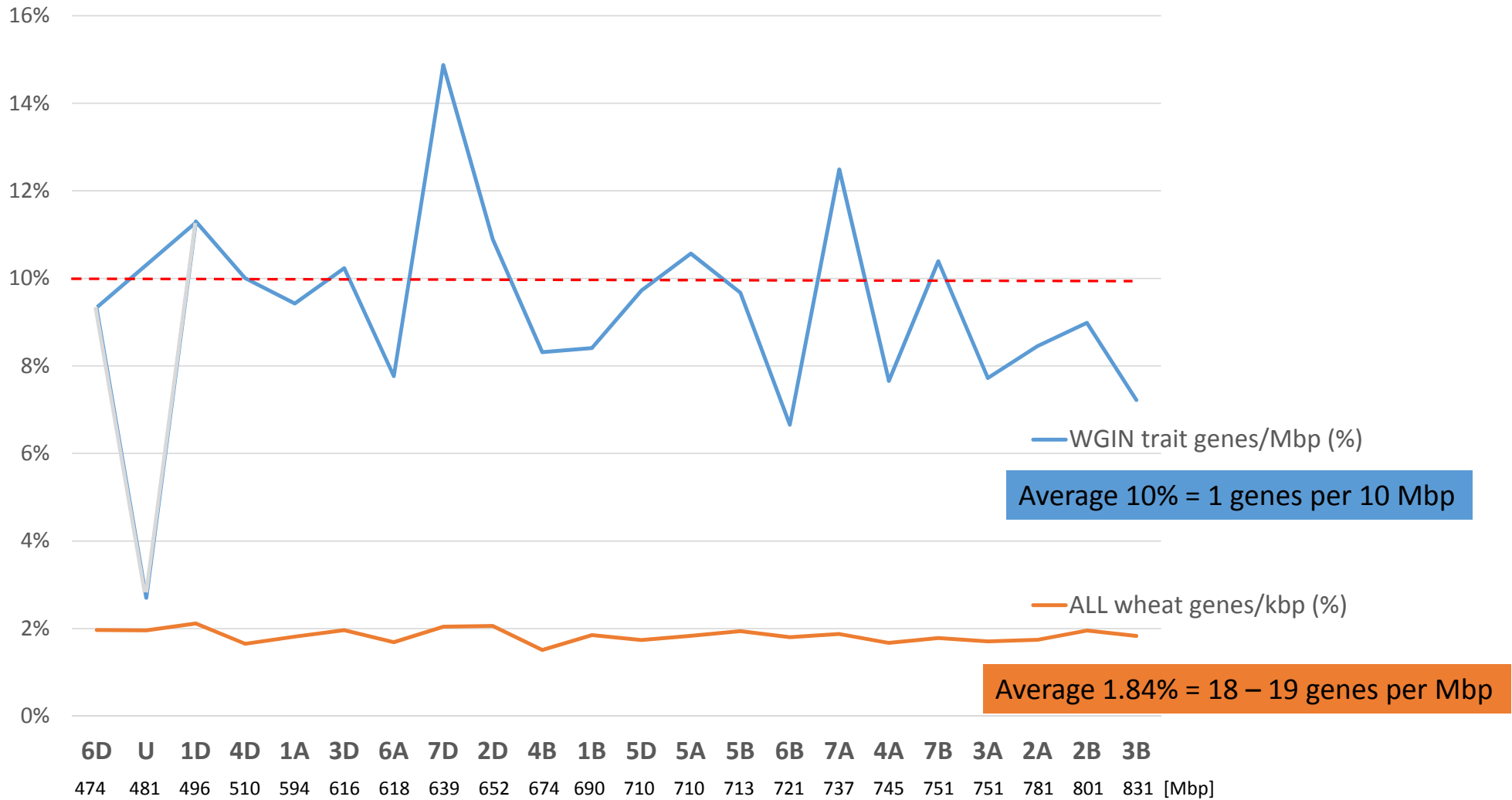
## Grain Composition (Trait 2) – **Physical** Position of Chromosome 5 Homoeologues



## Biotic Stress (Trait 4) – Physical Position of Genes on A Genome



## Distribution of WGIN Trait Genes in Wheat Genome is NOT even





## MYbaits filtration to find baits (primers) specific for each homoeologue

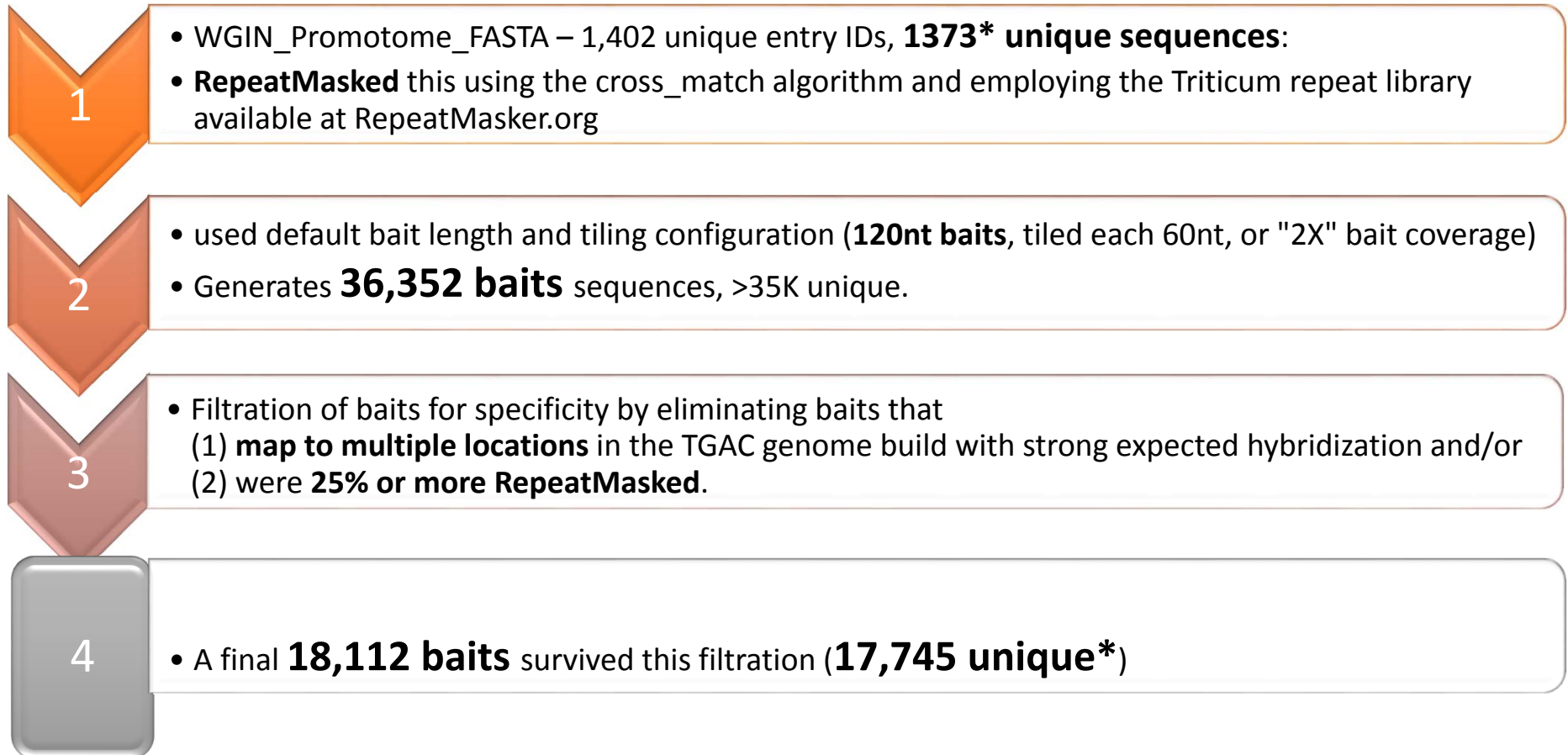
### MYbaits Coverage

How many baits are **specific** for each individual promoter sequence?

trait&gene no.	gene	WGIN Promotome ID	MYbaits	MYbaits coverage (%)	homoeologue ID (%)
T4-1	TaCERK1	7A-1826280-T4-1	18	74.4	29.8
T4-1	TaCERK1	7B-1934340-T4-1	11	51.3	
T4-1	TaCERK1	7D-2045100-T4-1	9	39.8	
T4-10	TaMPK3	4A-1008960-T4-10r	7	41.6	57.5 but 84.7 to [-800]
T4-10	TaMPK3	4B-1033300-T4-10	14	65.2	
T4-10	TaMPK3	4D-1148960-T4-10r	12	58.4	
T4-11	TaMPK6	7A-1834630-T4-11r	4	21.6	45.5 but 69.9 to [-800]
T4-11	TaMPK6	7B-1954300-T4-11r	5	25.4	
T4-11	TaMPK6	7D-2040040-T4-11r	6	32.5	
T4-12	TaOCP3	7A-1770880-T4-12	13	52.2	24.3
T4-12	TaOCP3	7B-1867300-T4-12r	18	77.6	
T4-12	TaOCP3	7D-2004100-T4-12r	22	91.5	
T4-13	TaOPR3	7A-1801960-T4-13	8	43.3	49.1
T4-13	TaOPR3	7B-1860490-T4-13	7	29.6	T4-13 = T5-14???
T4-13	TaOPR3	7D-2008410-T4-13	6	25.9	
T4-14	TaNH1	3A-0688550-T4-14	14	65.9	42
T4-14	TaNH1	3B-0744050-T4-14	10	43.1	
T4-14	TaNH1	3D-0914040-T4-14	12	58.4	
T4-15	TaBI-1	1D-0240360-T4-15	7	28.5	25.1
T4-15	TaBI-1	6A-1573990-T4-15	15	76.2	
T4-15	TaBI-1	6B-1671980-T4-15r	17	82.6	
T4-16	TaLSD1	7A-1798270-T4-16	6	36	46.5 but 86.9 to [-580] and <b>100%</b> (1080-1176)
T4-16	TaLSD1	7B-1864900-T4-16	15	72.9	
T4-16	TaLSD1	7D-1974840-T4-16	14	71.1	

## Mybaits Filtration

(performed by Dr. Jacob Enk @ MYcroarray)



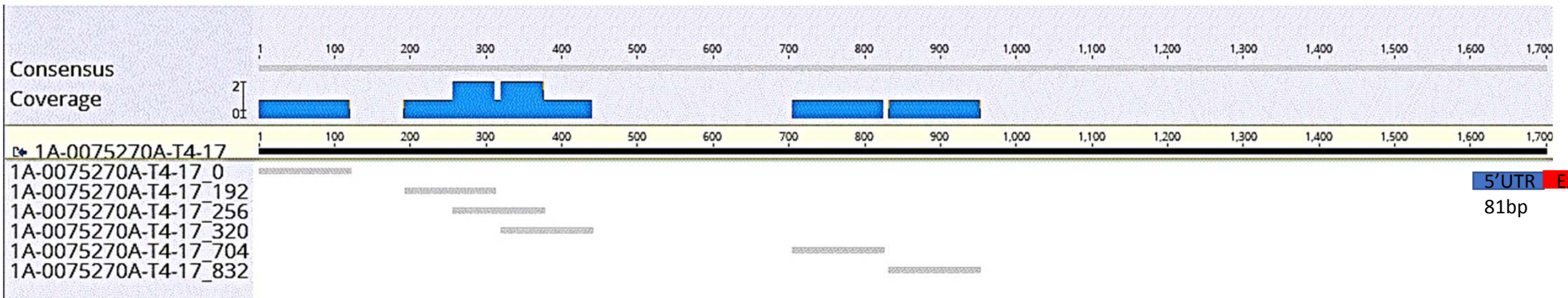
\* differences: same gene ID provided for multiple traits -> seqs removed from all but one trait

**MYbaits filtration for 1A-0075270A-T4-17**  
**only 6 Baits pass stringent criteria**

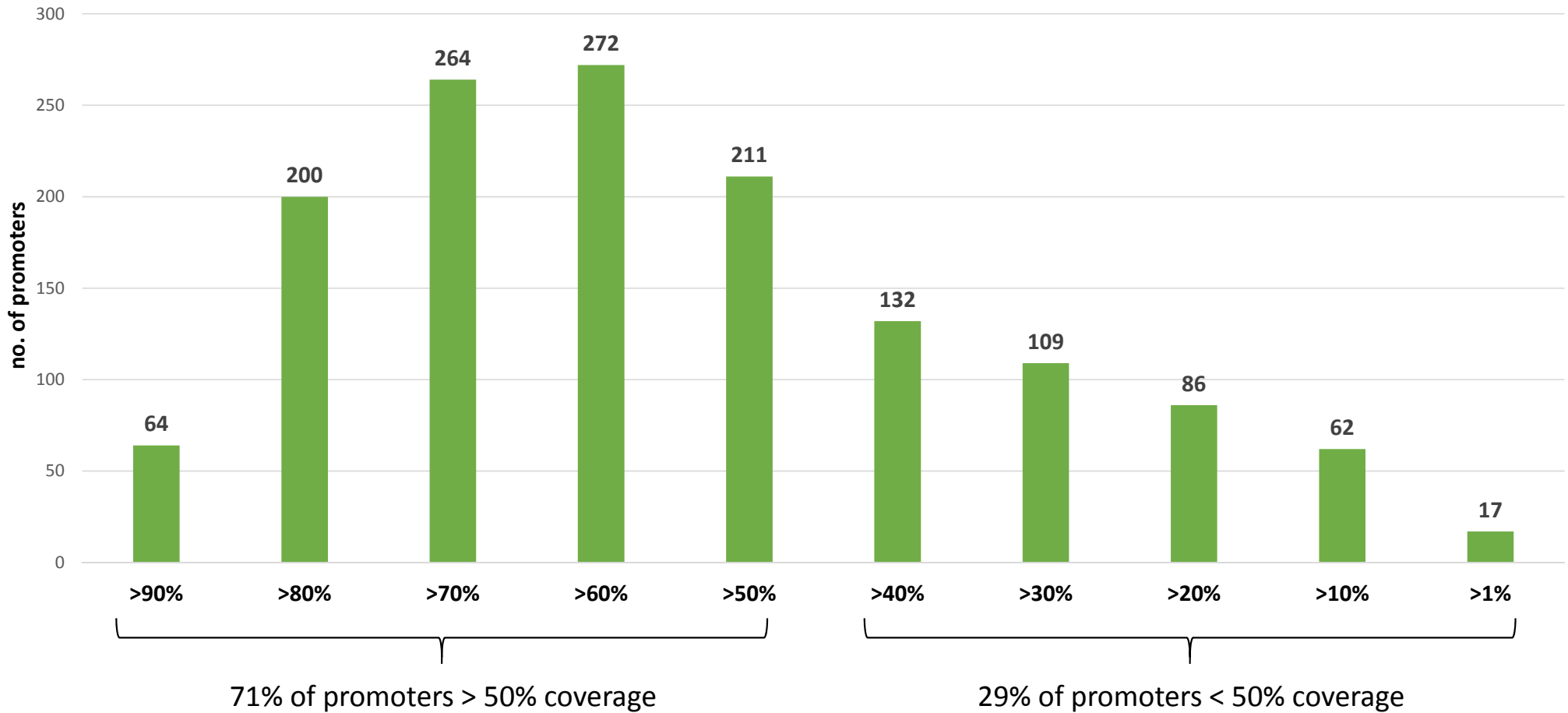
Bait	%GC	%RM	BLAST-Hits	40-60C	60-62.5C	62.5-65C	65-67.5C	67.5-70C	>70C	Stringent	Moderate	Relaxed	Sequenc
1A-0075270A-T4-17_0	35	0	32	17	0	6	0	0	0	pass	pass	pass	TGCTATGATGC
1A-0075270A-T4-17_64	34.2	0	25	13	1	7	1	0	0	fail	pass	pass	AAATATTTTCA
1A-0075270A-T4-17_128	35.8	0	40	31	1	6	0	0	0	fail	fail	fail	CCTCTATTCTC
1A-0075270A-T4-17_192	34.2	0	39	25	1	8	0	0	0	pass	pass	pass	TACTTGATTTT
1A-0075270A-T4-17_256	29.2	0	21	9	5	4	0	0	0	pass	pass	pass	AATCTAGGTAT
1A-0075270A-T4-17_320	29.2	0	22	8	9	3	0	0	0	pass	pass	pass	ATGAGGATAA
1A-0075270A-T4-17_384	40.8	0	33	15	2	1	6	3	0	fail	fail	fail	CATTAATGCTA
1A-0075270A-T4-17_448	39.2	0	15	3	1	4	6	0	0	fail	fail	fail	AAAAGTGCTA
1A-0075270A-T4-17_512	39.2	0	23	7	0	2	6	4	0	fail	fail	fail	TTTCAGATTCT
1A-0075270A-T4-17_576	45.8	0	17	2	0	0	1	10	1	fail	fail	fail	CAAAACCTTTG
1A-0075270A-T4-17_640	41.7	0	14	0	2	3	8	0	0	fail	fail	fail	TCTACTCACAT
1A-0075270A-T4-17_704	35	0	15	1	8	5	0	0	0	pass	pass	pass	TTACTTGCAAT
1A-0075270A-T4-17_768	35	0	22	3	2	6	6	0	0	fail	fail	fail	TAACTTATATC
1A-0075270A-T4-17_832	34.2	0	43	18	4	9	0	0	0	pass	pass	pass	GATACCAGTTI
1A-0075270A-T4-17_896	40.8	0	16	1	0	1	6	7	0	fail	fail	fail	AAAAATGGAG
1A-0075270A-T4-17_960	42.5	0	18	3	0	2	8	4	0	fail	fail	fail	GGAGAAGGCC
1A-0075270A-T4-17_1024	37.5	0	21	6	1	5	8	0	0	fail	fail	fail	ATACAGTTCCP
1A-0075270A-T4-17_1088	36.7	0	29	4	1	6	10	0	0	fail	fail	fail	AAGAAGAAGC
1A-0075270A-T4-17_1152	40.8	0	23	5	0	5	8	4	0	fail	fail	fail	AGTGCTGCA
1A-0075270A-T4-17_1216	41.7	0	25	8	4	9	3	0	0	fail	fail	pass	GAGGTGGTTA
1A-0075270A-T4-17_1280	39.2	0	25	6	3	14	1	0	0	fail	fail	fail	AGGTGGCTAA
1A-0075270A-T4-17_1344	36.7	0	24	4	4	9	6	0	0	fail	fail	fail	TGAATTTCTTA
1A-0075270A-T4-17_1408	35.8	0	27	5	7	13	0	0	0	fail	fail	fail	AATGTCATTGC
1A-0075270A-T4-17_1472	36.7	0	29	8	6	13	1	0	0	fail	fail	fail	TCTTATGCAGC
1A-0075270A-T4-17_1536	39.2	0	21	1	7	12	0	0	0	fail	fail	fail	AAAGTTGATCT
1A-0075270A-T4-17_1580	47.5	0	3	0	0	0	1	1	0	fail	fail	fail	CAAGGGCACT



## T4-17 (TaEDR1) A homoeologue – 38.5% Mybait coverage



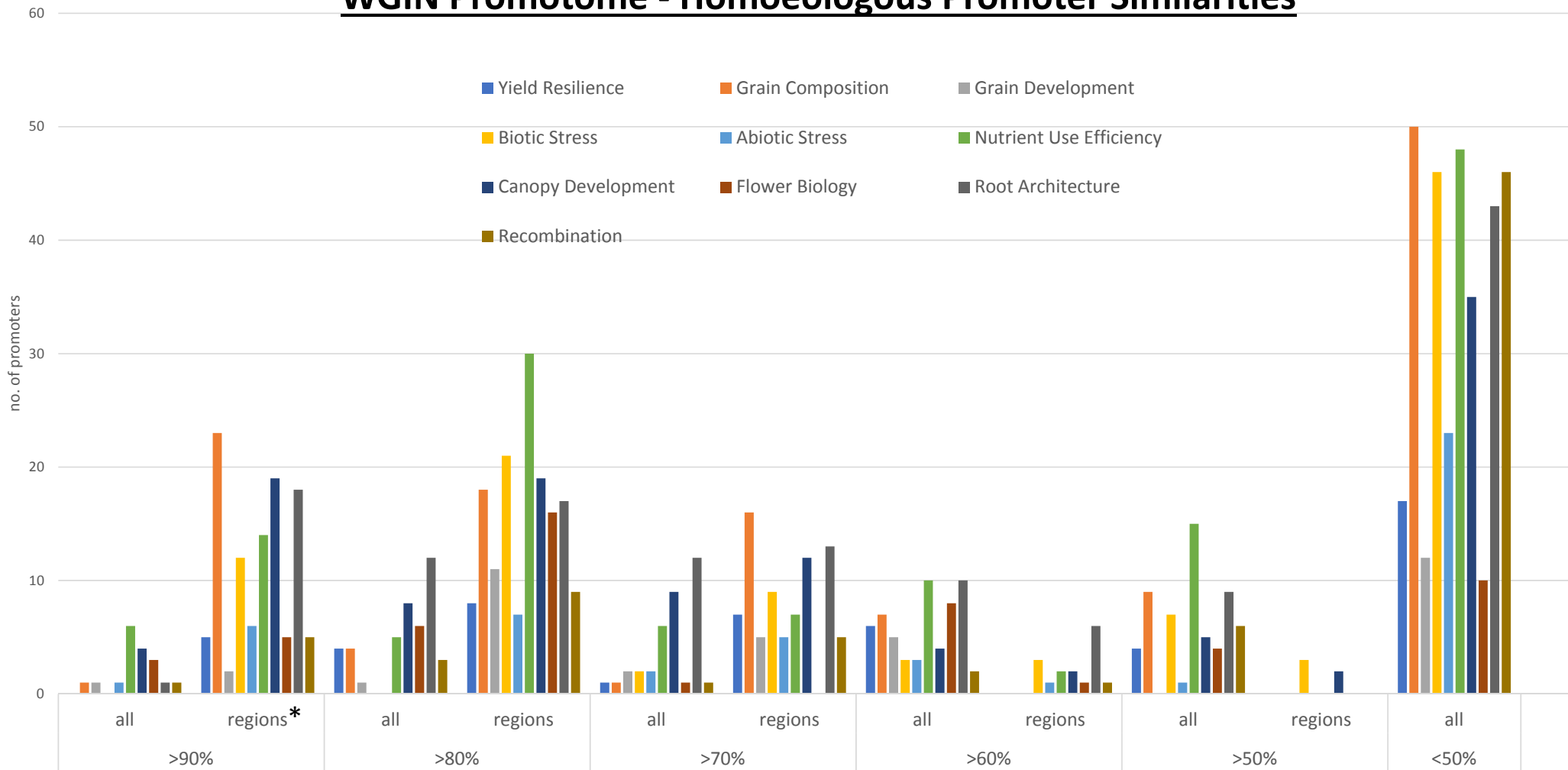
## WGIN Promotome MYbaits coverage



## Comparison of Homoeologous Promoter Sequences

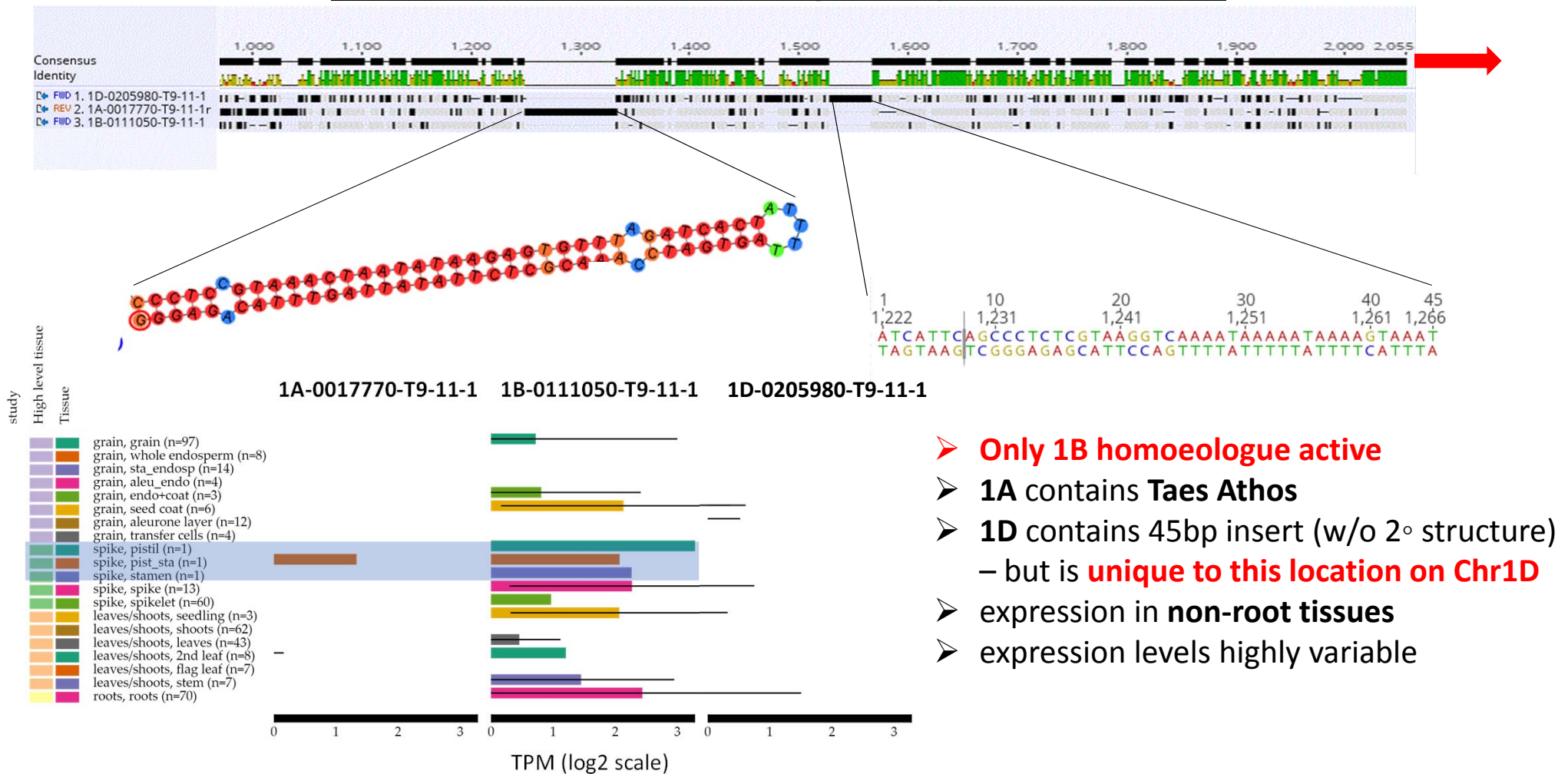
trait&gene no.	gene	WGIN Promotome ID	MYbaits	MYbaits coverage (%)	homoeologue ID (%)
T4-1	TaCERK1	7A-1826280-T4-1	18	74.4	29.8
T4-1	TaCERK1	7B-1934340-T4-1	11	51.3	
T4-1	TaCERK1	7D-2045100-T4-1	9	39.8	
T4-10	TaMPK3	4A-1008960-T4-10r	7	41.6	57.5 but 84.7 to [-800]
T4-10	TaMPK3	4B-1033300-T4-10	14	65.2	
T4-10	TaMPK3	4D-1148960-T4-10r	12	58.4	
T4-11	TaMPK6	7A-1834630-T4-11r	4	21.6	45.5 but 69.9 to [-800]
T4-11	TaMPK6	7B-1954300-T4-11r	5	25.4	
T4-11	TaMPK6	7D-2040040-T4-11r	6	32.5	
T4-12	TaOCP3	7A-1770880-T4-12	13	52.2	24.3
T4-12	TaOCP3	7B-1867300-T4-12r	18	77.6	
T4-12	TaOCP3	7D-2004100-T4-12r	22	91.5	
T4-13	TaOPR3	7A-1801960-T4-13	8	43.3	49.1
T4-13	TaOPR3	7B-1860490-T4-13	7	29.6	T4-13 = T5-14???
T4-13	TaOPR3	7D-2008410-T4-13	6	25.9	
T4-14	TaNH1	3A-0688550-T4-14	14	65.9	42
T4-14	TaNH1	3B-0744050-T4-14	10	43.1	
T4-14	TaNH1	3D-0914040-T4-14	12	58.4	
T4-15	TaBI-1	1D-0240360-T4-15	7	28.5	25.1
T4-15	TaBI-1	6A-1573990-T4-15	15	76.2	
T4-15	TaBI-1	6B-1671980-T4-15r	17	82.6	
T4-16	TaLSD1	7A-1798270-T4-16	6	36	46.5 but 86.9 to [-580] and 100% (1080-1176)
T4-16	TaLSD1	7B-1864900-T4-16	15	72.9	
T4-16	TaLSD1	7D-1974840-T4-16	14	71.1	

# WGIN Promotome - Homoeologous Promoter Similarities



\* regions  $\geq$  100bp

## T9-11-1 (Bonzaï3) Homoeologues Expression Patterns



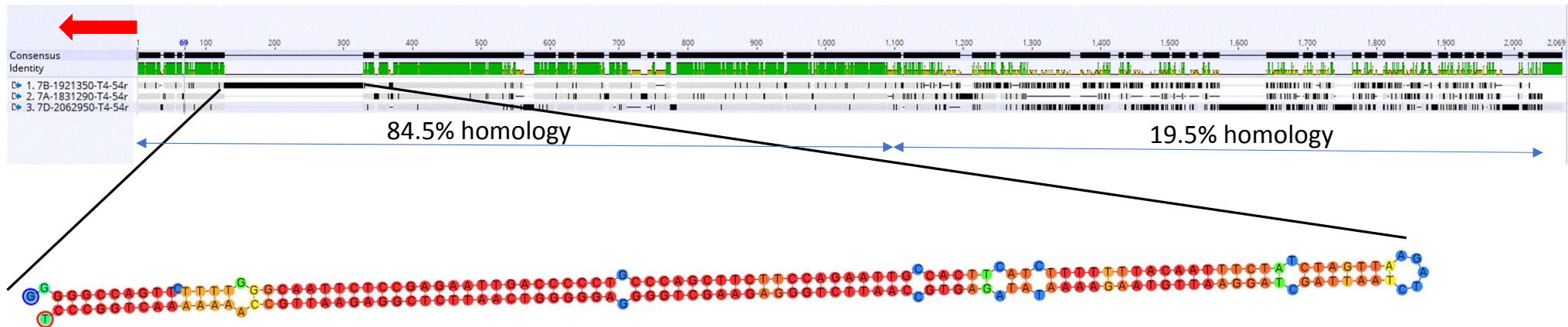
- **Only 1B homoeologue active**
- **1A contains Taes Athos**
- **1D contains 45bp insert (w/o 2° structure)**  
– but is **unique to this location on Chr1D**
- expression in **non-root tissues**
- expression levels highly variable

Expression graph modified from **Wheat Expression Browser** (Cristobal Uauy (JIC) <http://www.wheat-expression.com>)



## T4-54r (AK360626 – RING-H2 finger protein)

### B homoeologue insertion with stable secondary structure (palindrome)



- **Very stable 2° structure**
- **Possible MITE ?**
- **1,020 hits** in Wheat Genome
- **NO match** in TREP database & IWGSC\_refseqv1.0\_TransposableElements\_2017Mar13

**➔ New MITE?**

## Key Outcomes Prior to MYbaits Capture

**50 genes** were reassigned to different chromosomes (TGAC → IWGSC), in 32 cases completing the ABD Homoeologue sets

Very specific primers identified for **each homoeologue**

**71% of promoters** with >50% MYbaits coverage  
→ **full 1700bp sequences expected** for ALL of these

New **MITE** identified (let's call it **Taes WGIN**) – by aligning the homoeologous promoter/5'UTR sequences

Good evidence for **differential homoeologous expression** in Chinese Spring – vindicating the WGIN approach of capturing individual homoeologues

Chromosomes with highest densities for ALL Traits are 7D, 7A & 1D (243/1390)

## Galaxy Workflow for Sequence Assembly & SNP calling

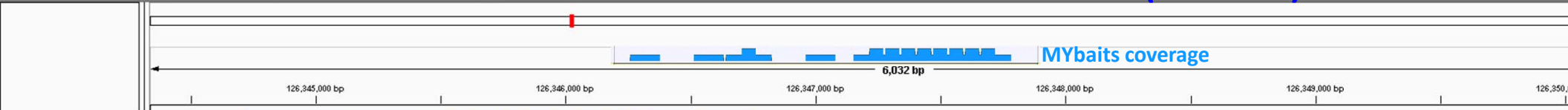


1. BWA-mem mapping
2. Filter for proper paired mapped reads
3. Left align for indels
4. Freebayes snp calling using min base quality of 10 and mapping quality of 10
5. Filter snp files using snpSift with min total coverage 10 and snp quality 30
6. vcf-merge to combine files to one file
7. vcf-stats for individual and combined file stat numbers, snpeff for combined file stats.

## Summary of Sequencing Data\* Obtained

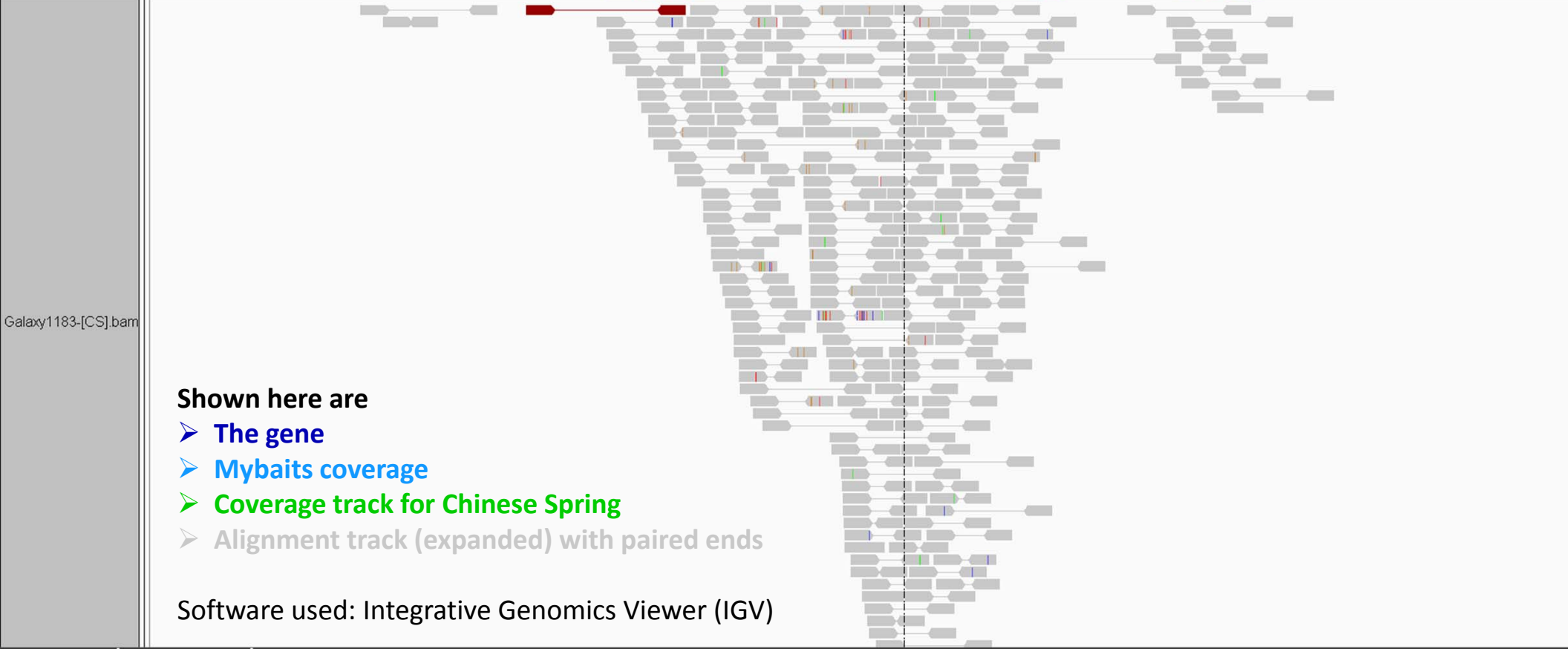
Chromosome	length (bp)	trait genes	sequence generated	95 cultivars	variants	1 SNP/InDel every xxx bp
1A	594,102,056	58	116,000	11,020,000	70,221	157
1B	689,851,870	58	116,000	11,020,000	30,590	360
1D	495,453,186	52	104,000	9,880,000	17,285	572
2A	780,798,557	64	128,000	12,160,000	102,178	119
2B	801,256,715	69	138,000	13,110,000	36,176	362
2D	651,852,609	72	144,000	13,680,000	24,012	570
3A	750,843,639	59	118,000	11,210,000	90,645	124
3B	830,829,764	61	122,000	11,590,000	40,422	287
3D	615,552,423	61	122,000	11,590,000	21,476	540
4A	744,588,157	56	112,000	10,640,000	83,862	127
4B	673,617,499	57	114,000	10,830,000	25,411	426
4D	509,857,067	50	100,000	9,500,000	16,159	588
5A	709,773,743	69	138,000	13,110,000	81,630	161
5B	713,149,757	64	128,000	12,160,000	40,261	302
5D	566,080,677	63	126,000	11,970,000	19,958	600
6A	618,079,260	48	96,000	9,120,000	87,516	104
6B	720,988,478	47	94,000	8,930,000	33,681	265
6D	473,592,718	45	90,000	8,550,000	16,034	533
7A	736,706,236	88	176,000	16,720,000	94,821	176
7B	750,620,385	75	150,000	14,250,000	35,872	397
7D	638,686,055	88	176,000	16,720,000	26,492	631
Un	480,980,714	13	26,000	2,470,000	16,250	152
<b>Total</b>	<b>14,547,261,565</b>	<b>1317</b>	<b>2,634,000</b>	<b>252,864,000</b>	<b>1,010,952</b>	

\* Paired End Reads with 100bp sequenced from each end of the captured DNA fragments (PE100)



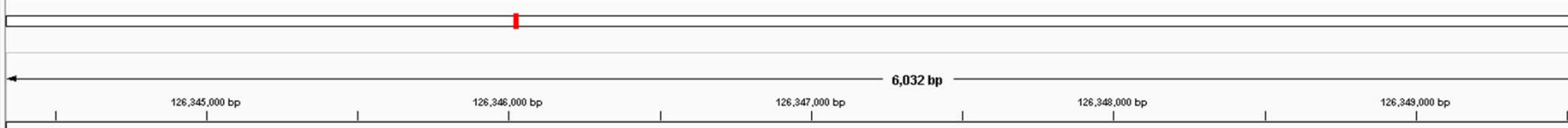
IWGS- HC Genes Chinese Spring

TraesCS6A02G146900

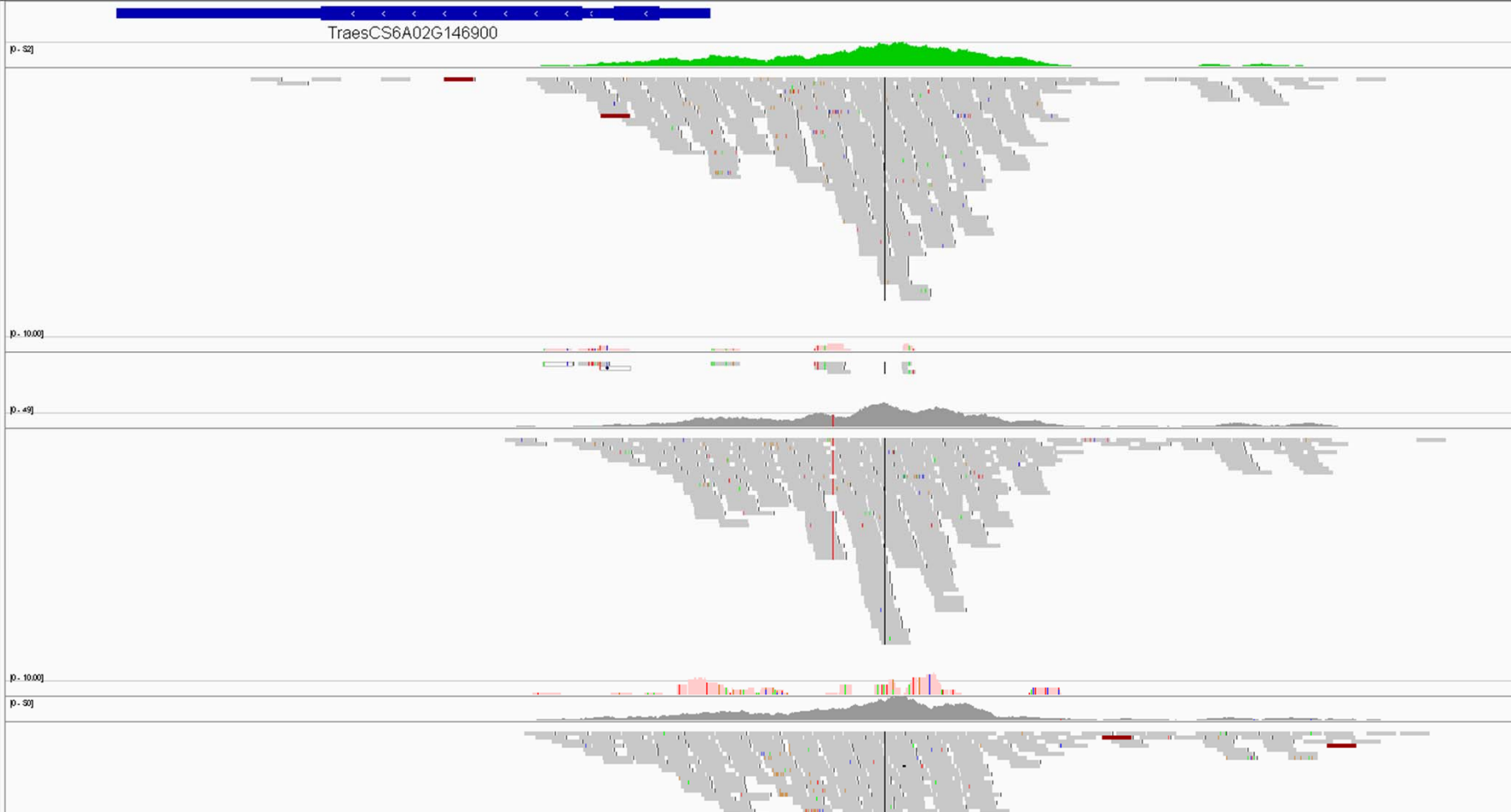


- Shown here are
- The gene
  - MYbaits coverage
  - Coverage track for Chinese Spring
  - Alignment track (expanded) with paired ends

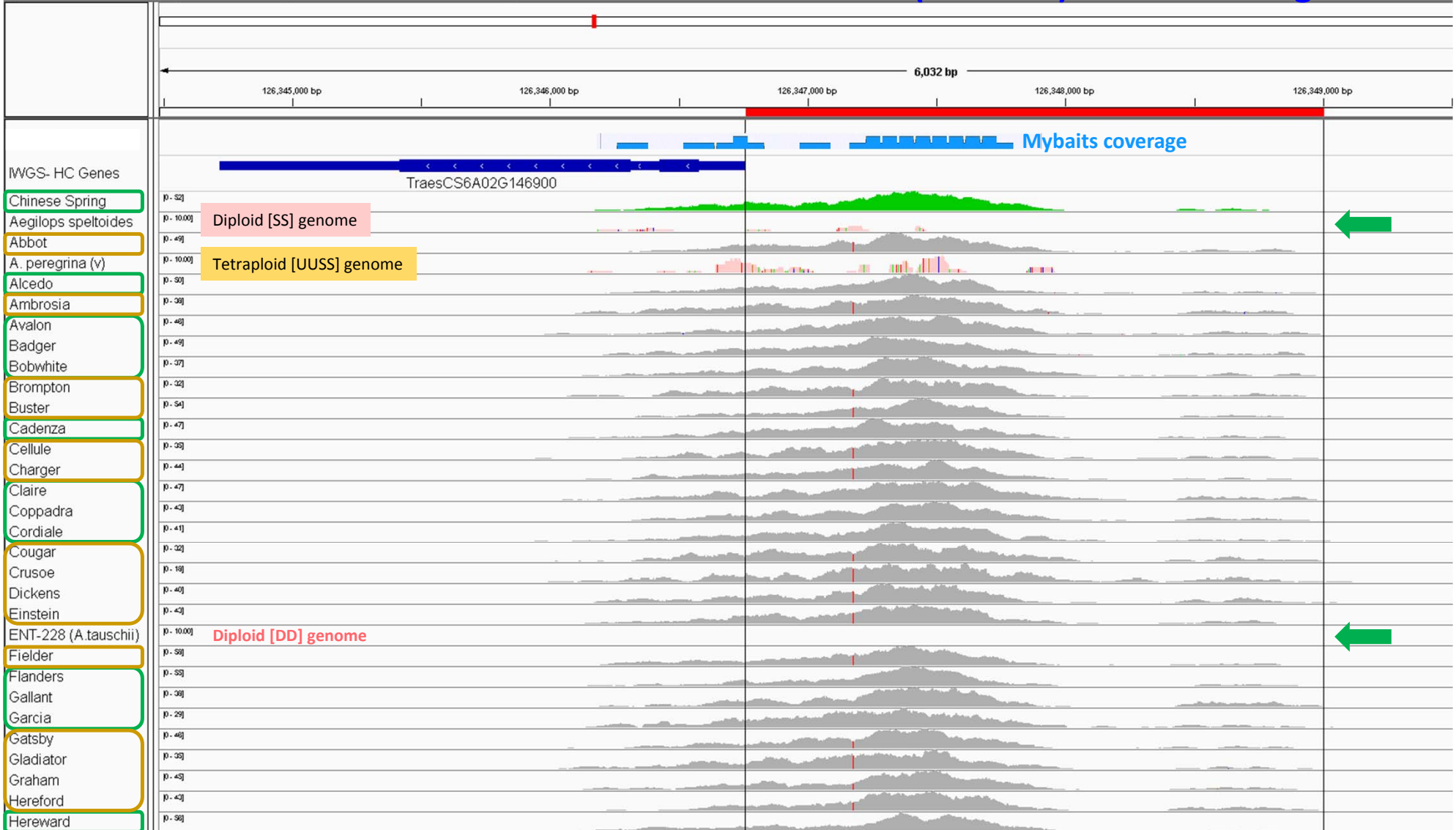
Software used: Integrative Genomics Viewer (IGV)



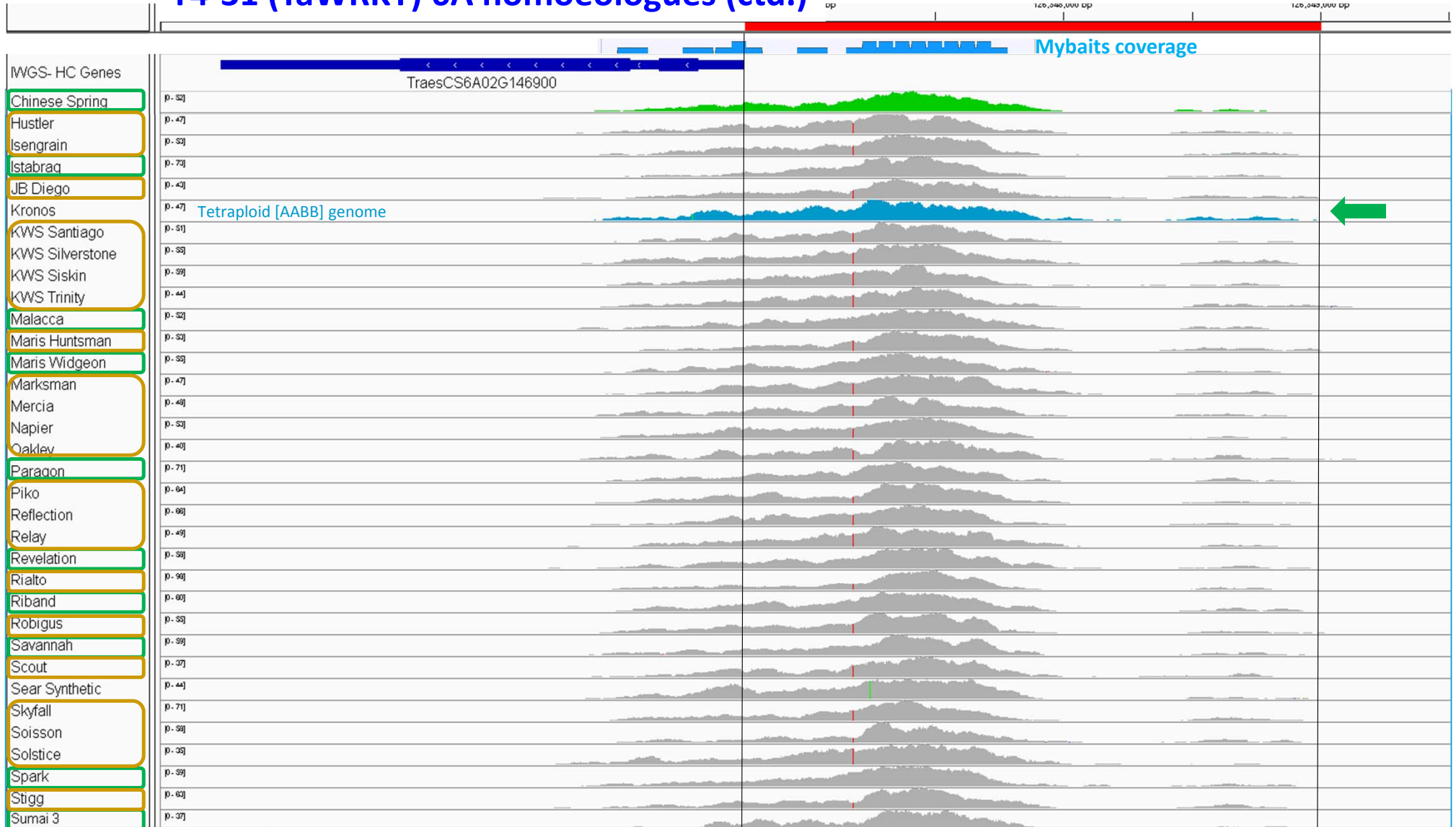
IWGS- HC Genes  
Chinese Spring  
Galaxy1183-[CS].bam  
Aegilops speltoides  
Galaxy1231-[Aspel].ba  
Abbot  
Galaxy1237-[AB].bam  
A. peregrina (v)  
Alcedo  
Galaxy1249-[AL].bam



# T4-31 (TaWRKY) 6A homoeologues

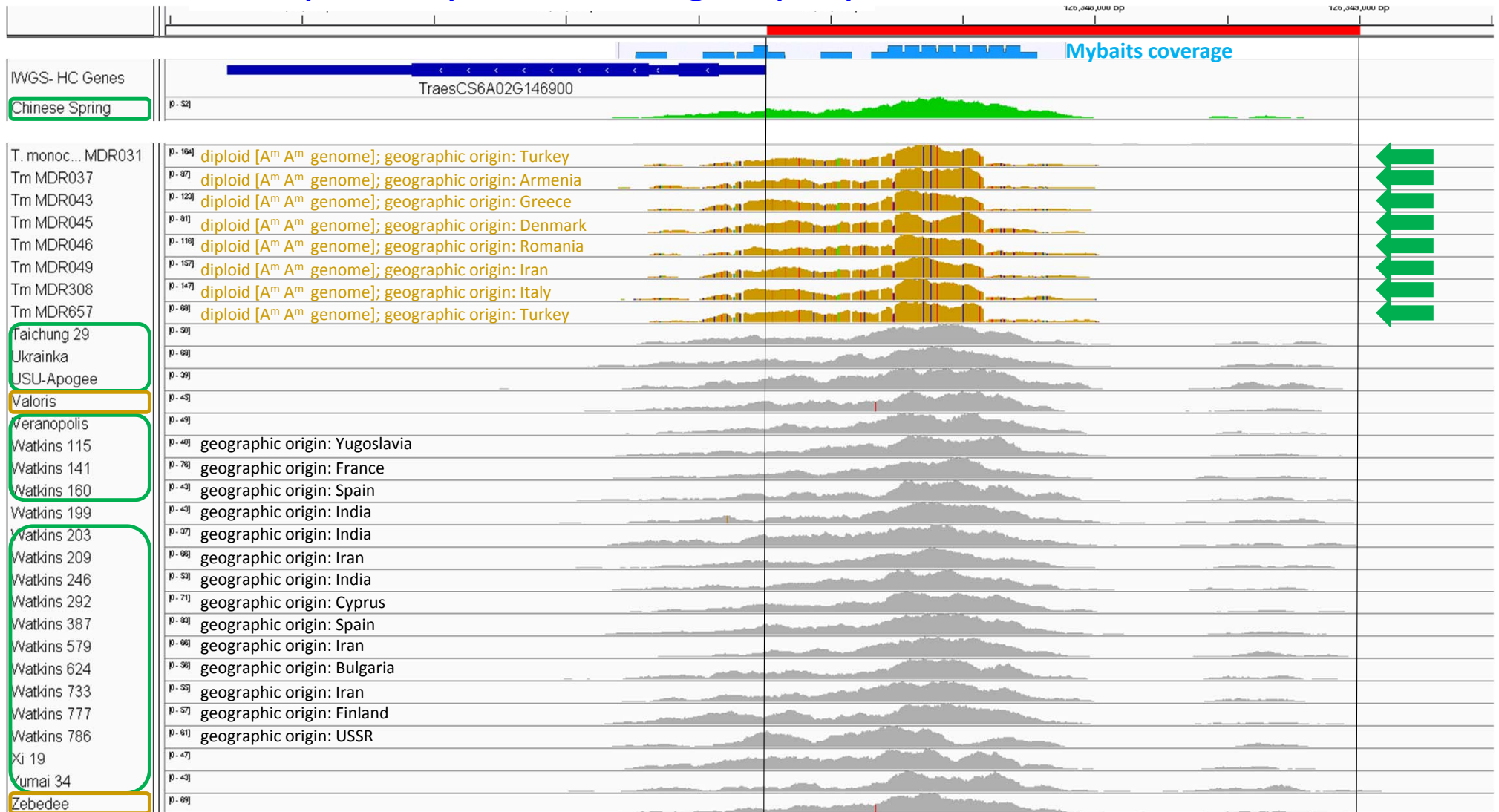


# T4-31 (TaWRKY) 6A homoeologues (ctd.)

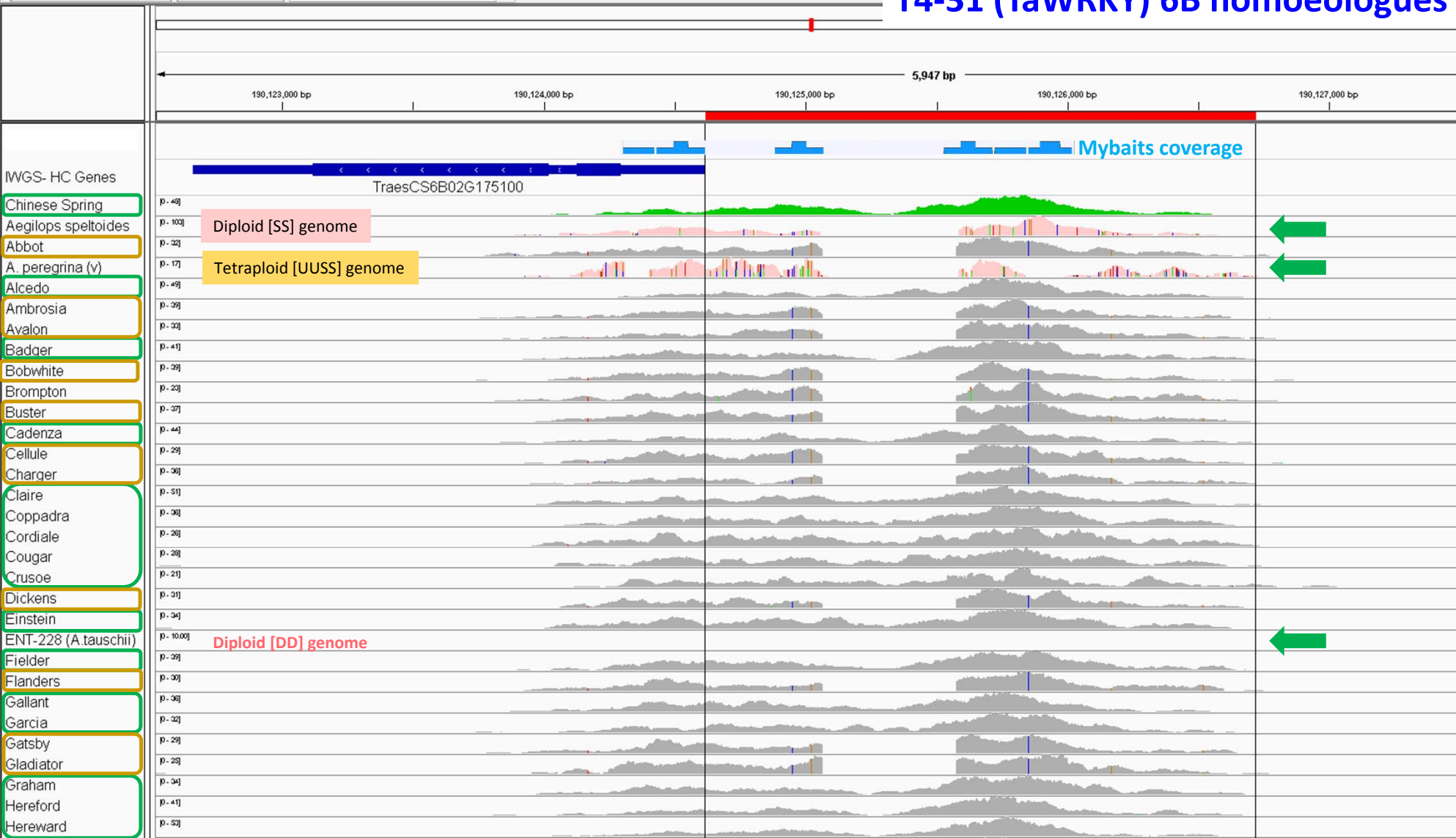




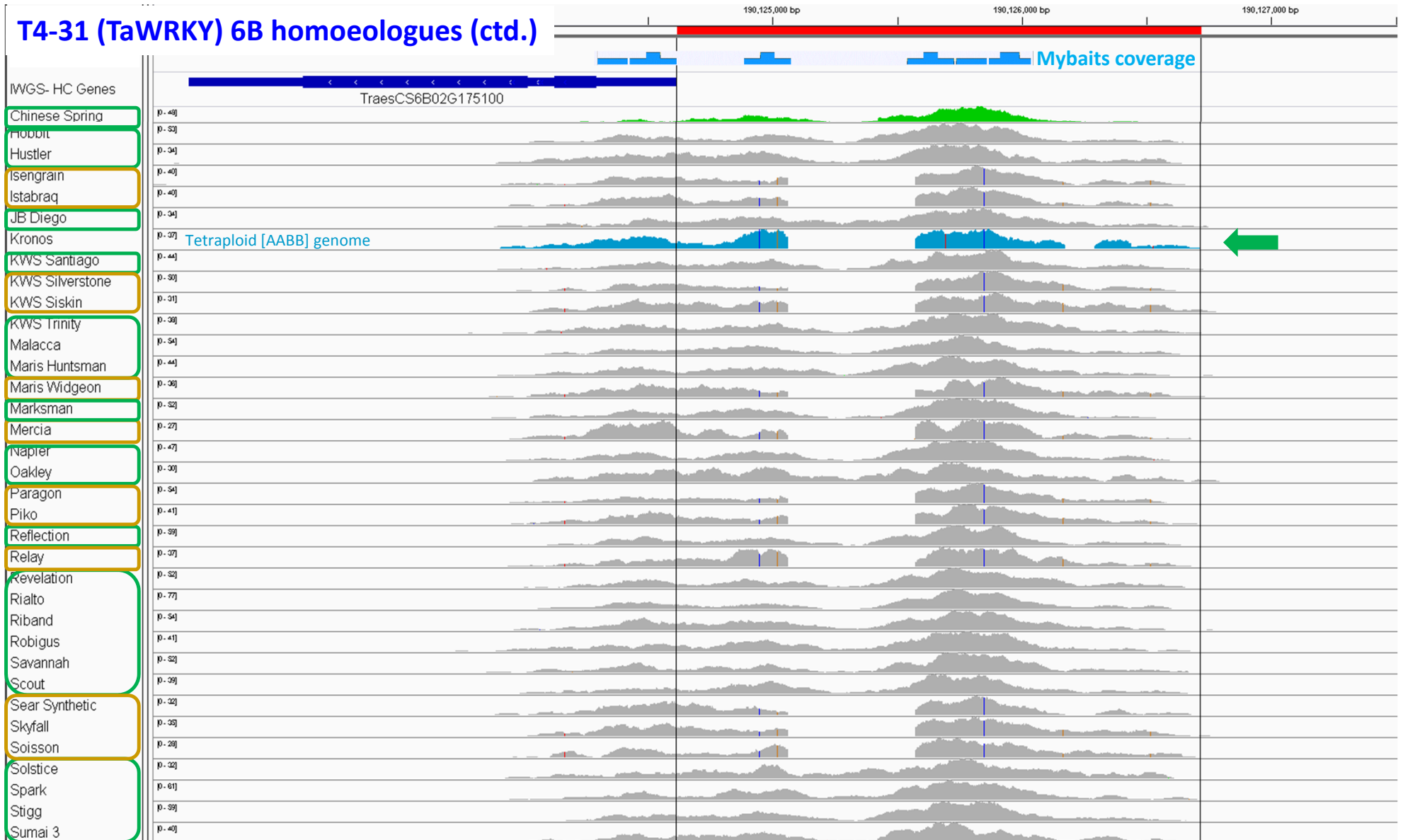
# T4-31 (TaWRKY) 6A homoeologues (ctd.)



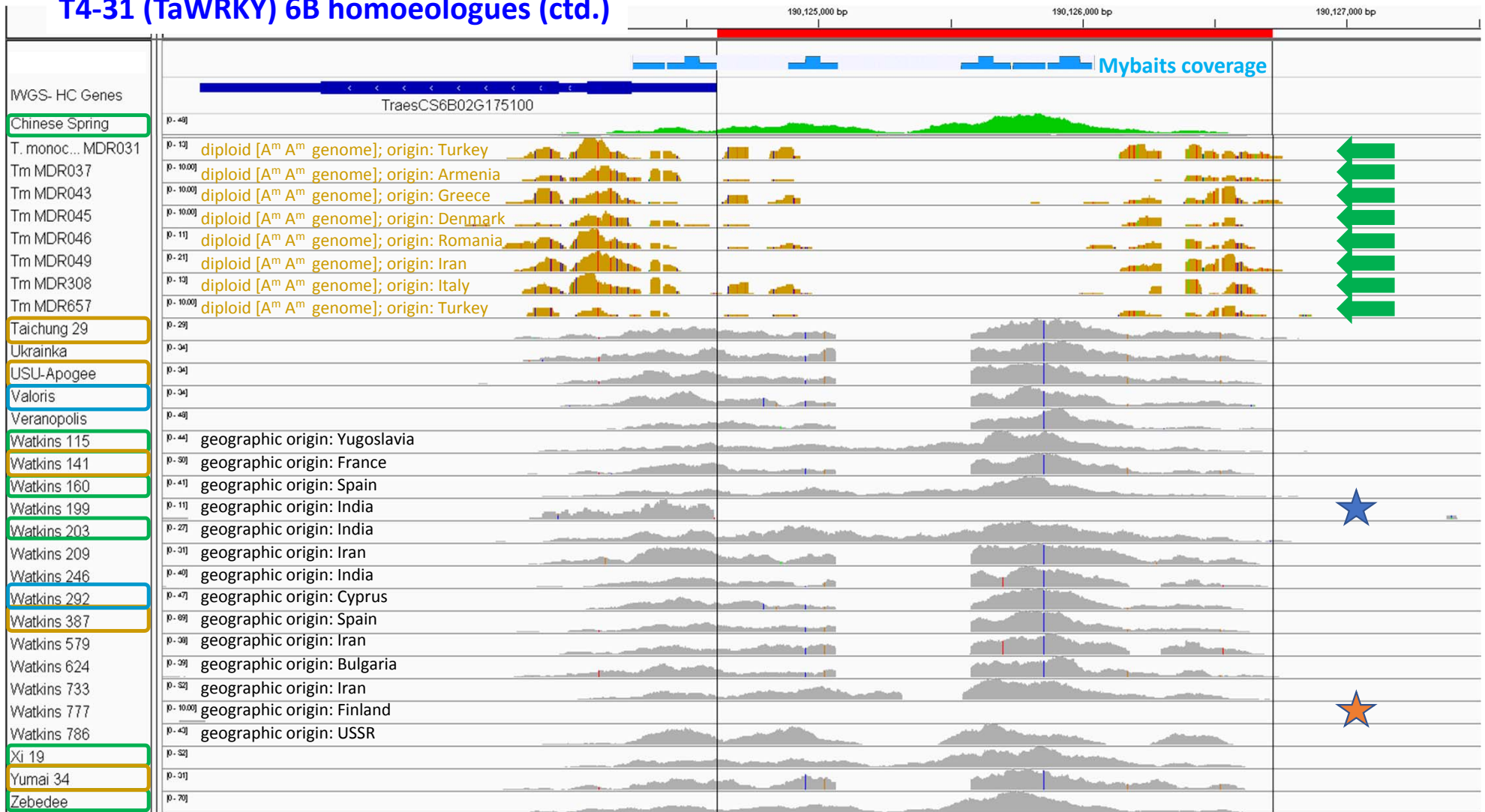
# T4-31 (TaWRKY) 6B homoeologues



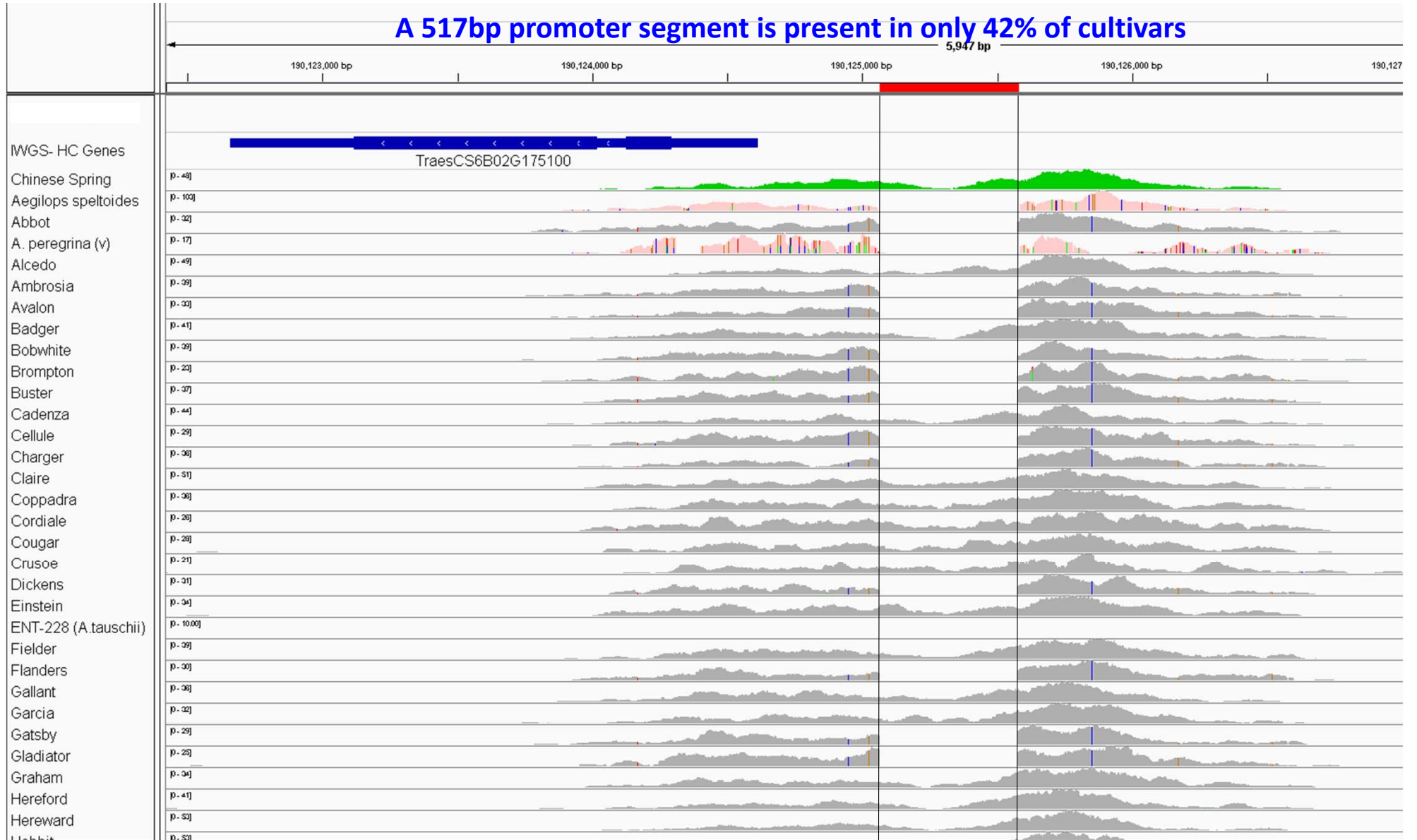
# T4-31 (TaWRKY) 6B homoeologues (ctd.)



# T4-31 (TaWRKY) 6B homoeologues (ctd.)



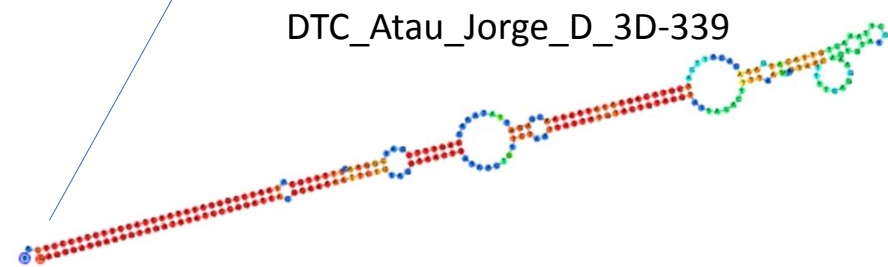
## A 517bp promoter segment is present in only 42% of cultivars



The sequence present in Chinese Spring and several other varieties in the T4-31(6B) promoter is a **chimeric sequence formed by two transposons** with DTC\_Atatau\_Jorge\_D\_3D-339 inserted inside DTH\_Taes/Tdur\_Coeus



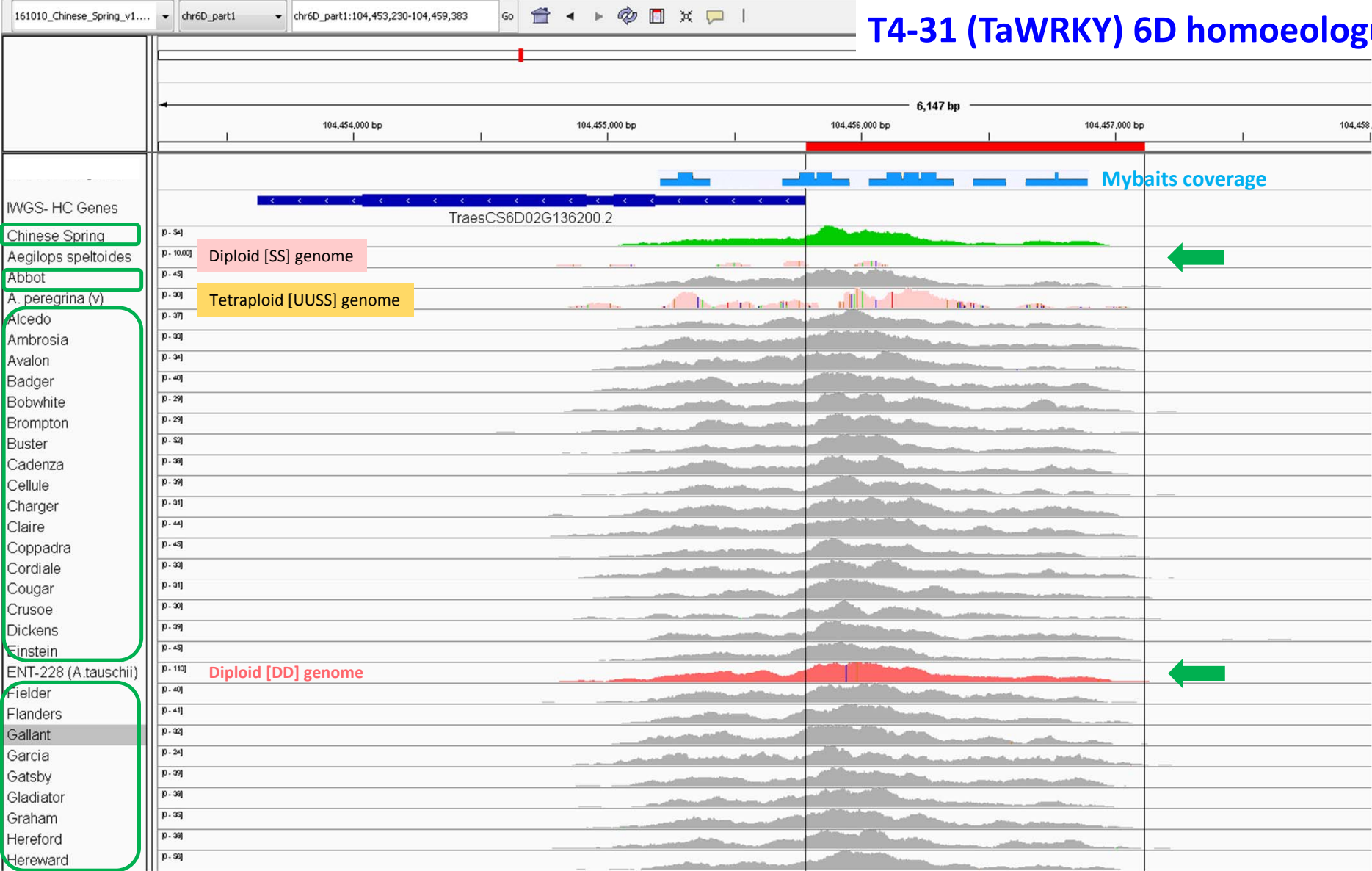
TP present	Varieties
Yes	CS, AL, BA, CA, CL, EI, FI, GL, GC, GA, GR, HE, HW, HU, JB, KSA, KTR, MA, MH, MK, NA, OA, RF, RV, RI, RB, RO, SA, SC, SO, SP, ST, SU, W115, W160, W203, W733*, W786, XI19, ZE
No	Aspel, AB, Apere, AM, AV, BW, BR, BU, CE, CH, DI, ENT**, FL, GT, GL, IS, IQ, KR, KSI, KSK, MW, ME, PA, PI, RL, SS, SF, SO, <b>ALL MDR***</b> , TA, UK, AP, VA, VE, W141, W199+, W209, W246, W292, W387, W579, W624, W777**, YU



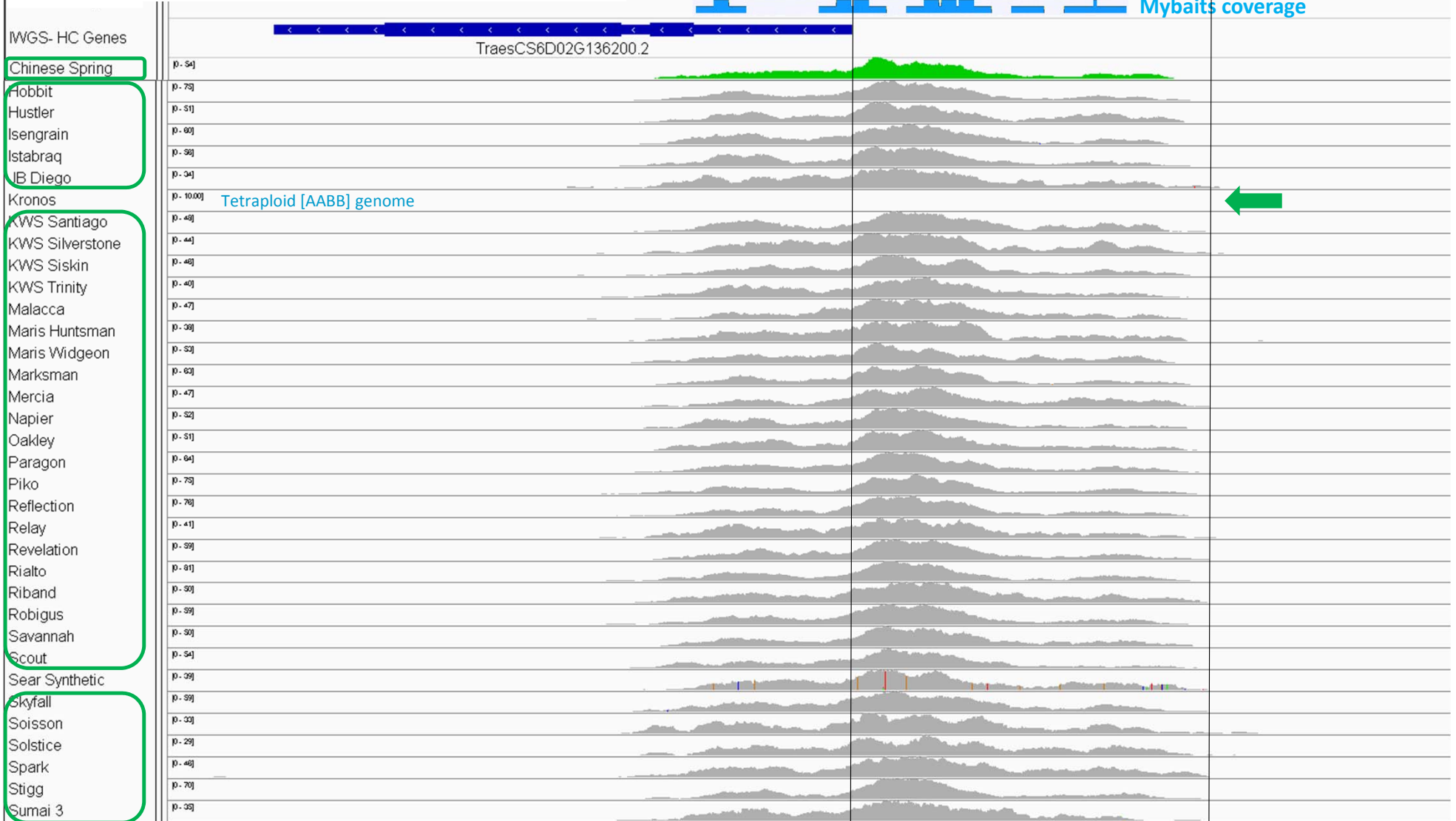
\*only DTH\_Taes/Tdur\_Coeus  
 \*\*\* very fragmented promoters

\*\* complete gene absent  
 + completely different promoter

# T4-31 (TaWRKY) 6D homoeologues

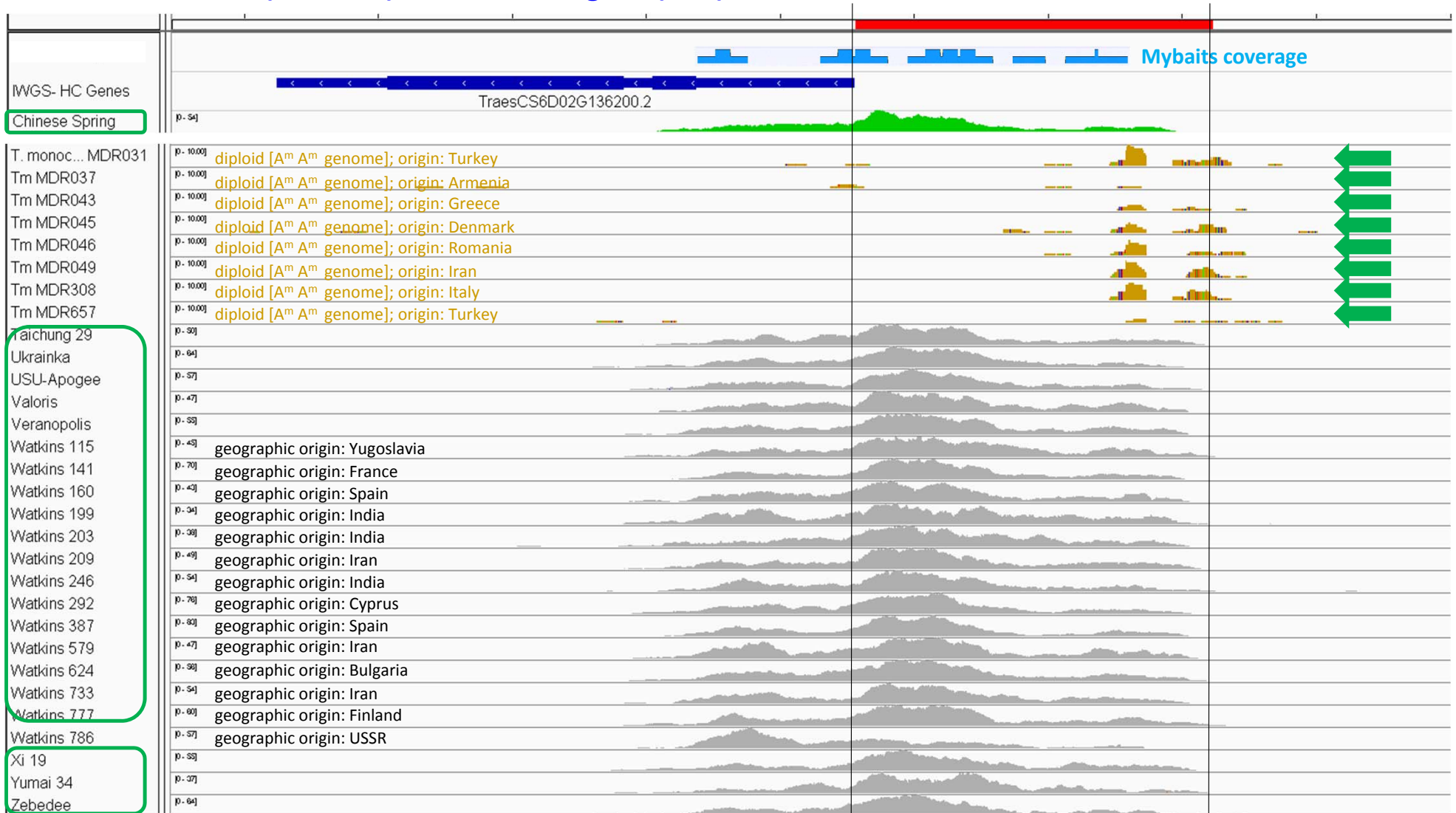


# T4-31 (TaWRKY) 6D homoeologues (ctd.)





## T4-31 (TaWRKY) 6D homoeologues (ctd.)



## Summary for T4-31 (TaWRKY)

homoeologue	Identical with Chinese Spring	Different with same SNPs/INDEL	Different with different SNPs
6A	41 varieties (49%)	40 varieties: only <b>1 SNP</b> (48%)	<ul style="list-style-type: none"> <li>• Sears Synthetic (1 SNP)</li> <li>• Watkins 199 (1 SNP)</li> </ul>
6B	41 varieties (49%)	28 varieties: 517bp INDEL & 4 SNPs (34%)	<ul style="list-style-type: none"> <li>• Valoris &amp; Watkins 292 : 2 add. SNPs (same INDEL)</li> <li>• Brompton, Kronos, Ukrainka, Verianopolis</li> <li>• Watkins W199, W209, W246, W579, W624, W733, W777, W786</li> </ul> <p>(17%)</p>
6D	82 varieties (99%)	NONE	Only Sears Synthetic w 12 SNPs

## WGIN Promotome Capture – initial conclusions

1. Technology has worked, using very stringent probe sets
2. Definitely possible to capture individual homoeologue promoters
3. Remarkable lack of diversity between Chinese Spring and other hexaploids (N=50)
4. Internal controls (non-hexaploid cultivars) have worked.
5. Evidence of at least two types of variation: a) SNPs frequently in conserved positions  
b) InDels: so far 2 transposon insertions in ONE promoter



## Acknowledgements

### ➤ Defra

- **The WGIN3 Management Team:** Andrew Riche (RRes), Clare Lister (JIC), David Feuerhelm (Syngenta), Dhan Bhandari (AHDB), Edward Flatman (Limagrain), 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), Peter Shewry (RRes), Richard Jennaway (saaten-union), Ruth Bryant (RAGT), Sarah Holdgate (NIAB), Simon Berry (Limagrain), Simon Griffiths (JIC), Simon Penson (Campden BRI), Stephen Smith (Elsoms), Vanessa McMillan (RRes)
- **Mycroarray (Arbor Biosciences):** Alison Devault, Jacob (Jake) Enk
- **Trait Co-ordinators & colleagues:** Kay Trafford (NIAB), Rowan Mitchell (RRes), Matthew Moscou (JIC), John Foulkes (Nottingham University), Peter Buchner (RRes), Nick Evens (RRes), James Connerton (JIC), Sigrid Heuer (RRes), Jose Fernandez (Nottingham University), Zoe Wilson (Nottingham University), Malcolm Bennett (Nottingham University), Laura Dixon (JIC), Scott Boden (JIC), James Higgins (Leicester University), Cristobal Uauy (JIC), Peter Jack (RAGT)
- **Provision of cultivars:** Mike Ambrose (GRU @ JIC), Simon Orford (JIC), Jacob Lage (KWS), Lesley Smart (RRes), Clare Lister (JIC), Nick Balaam (Senova Ltd), Kay Trafford (NIAB), Simon Berry (Limagrain)
- **Glasshouse etc:** Jill Maple (RRes), Jack Turner (RRes), Tom Yaxley (RRes)
- **Laboratory:** Carlos Bayon (RRes), Martin Urban (RRes)
- **Bioinformatics:** Rob King (RRes), Keywan Hassani-Pak (RRes)

