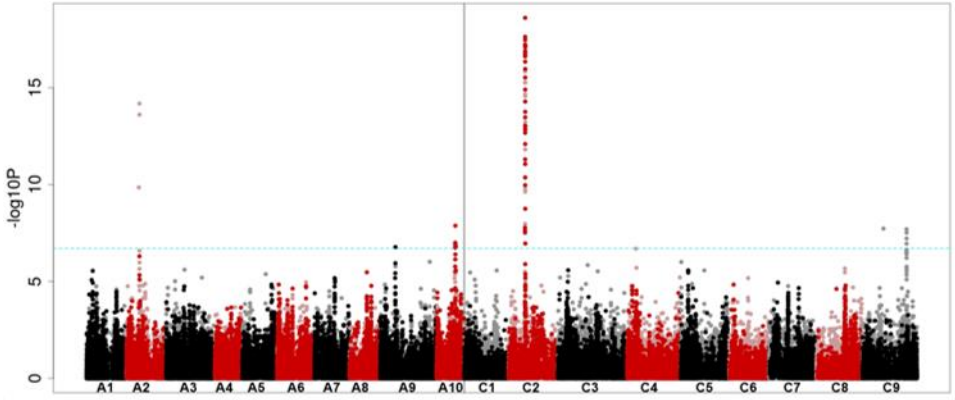
Introducing new genetic variation into crops

Many approaches are available for finding trait control genes via allelic variation

But what if the required variation doesn't exist in the available germplasm?

Novel variation traditionally introduced by:

- **Wide crossing (alien introgression)**
- **Mutagenesis (radiation, chemical)**



We can now use genomics to accelerate breeding based on knowledge of trait control genes

Genomics-led "traditional" approaches to crop genome engineering

Ian Bancroft

22nd May, 2019

- **Introduction to rapeseed genomics**
- **Genome structural variation by homoeologous exchange**
- **Genome structural variation by alien introgression**
- **Genome sequence and structural variation by gamma irradiation**

Rapeseed: a crop type of *Brassica napus*

Primary product is seed oil (~ €750/t) with feed meal co-product (~ €200/t)

Rapeseed is the third largest source of vegetable oil globally

EU produces ~10 million tonnes of rapeseed oil annually

Opportunities for enhancing further co-products, e.g. phytosterols, antioxidants

Breeding targets: tolerance to pests and diseases, nitrogen use efficiency, oil yield and fatty acid composition

Rapeseed: a crop type of *Brassica napus*

Rapeseed

Wheat

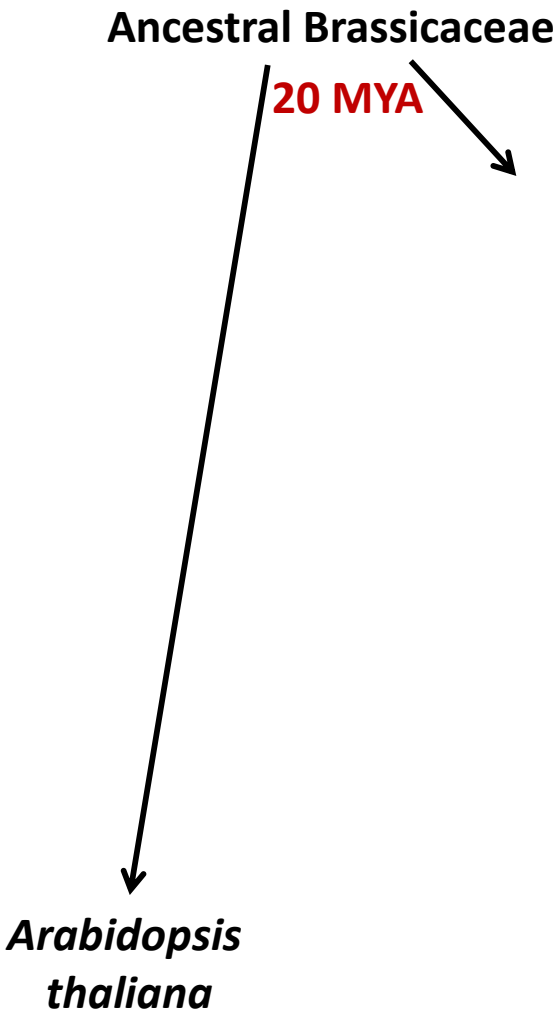
Potatoes

Crop rotation is essential for the control of pests, diseases and weeds

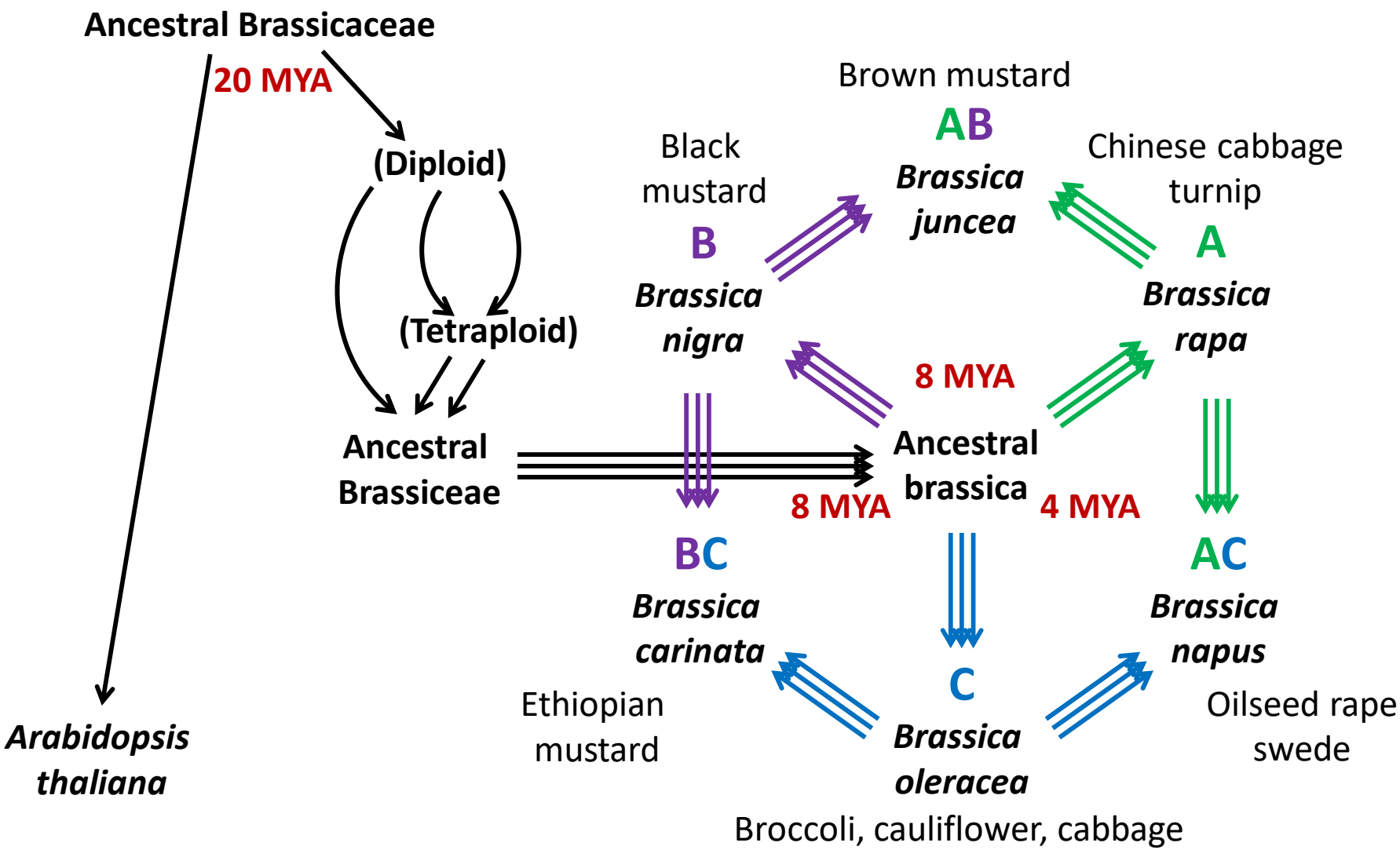
Field beans

Barley

Genome complexity in the cultivated *Brassica* species

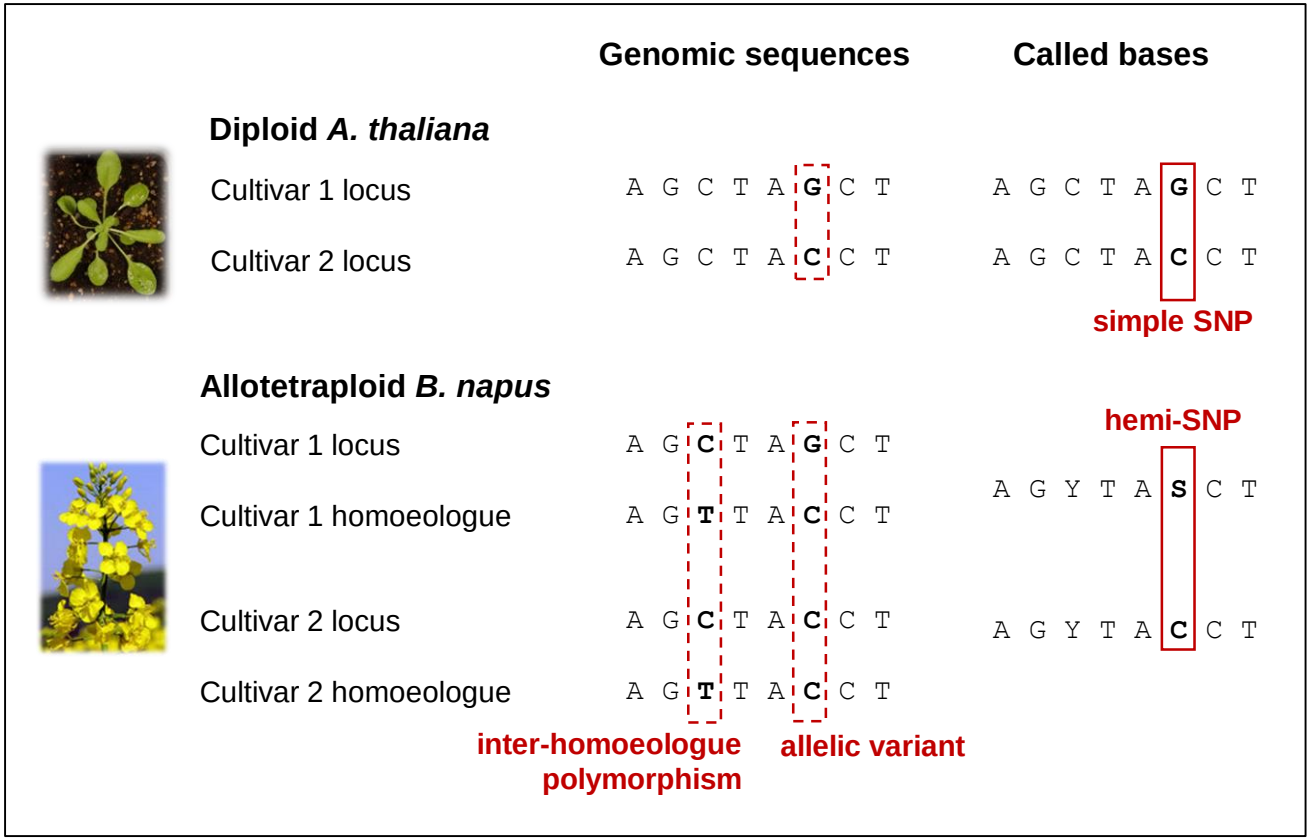


Genome complexity in the cultivated *Brassica* species



Genome complexity in polyploids complicates genotyping

Sequence polymorphisms between homoeologues (or paralogues) are much more abundant than allelic variation and can easily be misinterpreted, especially if sampling depth is insufficient to overcome stochastic effects.



Plant Biotechnology Journal

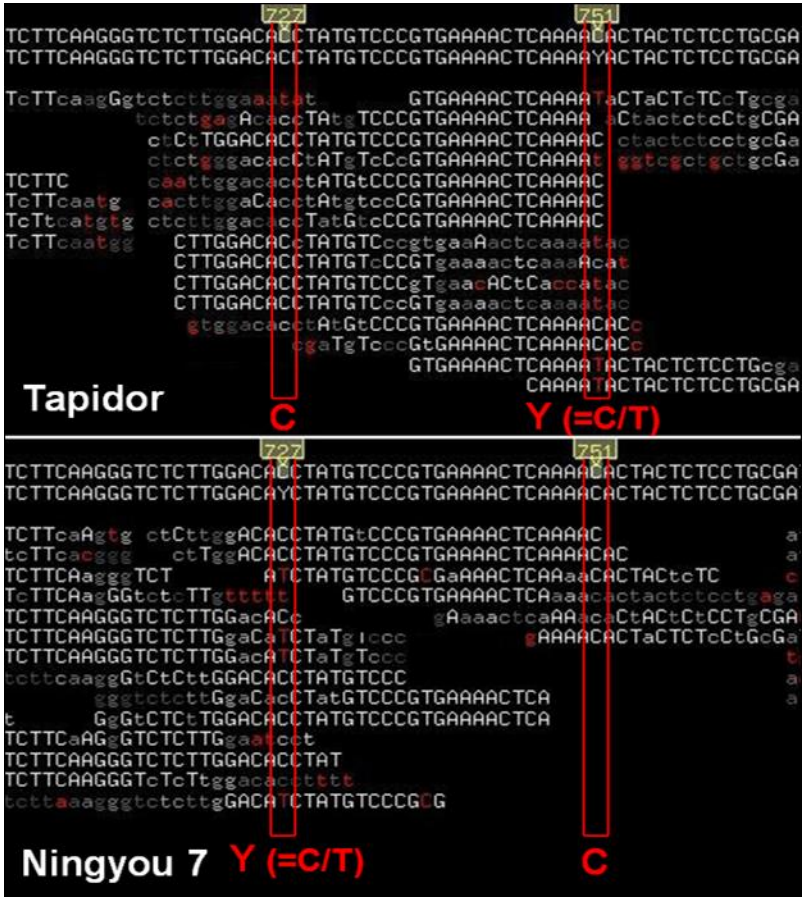
Trick, et al. SNP discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing Plant Biotechnology J. 7:334-346, 2009

“Functional genotypes” from mRNAseq data

mRNA →

3,000,000,000 bases of sequence per plant (~one human genome)

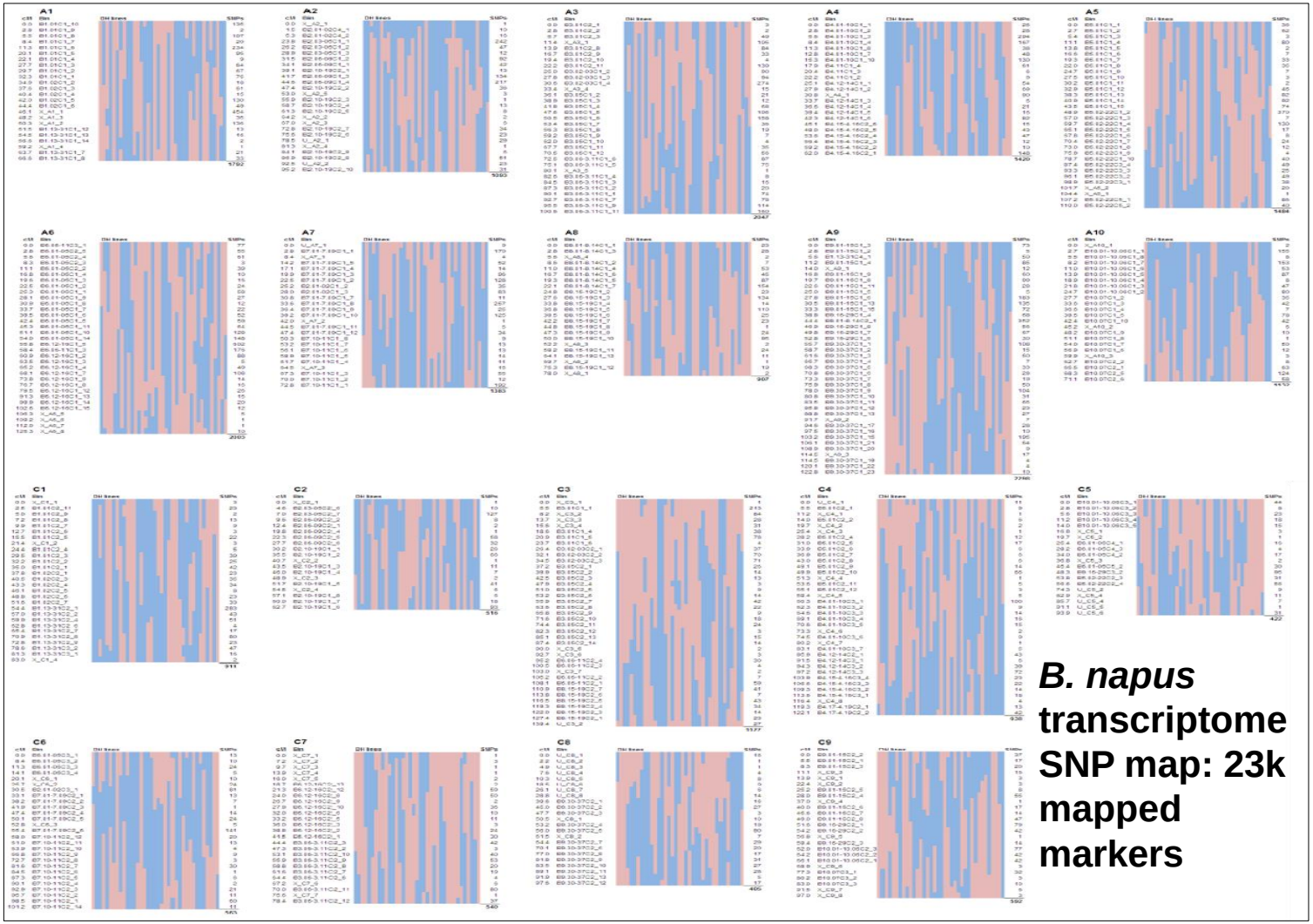
CDS or unigene reference sequence



Plant Biotechnology Journal

Trick, et al. SNP discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing Plant Biotechnology J. 7:334-346, 2009

Transcriptome SNPs as molecular markers in *B. napus*

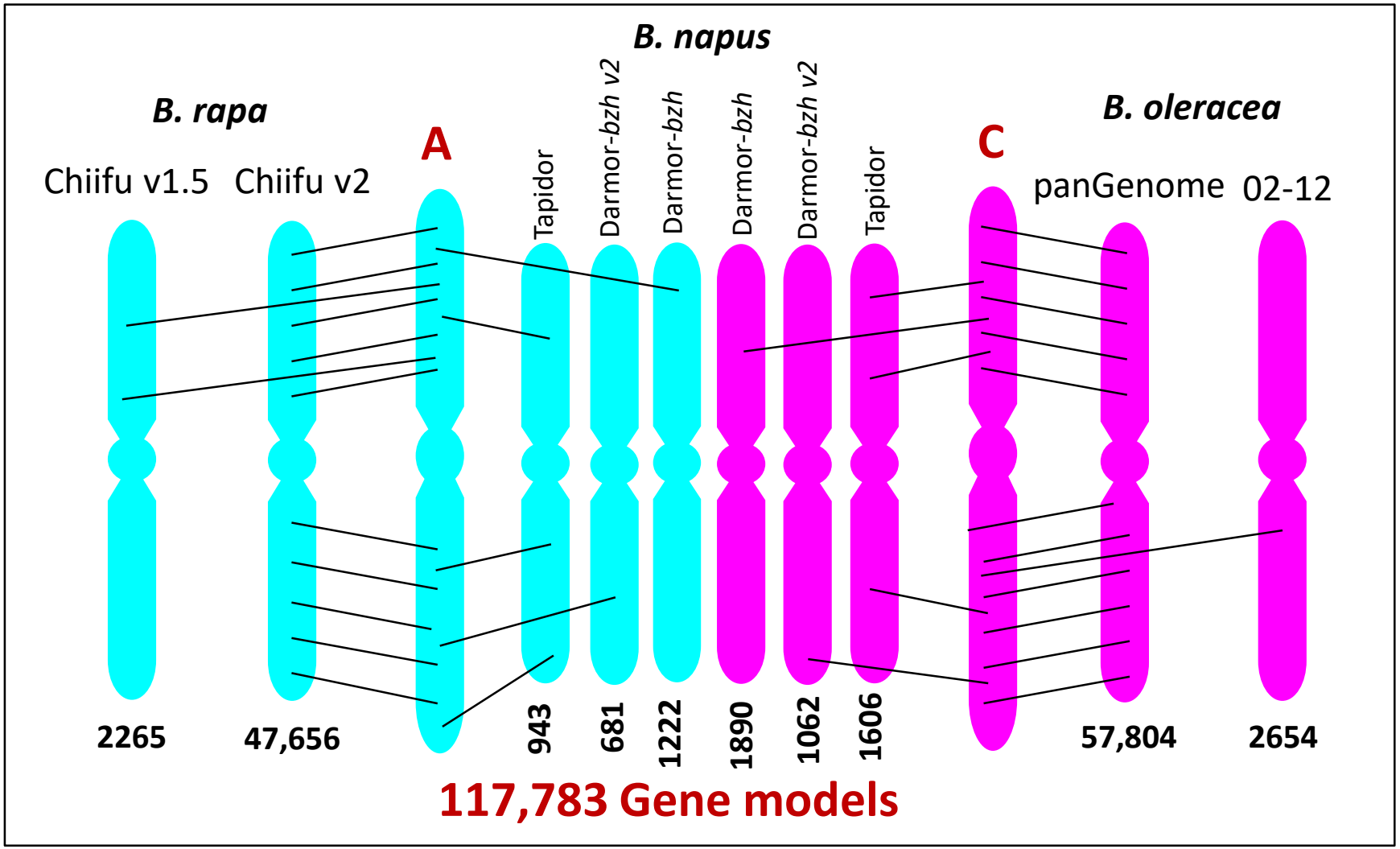


B. napus **transcriptome SNP map: 23k mapped markers**

nature biotechnology

Bancroft, et al. Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing Nature Biotechnology 29:762-6, 2011

AC pan-transcriptome reference sequence for *B. napus*

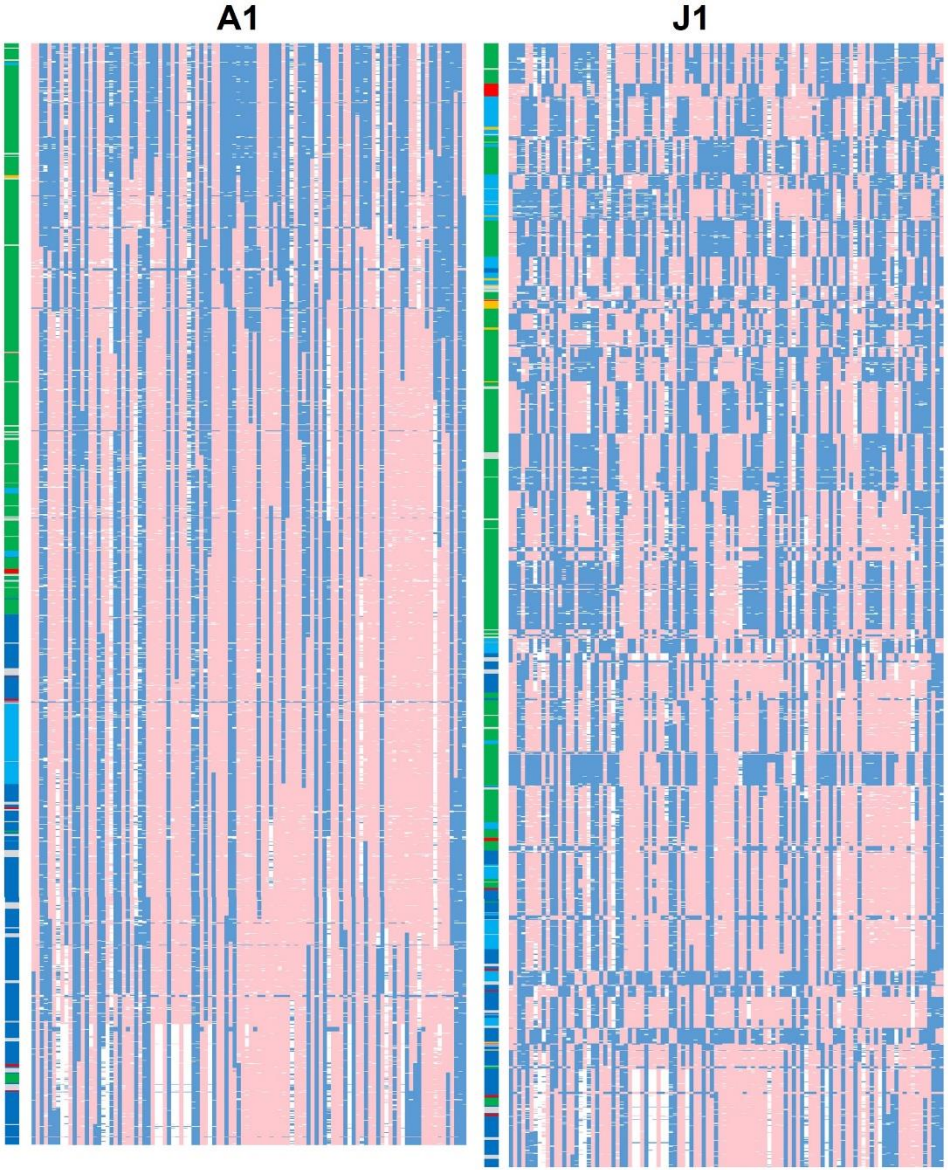


He, et al. Construction of *Brassica* A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research. *Data in brief* 4: 357-362, 2015.

Graphical genotypes displayed by gene order in a genome sequences provides a quality control for genome assemblies

Chimaeric and mis-mapped scaffolds are common in genome sequence assemblies

Genome Ordered Graphical Genotypes (GOGGs) help identify errors for correction



nature genetics

He and Bancroft. Organisation of the genome sequence of the polyploid crop species *Brassica juncea* Nature Genetics 50:1496-1498, 2018

Identifying homoeologous gene pairs for *B. napus*

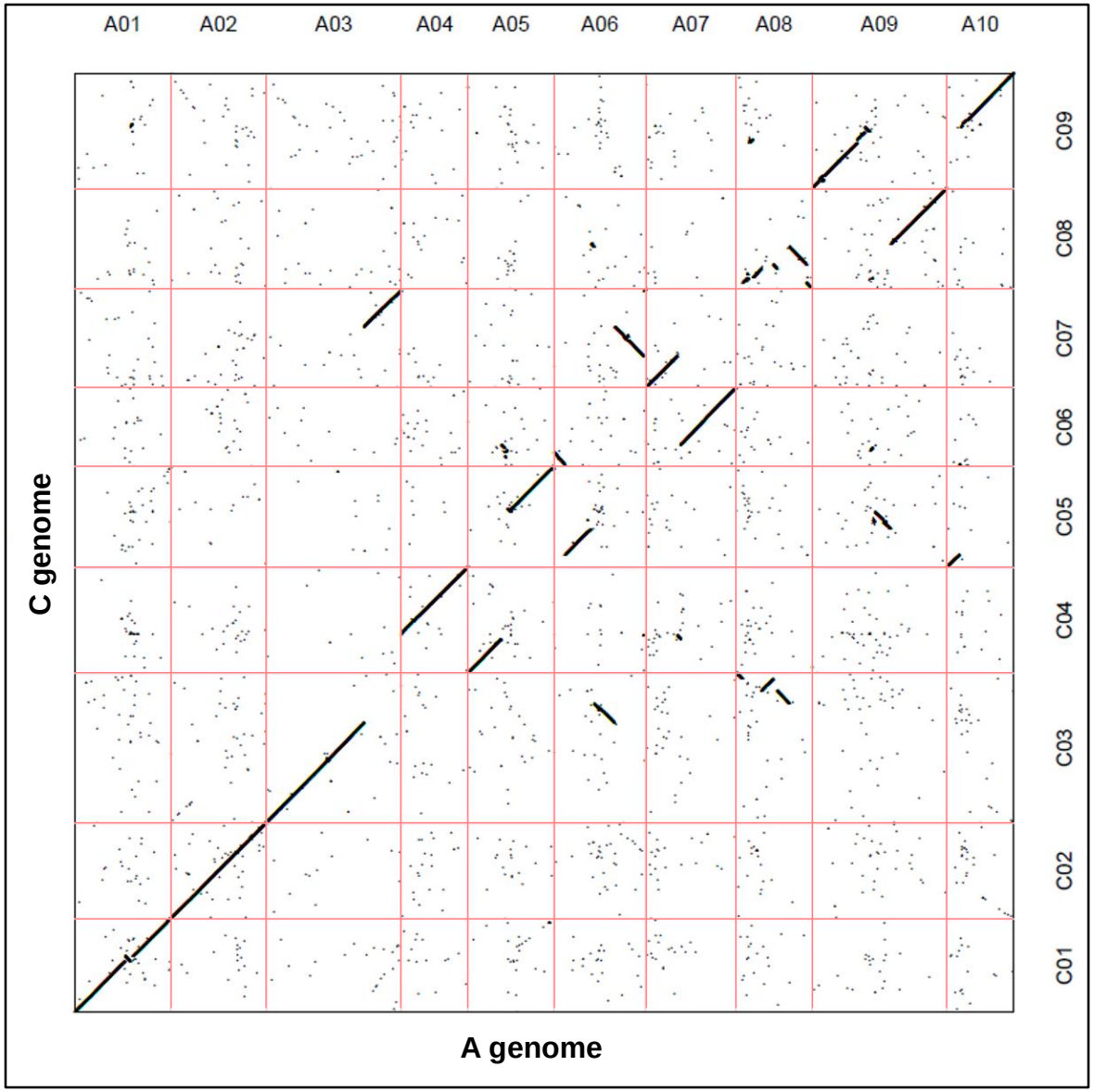
Identify pairs of homoeologues between *Brassica* A and C diploid genomes as reciprocal top BLAST hits between CDS gene models

A ⟷ C

35,618 reciprocal top BLAST hit pairs

1	A genome homoeologue	Position in A genome	Order in A	Order in C	Position in C genome	C genome homoeologue
2	Cab020597.1	A01_000007596_000004363	1	1	C01_000004096_000006192	Bo1g001030.1
3	Cab020596.1	A01_000009147_000008191	2	25377	C07_048354208_048355539	Bo7g120870.1
4	Cab020595.2	A01_000011403_000009778	3	2	C01_000007519_000009142	Bo1g001040.1
5	Cab020594.1	A01_000012765_000011917	4	3	C01_000009720_000010577	Bo1g001050.1
6	Cab020592.1	A01_000021944_000019956	5	5	C01_000037620_000038611	Bo1g001080.1
7	Cab020591.1	A01_000024271_000022281	6	6	C01_000039858_000040949	Bo1g001090.1
8	Cab020590.1	A01_000032446_000028836	7	7	C01_000046852_000050484	Bo1g001100.1
9	Cab020589.1	A01_000033293_000032595	8	30597	C09_021661539_021662053	Bo9g071600.1
10	Cab020588.1	A01_000035983_000033952	9	9	C01_000068965_000070059	Bo1g001140.1
11	Cab020581.1	A01_000063572_000063039	10	10	C01_000103201_000103782	Bo1g002150.1
12	Cab020579.1	A01_000079019_000078723	11	11	C01_000112982_000113278	Bo1g002190.1
13	Cab020577.1	A01_000088358_000088041	12	12	C01_000124351_000124668	Bo1g002200.1
14	Cab020576.1	A01_000094861_000093230	13	13	C01_000128563_000130200	Bo1g002210.1
15	Cab020575.1	A01_000097996_000095110	14	14	C01_000130440_000133386	Bo1g002220.1
16	Cab020574.1	A01_000100108_000098057	15	15	C01_000134318_000135251	Bo1g002230.1
17	Cab020573.1	A01_000105326_000101322	16	16	C01_000137577_000141372	Bo1g002250.1
18	Cab020571.1	A01_000115025_000112704	17	17	C01_000148961_000151258	Bo1g002280.1
19	Cab020570.1	A01_000119528_000116614	18	18	C01_000153250_000156261	Bo1g002300.1
20	Cab020569.1	A01_000129775_000126831	19	19	C01_000164958_000168351	Bo1g002310.1
21	Cab020568.1	A01_000135472_000131405	20	20	C01_000172779_000174206	Bo1g002330.1
22	Cab020567.1	A01_000143020_000139039	21	21	C01_000175196_000180272	Bo1g002340.1
23	BnaA01g05330D	A01_000146772_000145212.00	22	23	C01_000183323_000183774	Bo1g002360.1
24	BnaA01g05340D	A01_000146772_000145212.00	23	24	C01_000184847_000187483	Bo1g002370.1
25	Cab020565.1	A01_000146829_000145011	24	22	C01_000181177_000182520	Bo1g002350.1
26	Cab020563.1	A01_000155768_000155126	25	25	C01_000191317_000191852	Bo1g002400.1
27	Cab020562.1	A01_000158109_000156564	26	26	C01_000192637_000194204	Bo1g002410.1
28	Cab020560.1	A01_000162233_000159755	27	27	C01_000195435_000197330	Bo1g002420.1
29	Cab020558.1	A01_000164388_000163736	28	29	C01_000202463_000202892	Bo1g002450.1
30	Cab020557.1	A01_000167670_000166905	29	30	C01_000205351_000206116	Bo1g002460.1
31	Cab020556.1	A01_000168725_000168246	30	31	C01_000206881_000208172	Bo1g002470.1
32	Cab020555.1	A01_000170941_000170440	31	32	C01_000209076_000210429	Bo1g002480.1
33	Cab020554.1	A01_000171955_000171563	32	33	C01_000210647_000211042	Bo1g002490.1
34	Cab020553.1	A01_000180461_000175822	33	34	C01_000216852_000221622	Bo1g002500.1
35	Cab020552.1	A01_000184155_000183811	34	35	C01_000223558_000223902	Bo1g002510.1
36	Cab020550.1	A01_000190655_000188381	35	37	C01_000228012_000229293	Bo1g002520.1
37	Cab020549.1	A01_000199980_000195942	36	38	C01_000237167_000240760	Bo1g002530.1
38	Cab020547.1	A01_000210518_000207773	37	39	C01_000268432_000270656	Bo1g002540.1
39	Cab020546.2	A01_000233705_000220053	38	40	C01_000278884_000293749	Bo1g002550.1
40	Cab020545.1	A01_000236573_000235062	39	41	C01_000294476_000298099	Bo1g002560.1
41	Cab020544.1	A01_000240300_000237136	40	42	C01_000298888_000301646	Bo1g002570.1
42	Cab020543.1	A01_000244118_000242423	41	43	C01_000305351_000307099	Bo1g002590.1
43	Cab020540.1	A01_000247210_000246059	42	44	C01_000312999_000314150	Bo1g002620.1
44	Cab020539.1	A01_000249298_000248129	43	45	C01_000315809_000316990	Bo1g002630.1

Homoeology relationships between A and C genomes



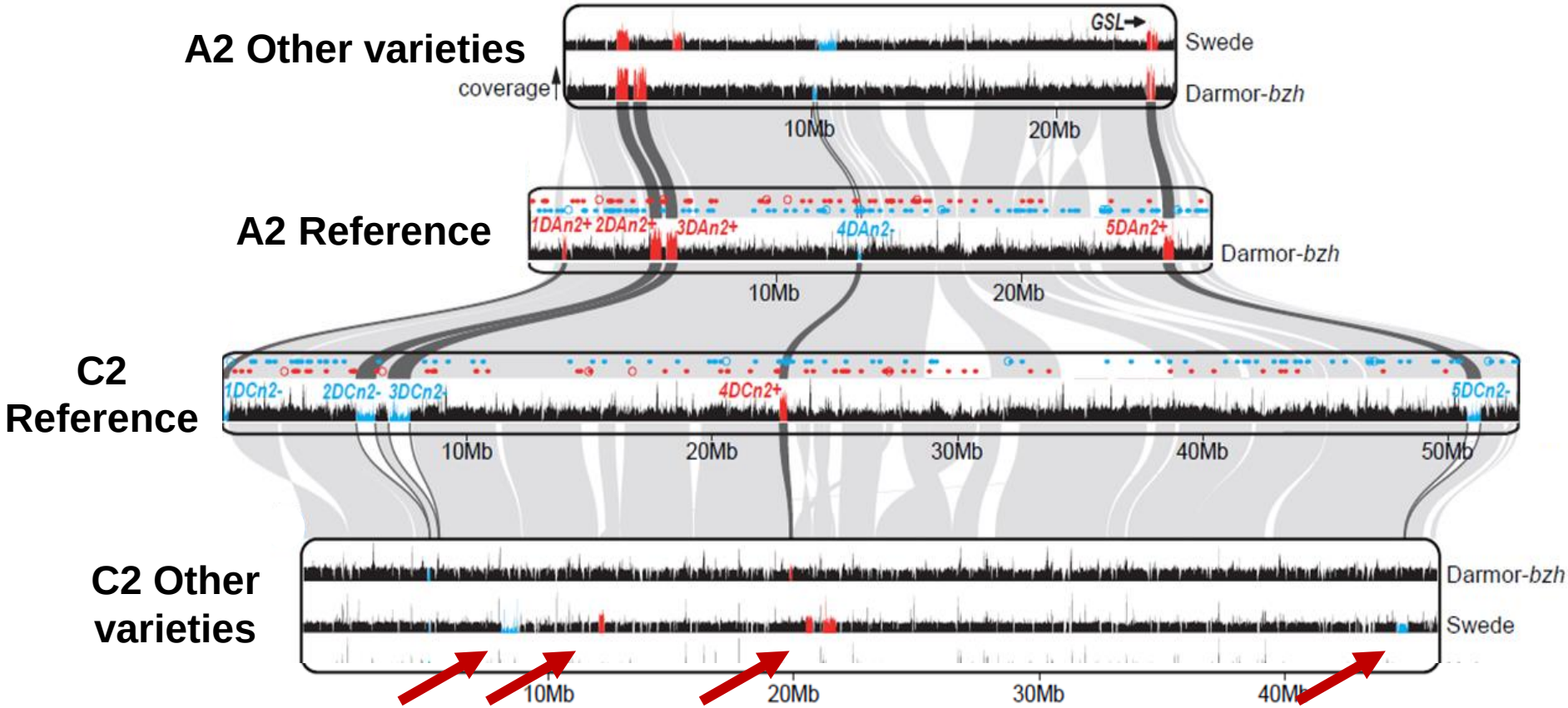
Genomics-led "traditional" approaches to crop genome engineering

Ian Bancroft

22nd May, 2019

- Introduction to rapeseed genomics
- **Genome structural variation by homoeologous exchange**
- Genome structural variation by alien introgression
- Genome sequence and structural variation by gamma irradiation

Homoeologous exchanges (HEs) identified in *B. napus* accessions



Genome dosage difficult to determine, especially when HEs in reference

Science AAAS

Chalhoub, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950-953, 2015

Visualizing genome dosage variation using mRNAseq data

Quantify transcript abundance (RPKM) using AC pan-transcriptome reference sequence

↓

Filter data for reciprocal top BLAST hit gene pairs

↓

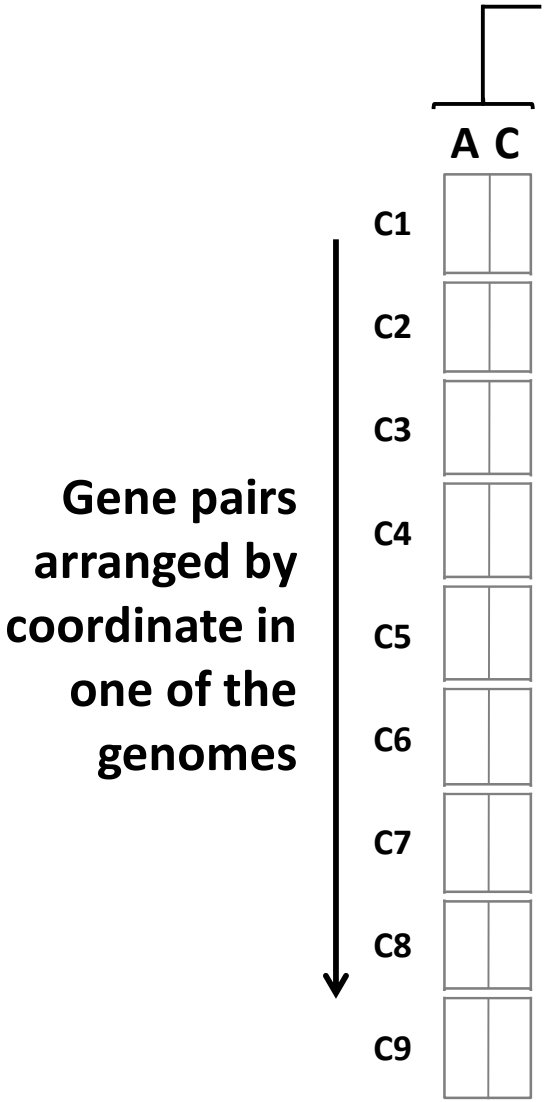
Code intensity of shading based on RPKM values

↓

Visualise using tile plots with gene pairs ordered by one genome

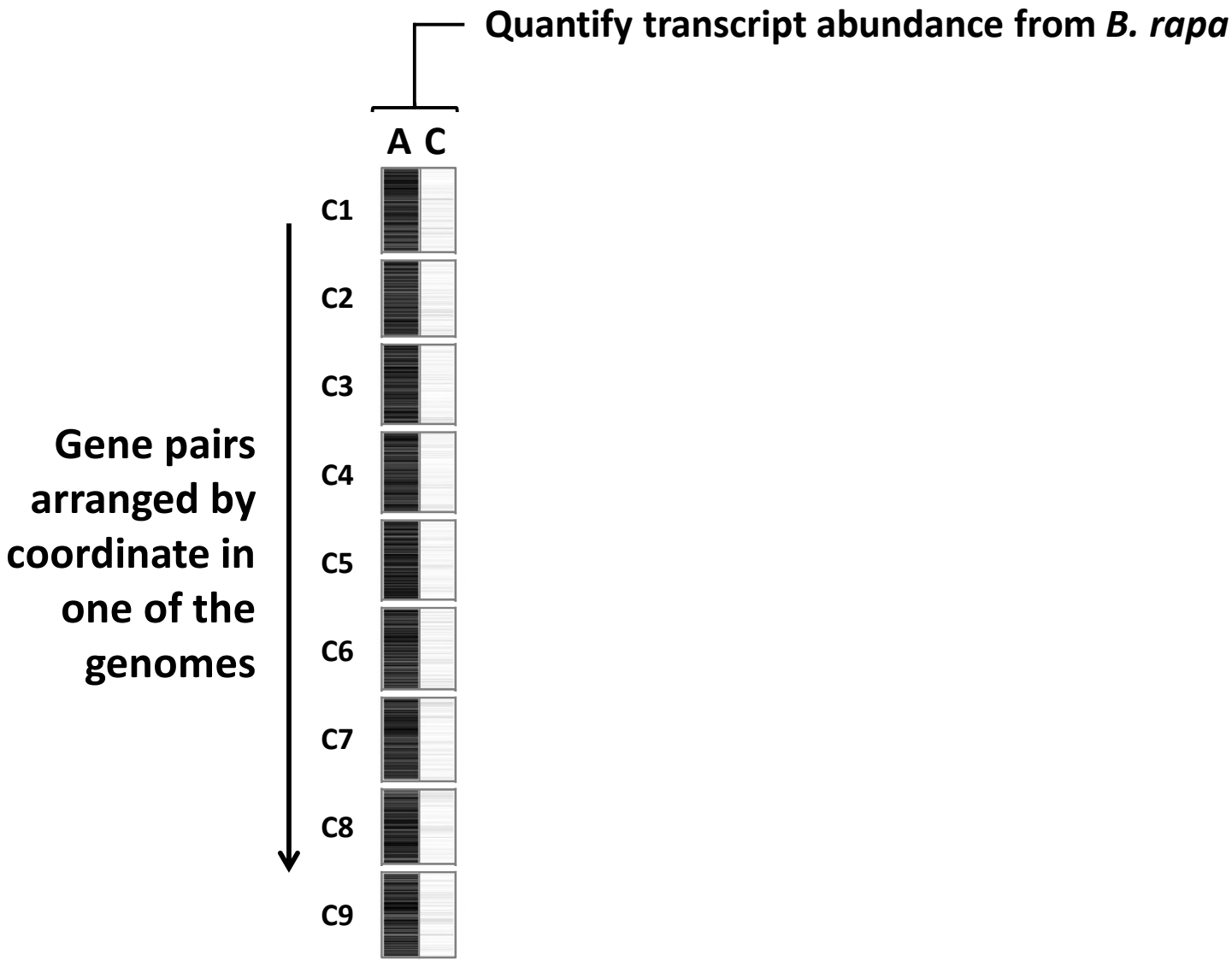
Visualizing genome dosage variation using mRNAseq data

Positions of 31,591 homoeologous gene pairs, arranged by the order in the C genome of the C genome homoeologoue

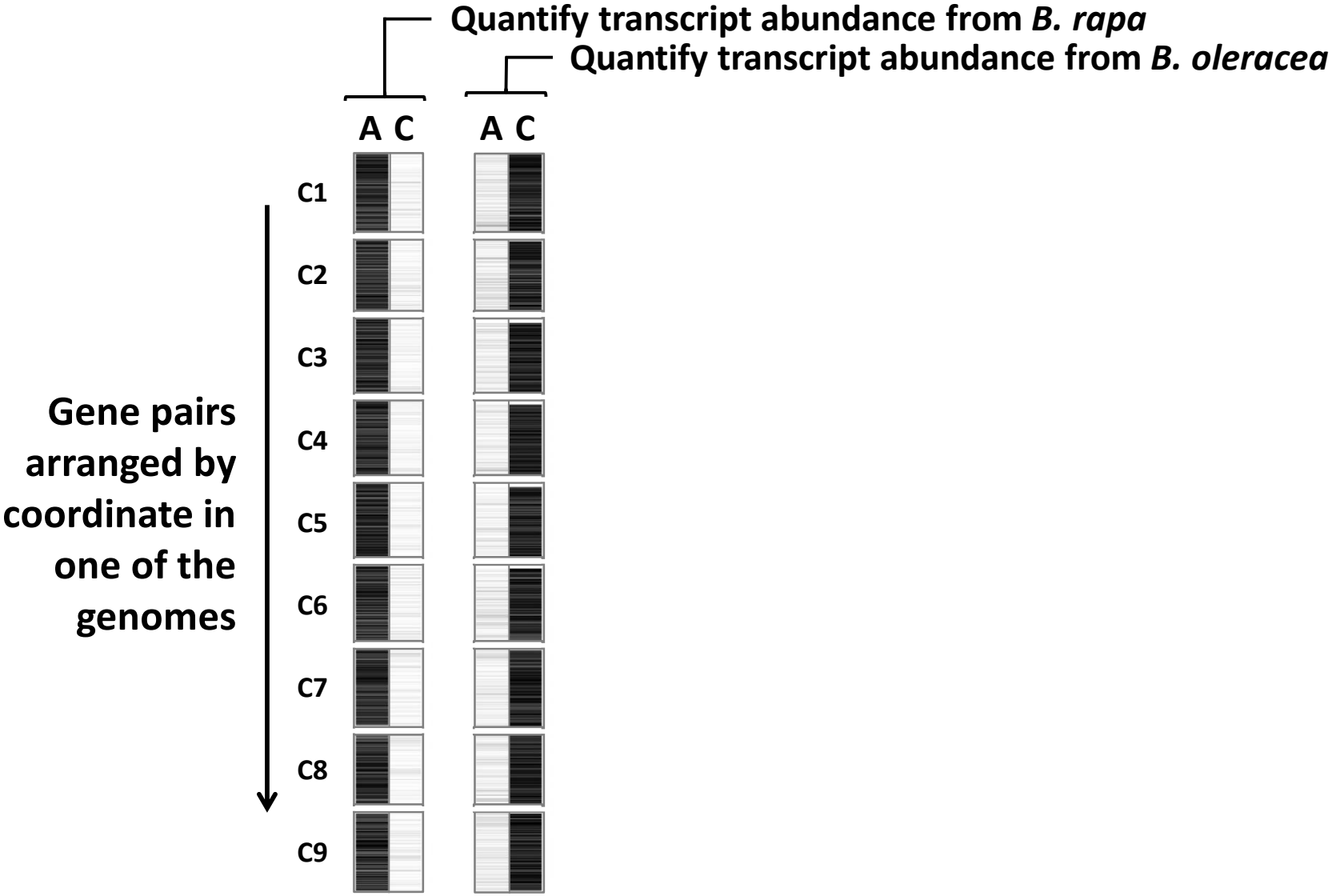


	A genome homoeologue	Position in A genome	Order in A	Order in C	Position in C genome	C genome homoeologue
2	Cab020597.1	A01_000007596_000004363	1	1	C01_000004096_000006192	Bo1g001030.1
3	Cab020596.1	A01_000009147_000008191	2	25377	C07_048354208_048355539	Bo7g120870.1
4	Cab020595.2	A01_000011403_000009778	3	2	C01_000007519_000009142	Bo1g001040.1
5	Cab020594.1	A01_000012765_000011917	4	3	C01_000009720_000010577	Bo1g001050.1
6	Cab020592.1	A01_000021944_000019956	5	5	C01_000037620_000038611	Bo1g001080.1
7	Cab020591.1	A01_000024271_000022281	6	6	C01_000039858_000040949	Bo1g001090.1
8	Cab020590.1	A01_000032446_000028836	7	7	C01_000046852_000050484	Bo1g001100.1
9	Cab020589.1	A01_000033293_000032595	8	30597	C09_021661539_021662053	Bo9g071600.1
10	Cab020588.1	A01_000035983_000033952	9	9	C01_000068965_000070059	Bo1g001140.1
11	Cab020581.1	A01_000063572_000063039	10	10	C01_000103201_000103782	Bo1g002150.1
12	Cab020579.1	A01_000079019_000078723	11	11	C01_000112982_000113278	Bo1g002190.1
13	Cab020577.1	A01_000088358_000088041	12	12	C01_000124351_000124668	Bo1g002200.1
14	Cab020576.1	A01_000094861_000093230	13	13	C01_000128563_000130200	Bo1g002210.1
15	Cab020575.1	A01_000097996_000095110	14	14	C01_000130440_000133386	Bo1g002220.1
16	Cab020574.1	A01_000100108_000098057	15	15	C01_000134318_000135251	Bo1g002230.1
17	Cab020573.1	A01_000105326_000101322	16	16	C01_000137577_000141372	Bo1g002250.1
18	Cab020571.1	A01_000115025_000112704	17	17	C01_000148961_000151258	Bo1g002280.1
19	Cab020570.1	A01_000119528_000116614	18	18	C01_000153250_000156261	Bo1g002300.1
20	Cab020569.1	A01_000129775_000126831	19	19	C01_000164958_000168351	Bo1g002310.1
21	Cab020568.1	A01_000135472_000131405	20	20	C01_000172779_000174206	Bo1g002330.1
22	Cab020567.1	A01_000143020_000139039	21	21	C01_000175196_000180272	Bo1g002340.1
23	BnaA01g05330D	A01_000146772_000145212.00	22	23	C01_000183323_000183774	Bo1g002360.1
24	BnaA01g05340D	A01_000146772_000145212.00	23	24	C01_000184847_000187483	Bo1g002370.1
25	Cab020565.1	A01_000146829_000145011	24	22	C01_000181177_000182520	Bo1g002350.1
26	Cab020563.1	A01_000155768_000155126	25	25	C01_000191317_000191852	Bo1g002400.1
27	Cab020562.1	A01_000158109_000156564	26	26	C01_000192637_000194204	Bo1g002410.1
28	Cab020560.1	A01_000162233_000159755	27	27	C01_000195435_000197330	Bo1g002420.1
29	Cab020558.1	A01_000164388_000163736	28	29	C01_000202463_000202892	Bo1g002450.1
30	Cab020557.1	A01_000167670_000166905	29	30	C01_000205351_000206116	Bo1g002460.1
31	Cab020556.1	A01_000168725_000168246	30	31	C01_000206881_000208172	Bo1g002470.1
32	Cab020555.1	A01_000170941_000170440	31	32	C01_000209076_000210429	Bo1g002480.1
33	Cab020554.1	A01_000171955_000171563	32	33	C01_000210647_000211042	Bo1g002490.1
34	Cab020553.1	A01_000180461_000175822	33	34	C01_000216852_000221622	Bo1g002500.1
35	Cab020552.1	A01_000184155_000183811	34	35	C01_000223558_000223902	Bo1g002510.1
36	Cab020550.1	A01_000190655_000188381	35	37	C01_000228012_000229293	Bo1g002520.1
37	Cab020549.1	A01_000199980_000195942	36	38	C01_000237167_000240760	Bo1g002530.1
38	Cab020547.1	A01_000210518_000207773	37	39	C01_000268432_000270656	Bo1g002540.1
39	Cab020546.2	A01_000233705_000220053	38	40	C01_000278884_000293749	Bo1g002550.1
40	Cab020545.1	A01_000236573_000235062	39	41	C01_000294476_000298099	Bo1g002560.1
41	Cab020544.1	A01_000240300_000237136	40	42	C01_000298888_000301646	Bo1g002570.1
42	Cab020543.1	A01_000244118_000242423	41	43	C01_000305351_000307099	Bo1g002590.1
43	Cab020540.1	A01_000247210_000246059	42	44	C01_000312999_000314150	Bo1g002620.1
44	Cab020539.1	A01_000249298_000248129	43	45	C01_000315809_000316990	Bo1g002630.1

Visualizing genome dosage variation using mRNAseq data



Visualizing genome dosage variation using mRNAseq data



Visualizing genome dosage variation using mRNAseq data

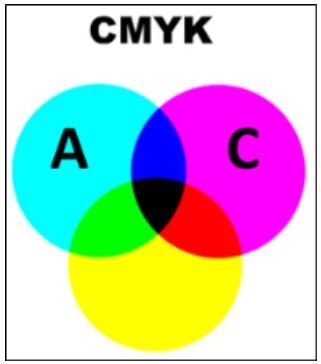
Quantify transcript abundance (RPKM) using
AC pan-transcriptome reference sequence

↓

Filter data for reciprocal top BLAST hit gene pairs

↓

Use CMYK colour-coding based on RPKM values



↓

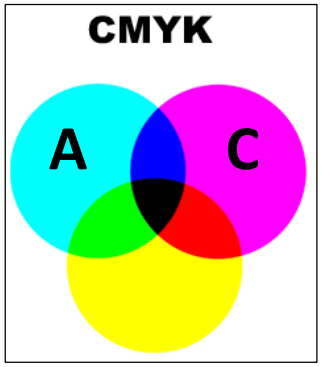
Visualise using tile plots with gene pairs ordered by one genome

Plant Biotechnology Journal

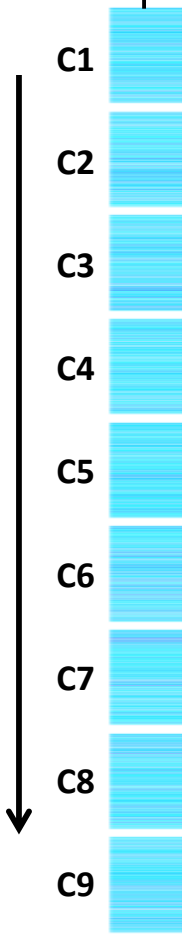
Harper, et al. Genome distribution of differential homoeologue contributions to leaf gene expression in bread wheat. Plant Biotechnology Journal 14:1207-1214, 2016

Visualizing genome dosage variation using mRNAseq data

A (*B. rapa*) = cyan



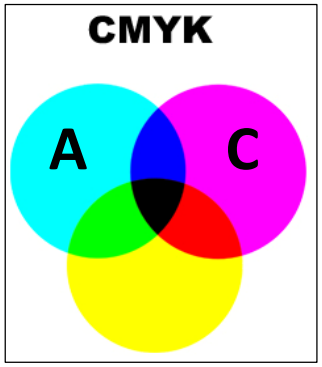
Gene pairs arranged by coordinate in one of the genomes



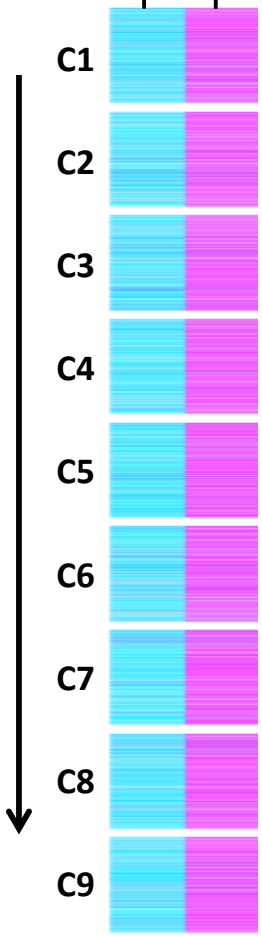
Visualizing genome dosage variation using mRNAseq data

A (*B. rapa*) = cyan

C (*B. oleracea*) = magenta



Gene pairs arranged by coordinate in one of the genomes

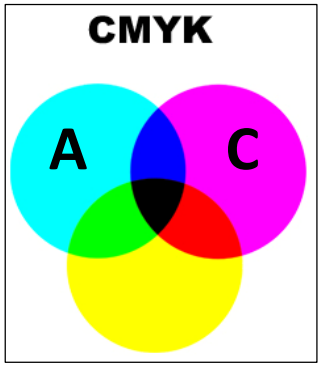


Visualizing genome dosage variation using mRNAseq data

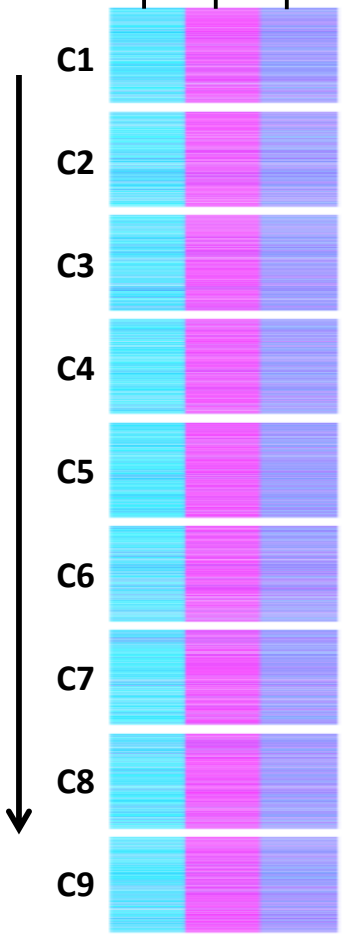
A (*B. rapa*) = cyan

C (*B. oleracea*) = magenta

in silico A + C (*B. napus*) = blue



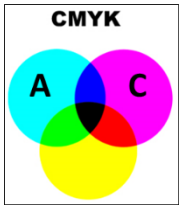
Gene pairs arranged by coordinate in one of the genomes

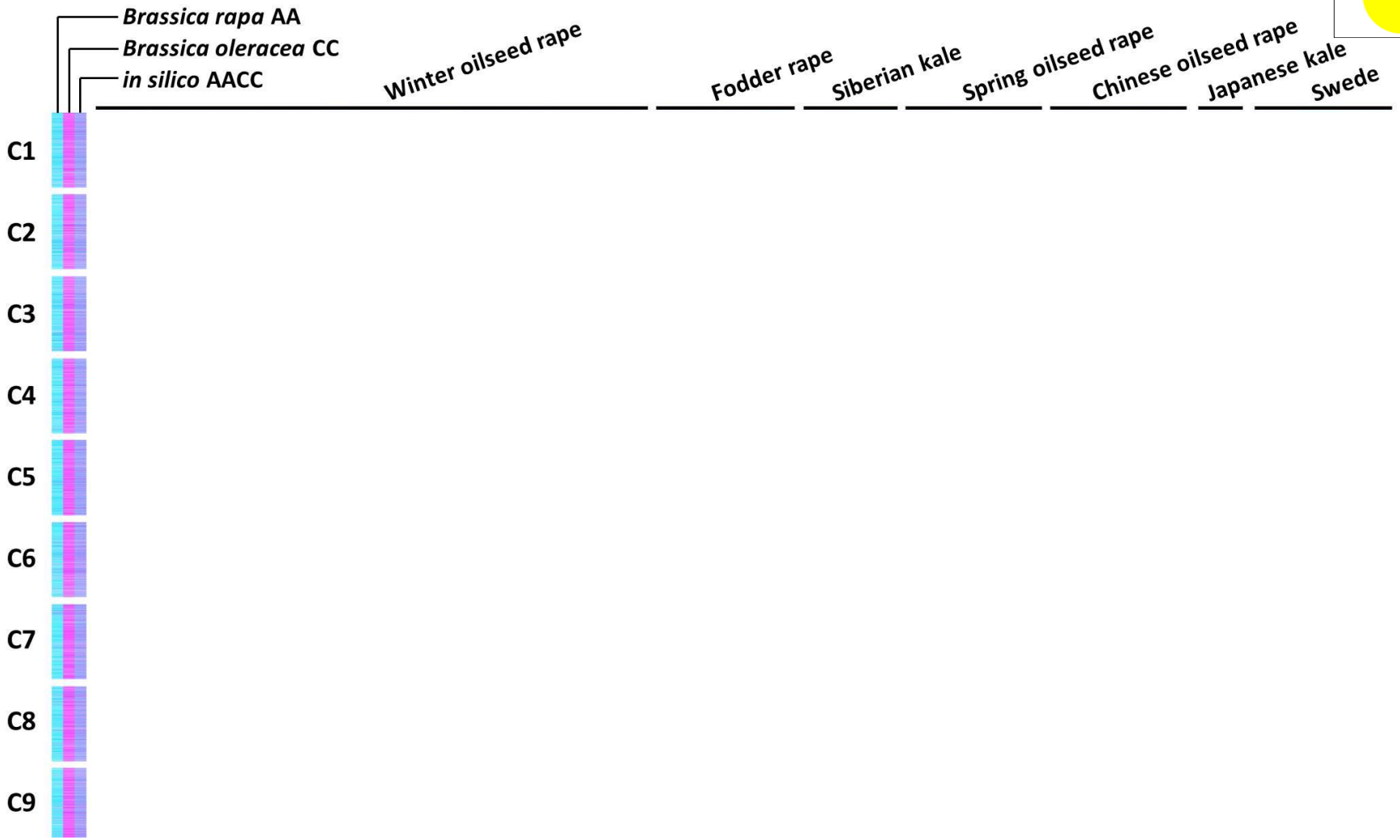


Assessment of a panel of *B. napus* varieties by mRNAseq

27 varieties selected across all crop types; x 4 biological replicates

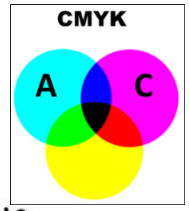
Expect to see blue signal throughout the genome

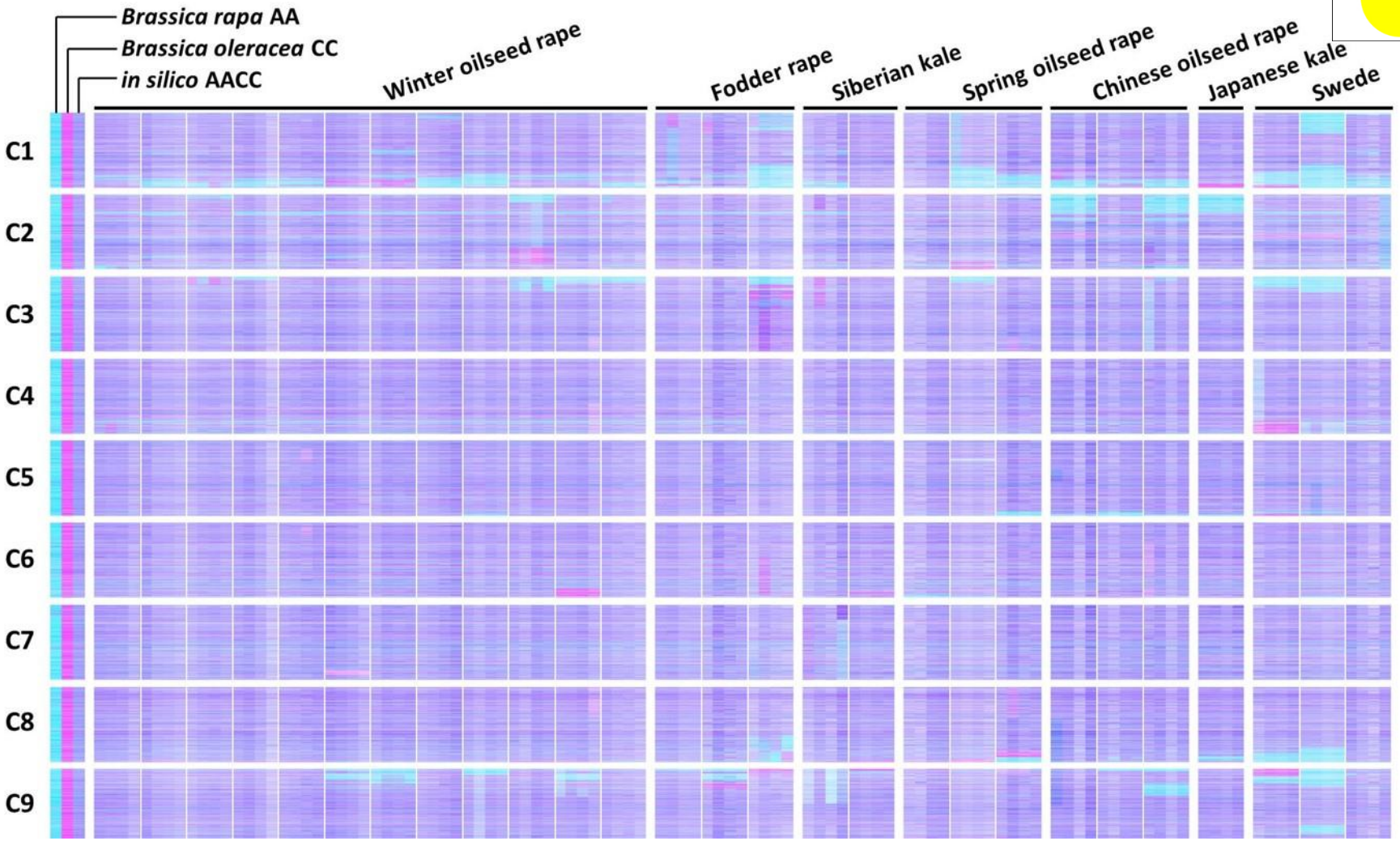




Assessment of a panel of *B. napus* varieties by mRNAseq

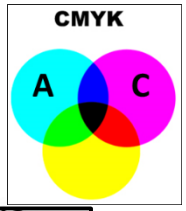
Homoeologous exchanges duplicating the A genome appear cyan, those duplicating the C genome appear magenta

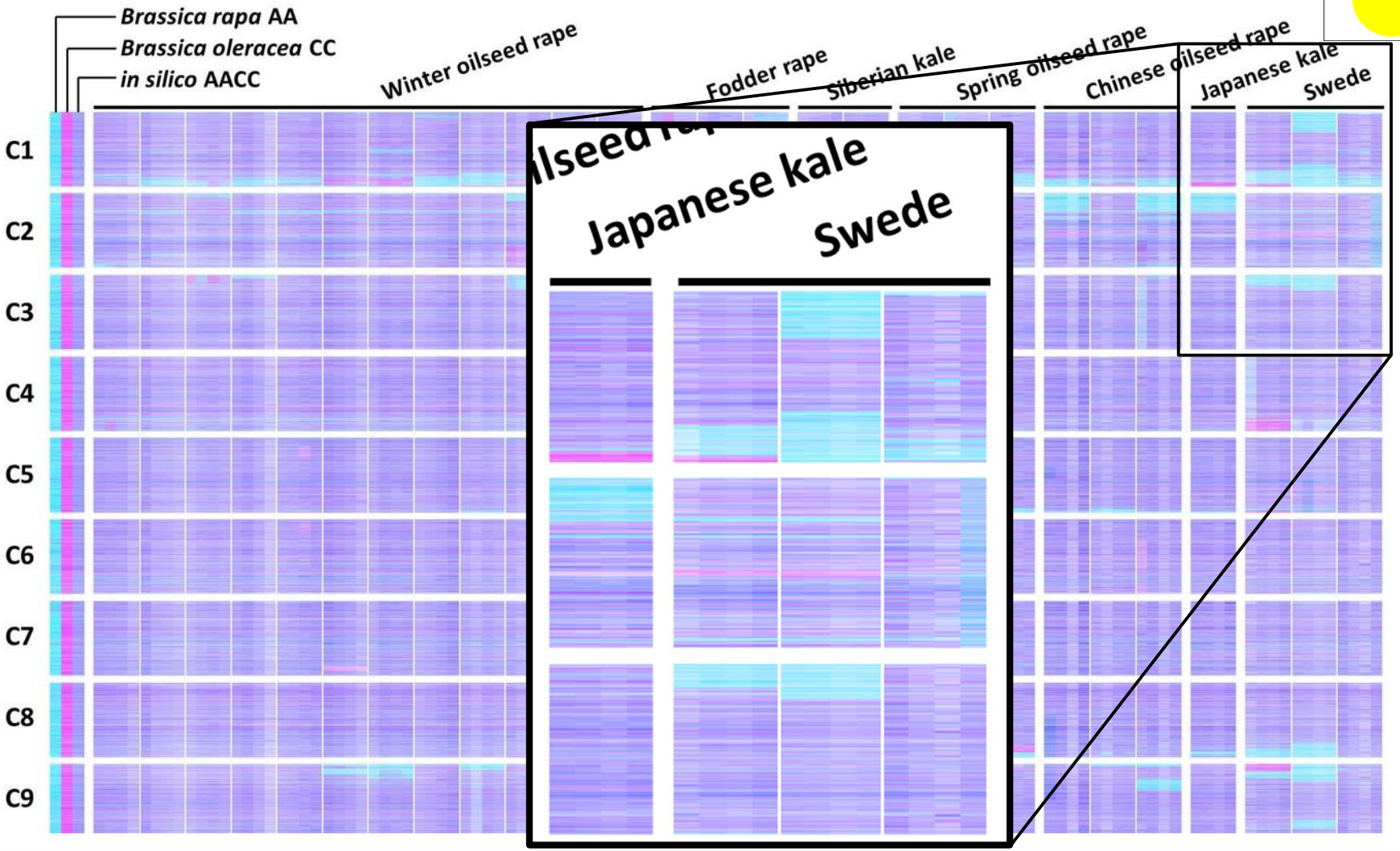




Assessment of a panel of *B. napus* varieties by mRNAseq

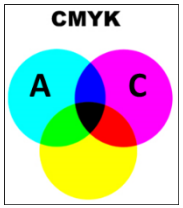
Homoeologous exchanges duplicating the A genome appear cyan, those duplicating the C genome appear magenta

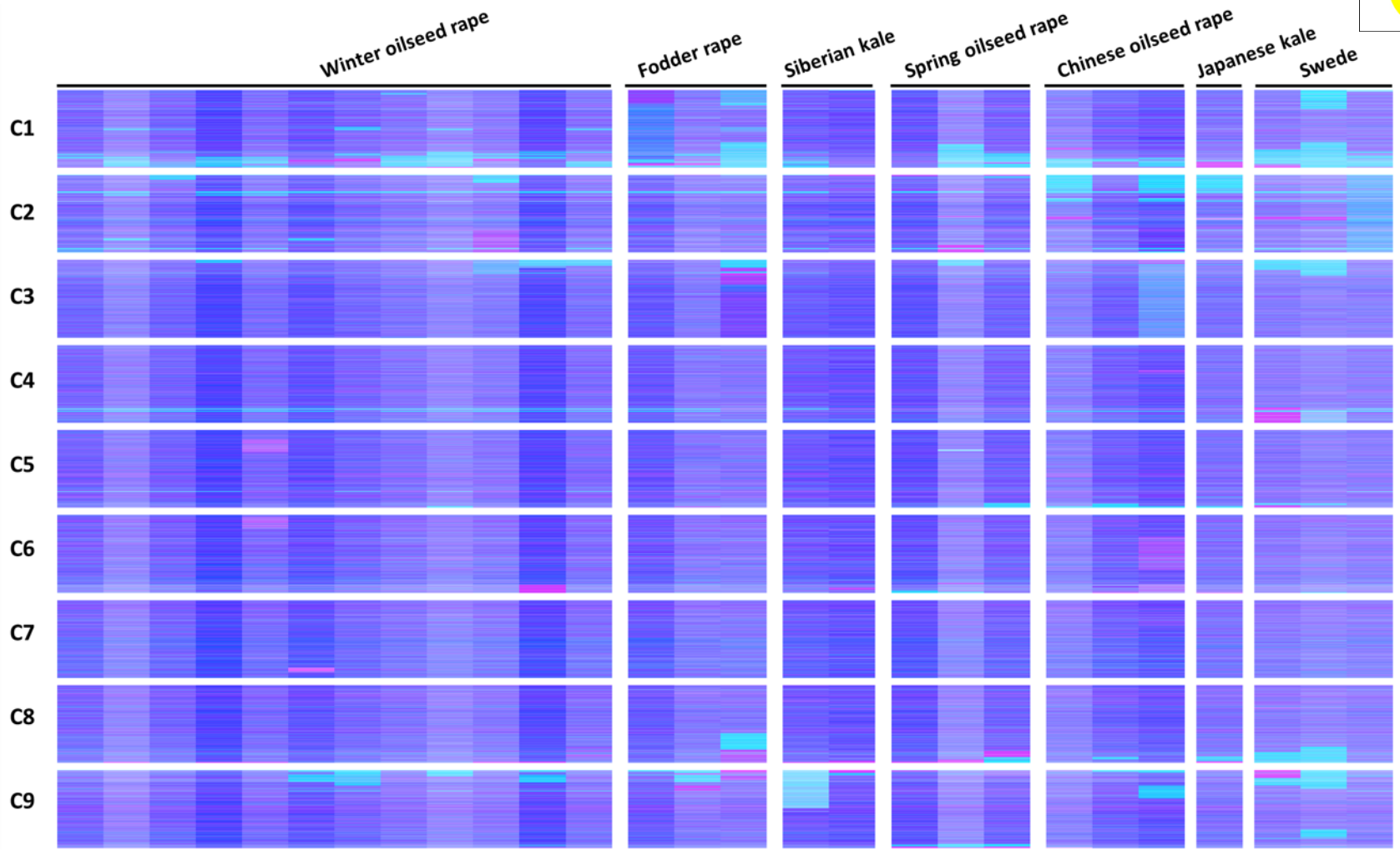




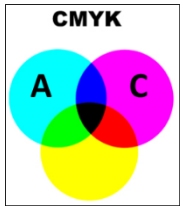
Confirmed by genome sequencing as structural variation

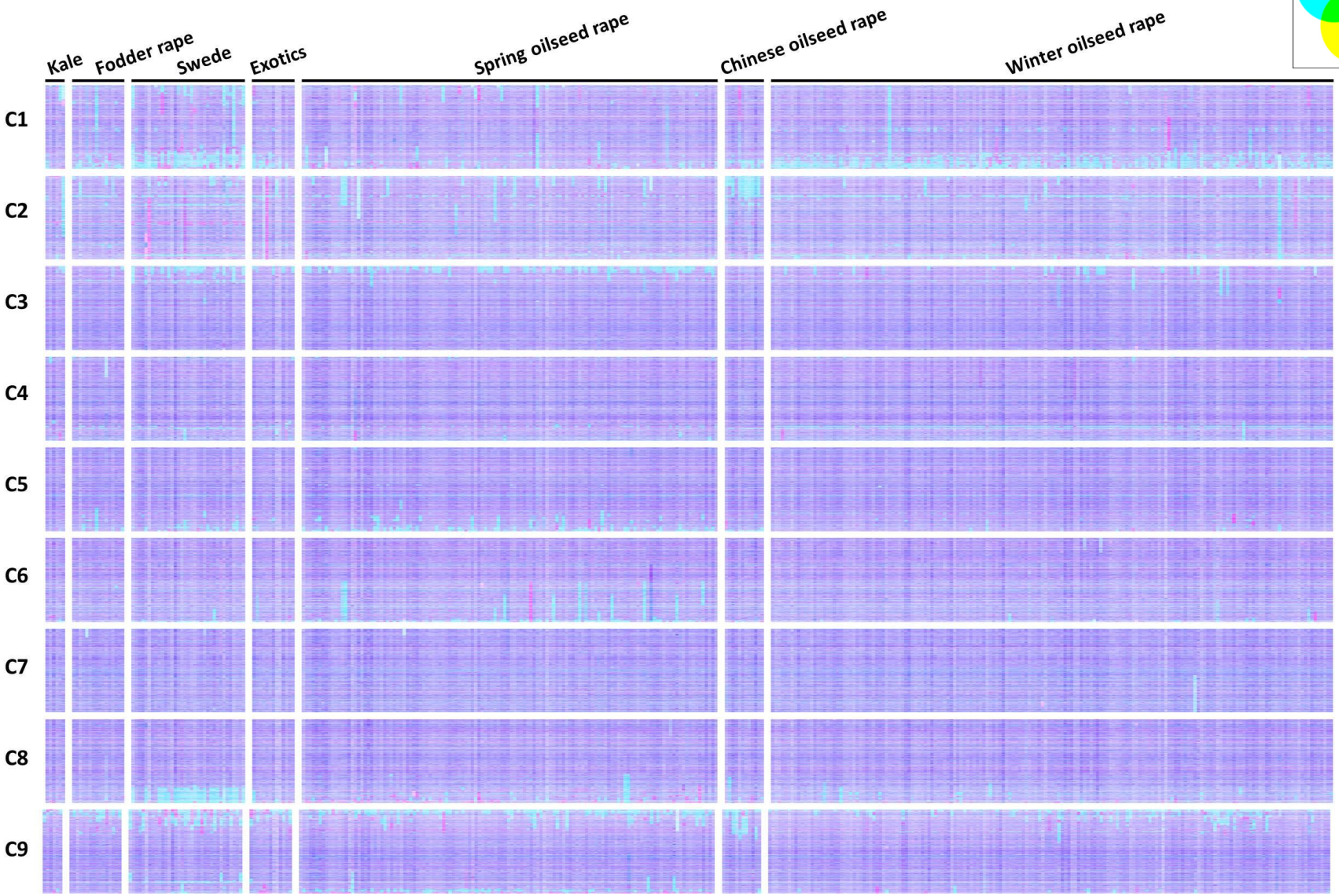
Homoeologous exchanges confirmed by genome skim sequencing





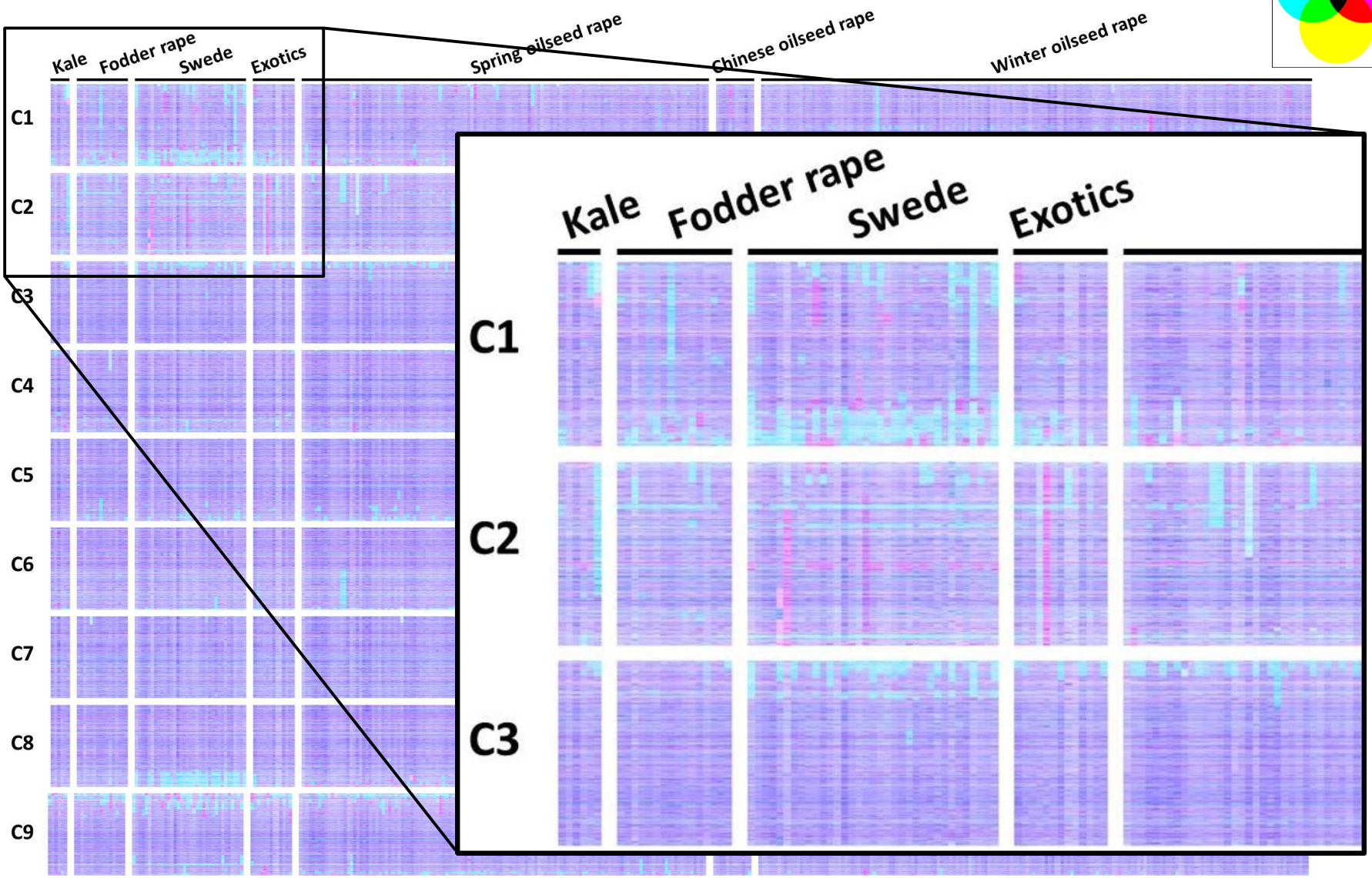
Homoeologous exchanges in *B. napus* RIPR diversity panel





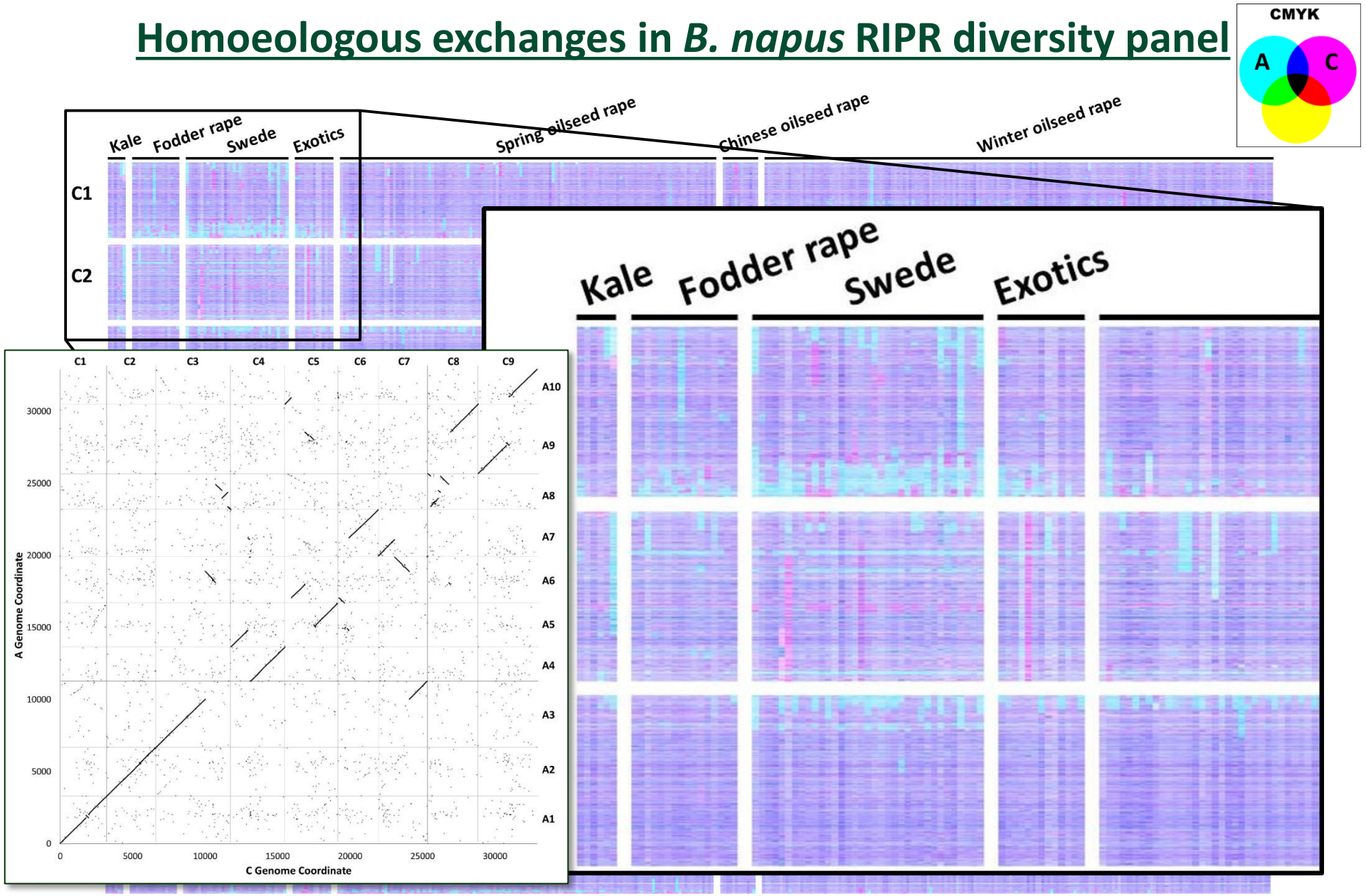
383 *B. napus* accessions

Homoeologous exchanges in *B. napus* RIPR diversity panel



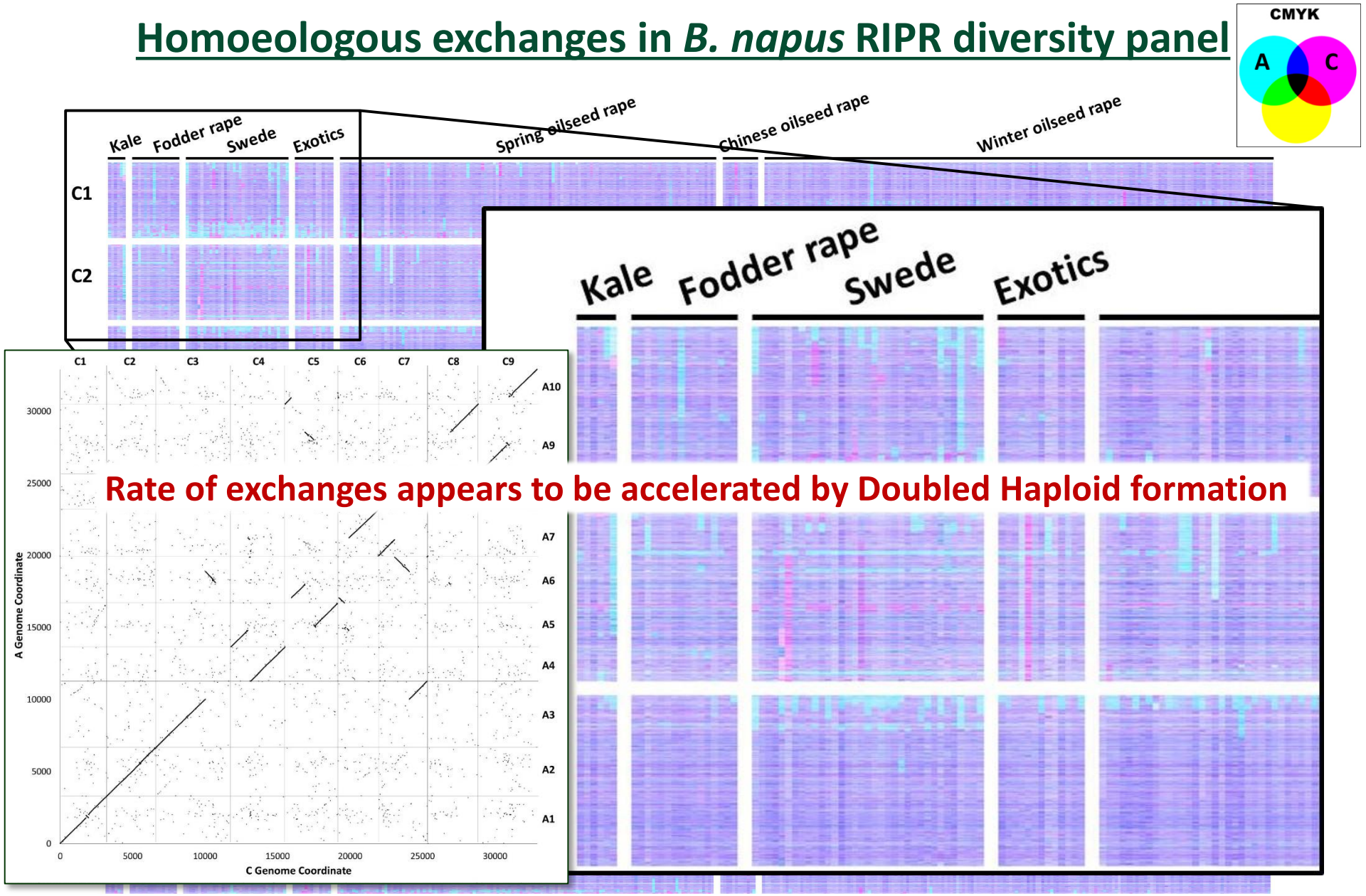
383 *B. napus* accessions

Homoeologous exchanges in *B. napus* RIPR diversity panel



383 *B. napus* accessions

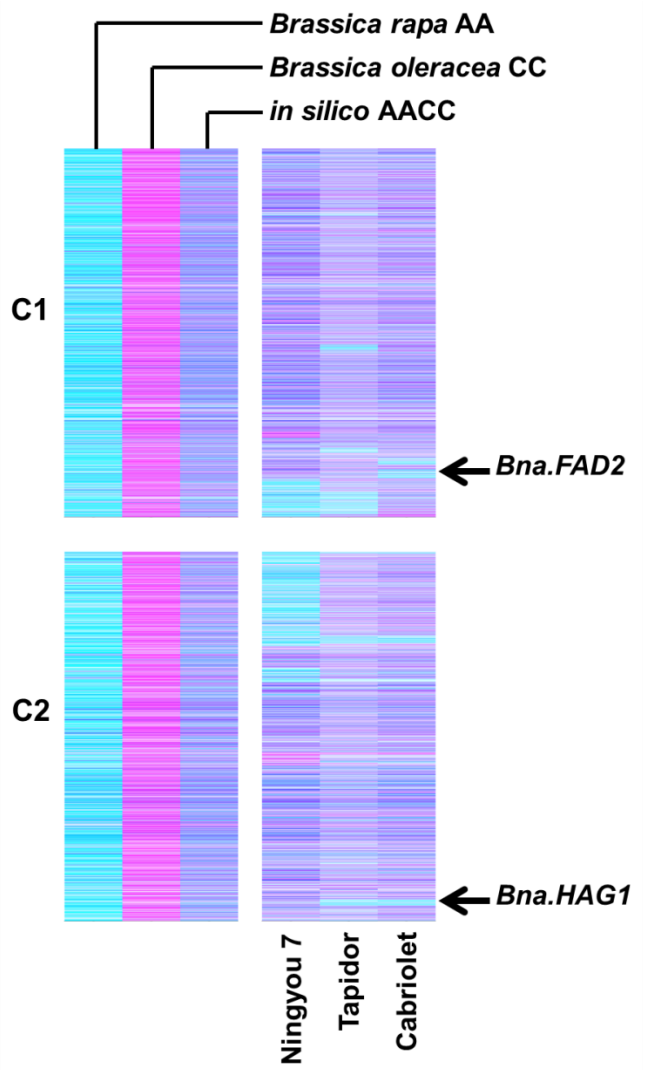
Homoeologous exchanges in *B. napus* RIPR diversity panel



Rate of exchanges appears to be accelerated by Doubled Haploid formation

383 *B. napus* accessions

Homoeologous exchanges affect crop traits



Substitution of active orthologue of *FAD2* by inactive orthologue reduces seed oil PUFA content

Substitution of active orthologue of *HAG1* by inactive orthologue reduces seed glucosinolate content

Plant Biotechnology Journal

Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotechnology J. 15:594-604, 2017

Genomics-led "traditional" approaches to crop genome engineering

Ian Bancroft

22nd May, 2019

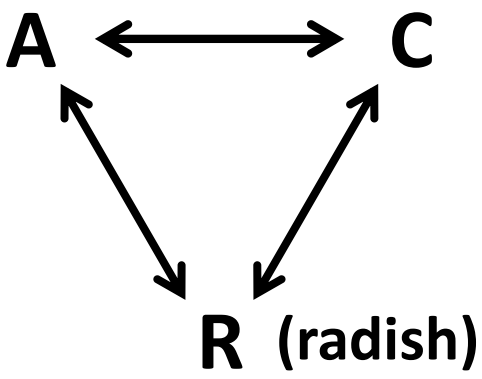
- Introduction to rapeseed genomics
- Genome structural variation by homoeologous exchange
- **Genome structural variation by alien introgression**
- Genome sequence and structural variation by gamma irradiation

"Alien introgression": Ogura cytoplasmic male

- Male sterility system widely used in F_1 hybrid rapeseed varieties
- Cytoplasm derived from radish with mitochondrial gene *orf138*
- PPR-containing gene (*Rfo*) introgressed from radish restores fertility

Identifying orthologous triplets of genes

Identify triplets of orthologues between *Brassica* A, *Brassica* C and radish genome, as 3-way reciprocal top BLAST hits between CDS gene models



46,652 A genome models have hits in C genome
41,344 A genome models have hits in R genome
52,539 C genome models have hits in A genome
48,578 C genome models have hits in R genome
48,746 R genome models have hits in A genome
50,910 R genome models have hits in C genome

22,449 3-way reciprocal top BLAST hit pairs

Visualizing alien introgressions into *B. napus*

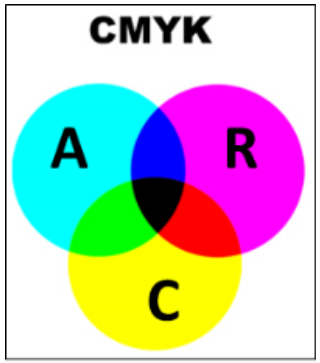
Quantify transcript abundance for each gene using ACR transcriptome reference sequence

↓

Filter data for reciprocal top BLAST hit pairs

↓

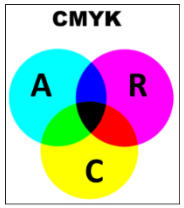
Use CMYK colour-coding based on RPKM values

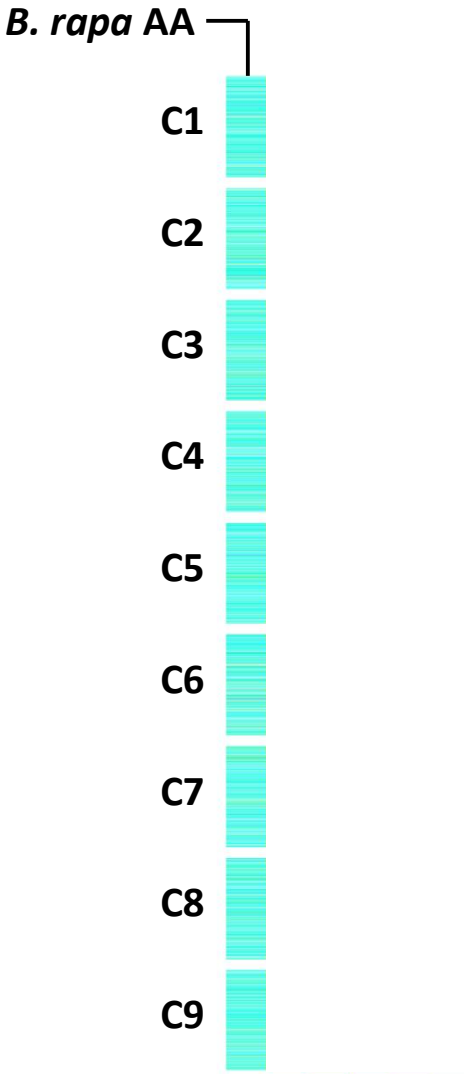


↓

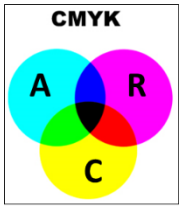
Visualise using tile plots with CDS pairs ordered by one genome

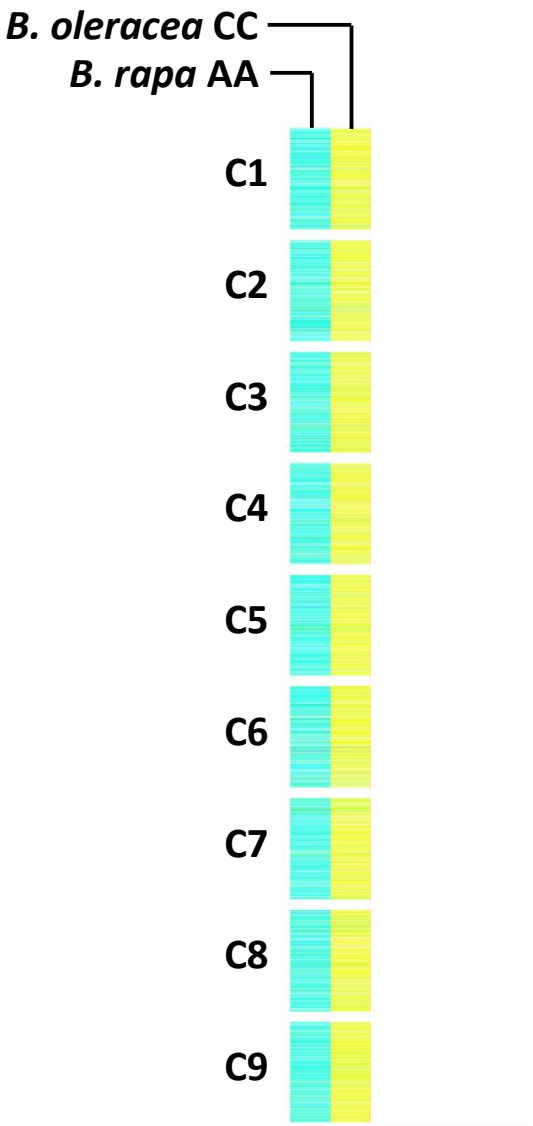
Visualization of a radish introgression in *B. napus*



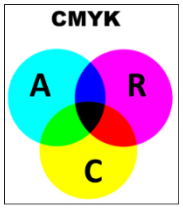


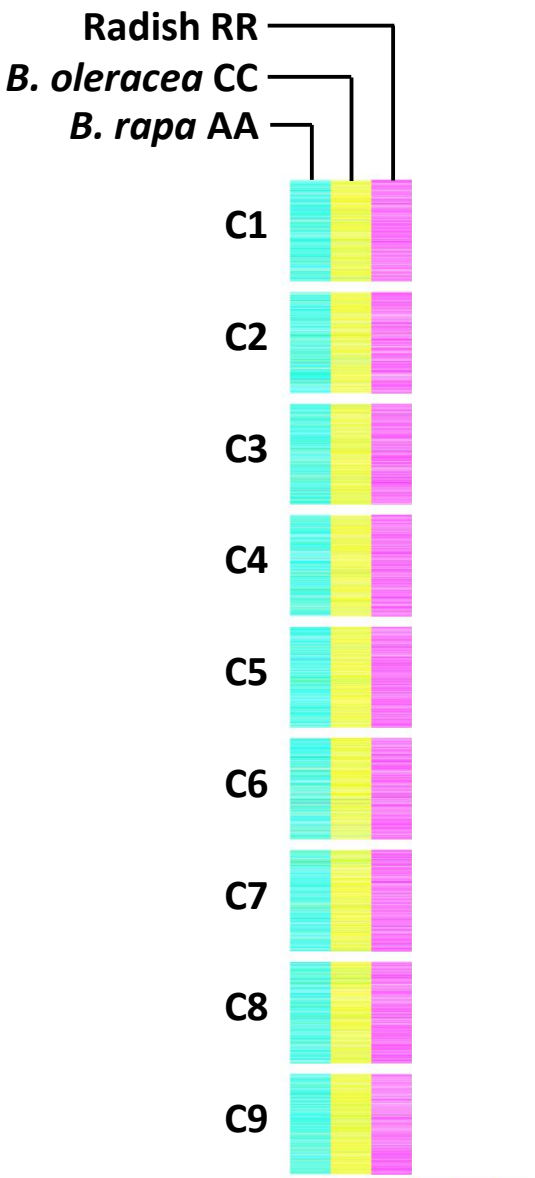
Visualization of a radish introgression in *B. napus*



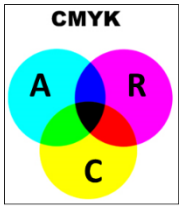


Visualization of a radish introgression in *B. napus*

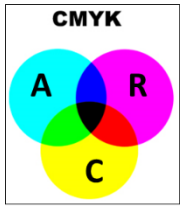


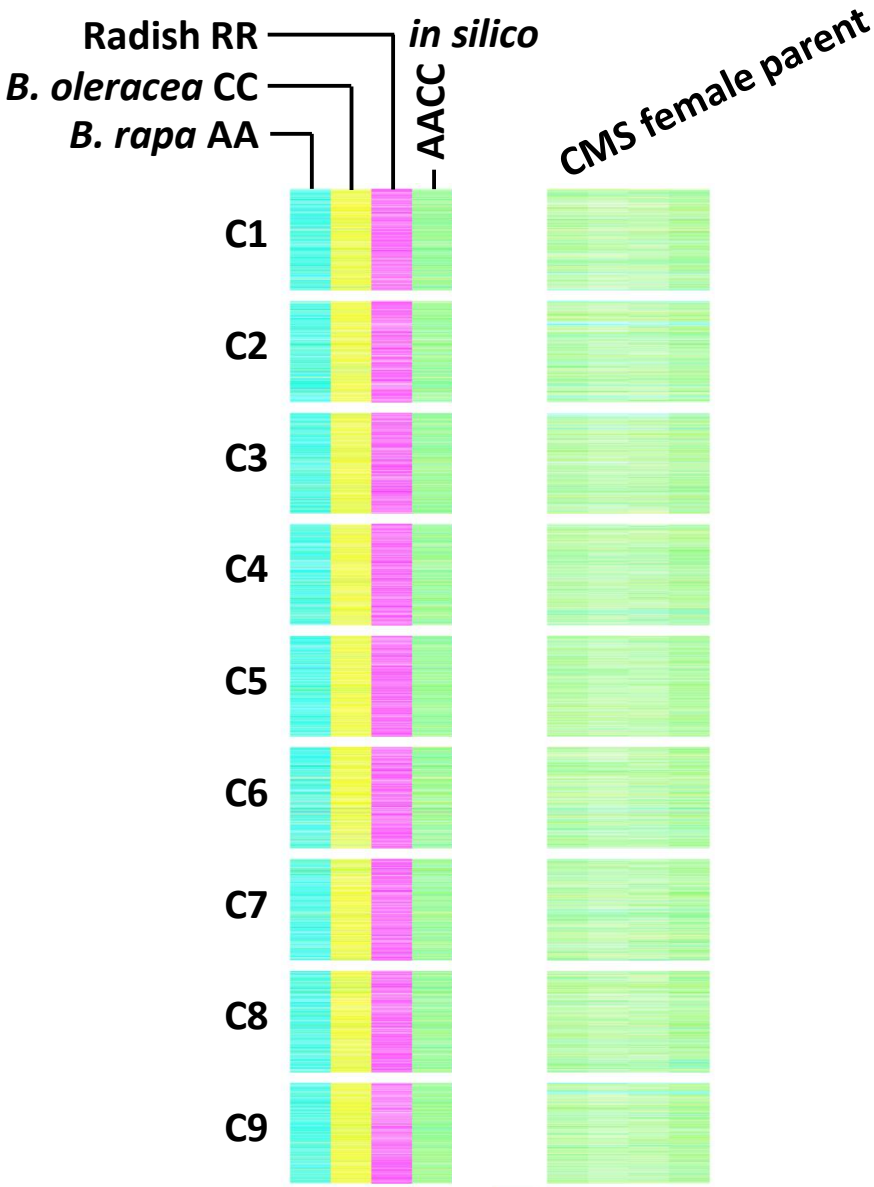


Visualization of a radish introgression in *B. napus*

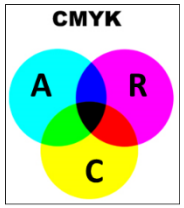


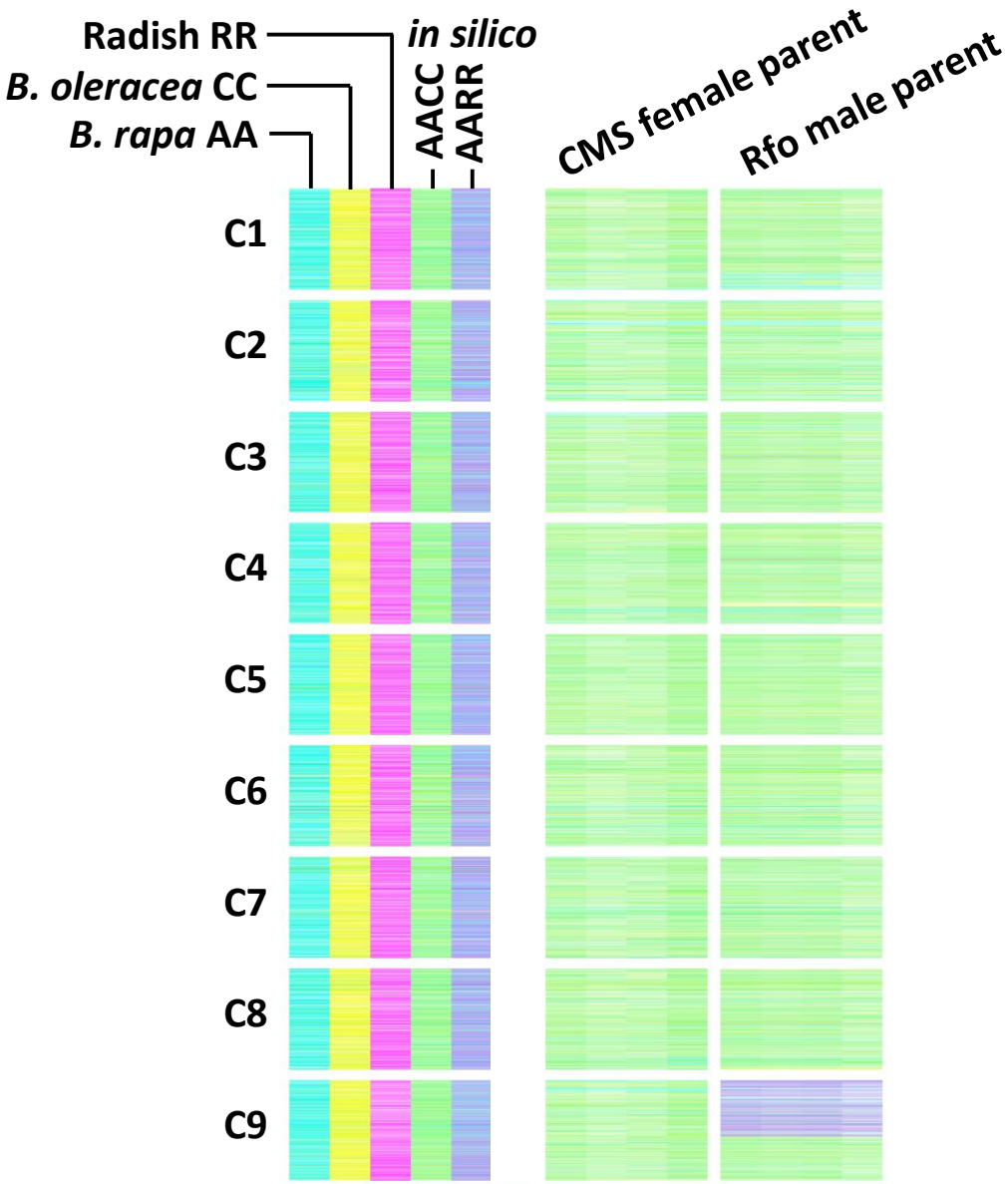
Visualization of a radish introgression in *B. napus*



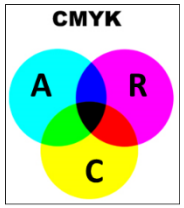


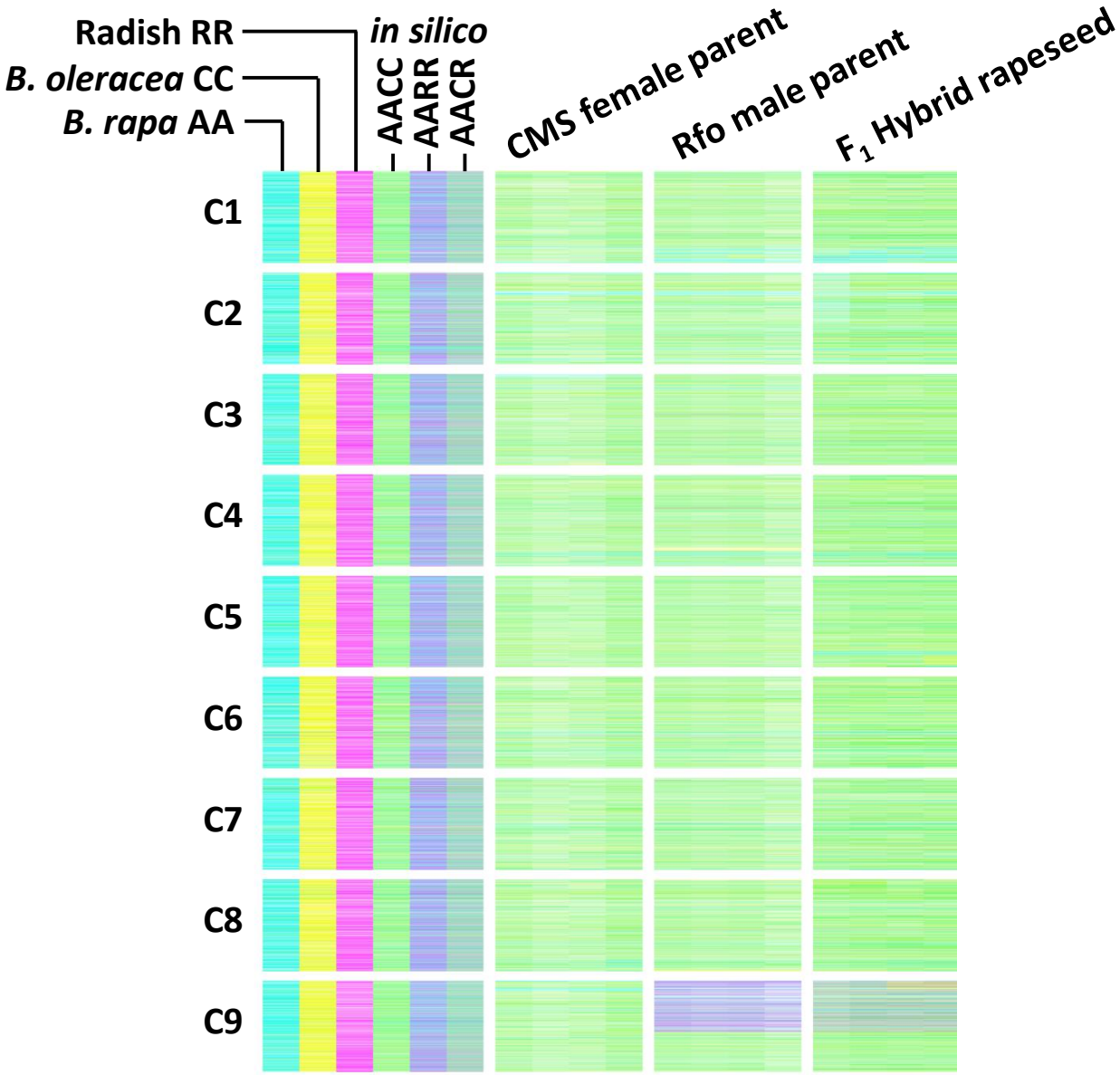
Visualization of a radish introgression in *B. napus*



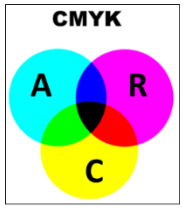


Visualization of a radish introgression in *B. napus*



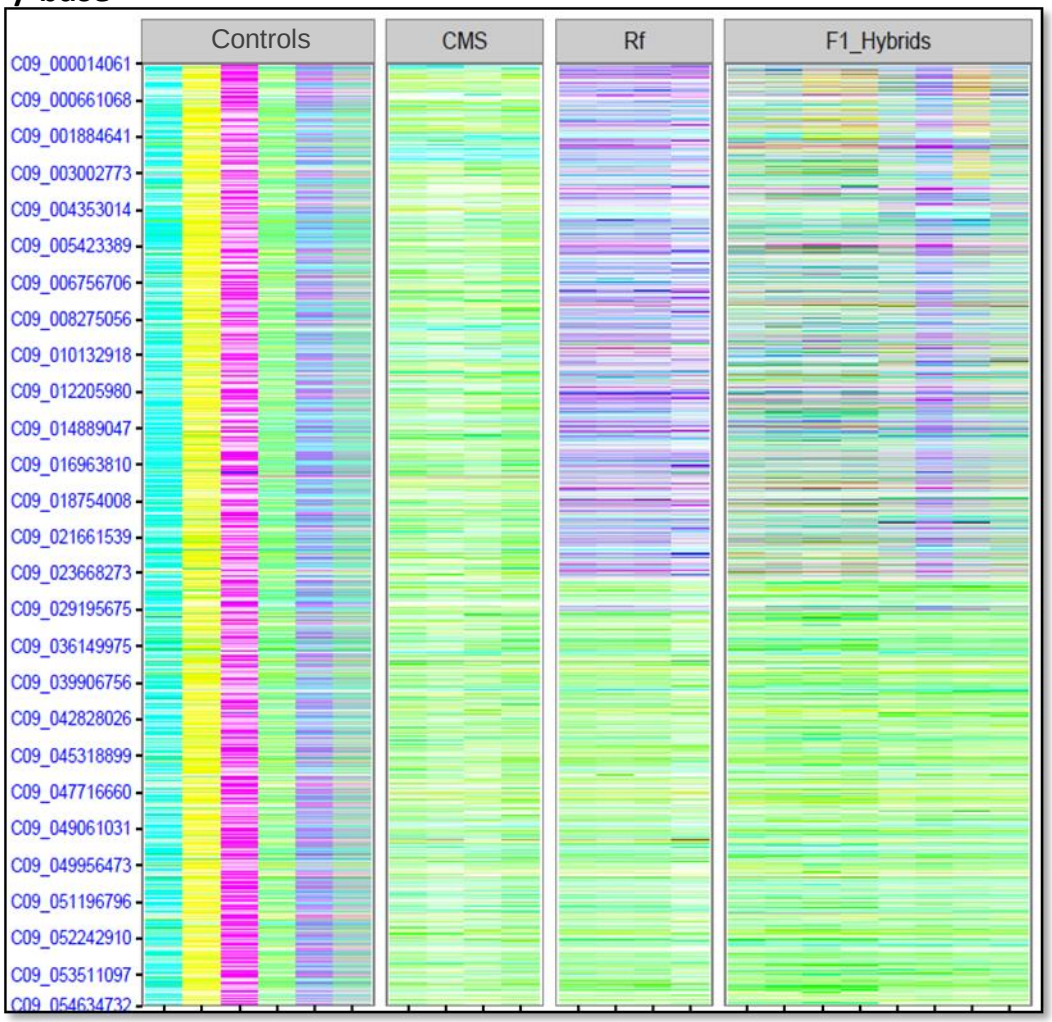


Visualization of a radish introgression in *B. napus*



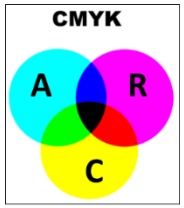
Coordinate / base

C9

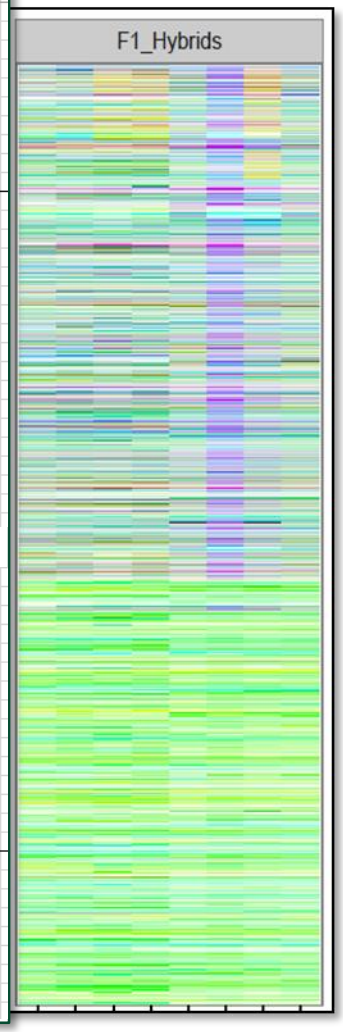


The introgression corresponds to the top half of chromosome C9

Visualization of a radish introgression in *B. napus*

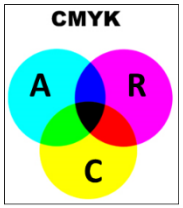


	A	B	C	D	E	F	AW	AX	AY	AZ	BB	BC
1	A	C	R	A_pos	C_pos	R_pos		Mean Rf_A	Mean Rf_C	mean Rf_R	Ratio R/C	
16853	Cab014603.1	Bo8g118150.	Rs252020	A09_045833808	C08_041638382	R05_023304204		0.1	0.83	0	0	
16854	Cab014604.1	Bo8g118160.	Rs252030	A09_045836539	C08_041641276	R05_023307826		0.4	0.57	0.07	0.12	
16855	Cab014605.4	Bo8g118180.	Rs252050	A09_045844458	C08_041647829	R05_023314225		0.03	0.83	0	0	
16856	Cab014606.1	Bo8g118190.	Rs252060	A09_045853826	C08_041656238	R05_023328802		0.1	0.77	0.1	0.13	
16857	Cab014608.1	Bo8g118210.	Rs252080	A09_045862640	C08_041670596	R05_023342540		0.23	0.73	0.13	0.18	
16858	Cab014609.1	Bo8g118220.	Rs252100	A09_045875521	C08_041678897	R05_023375798		0.13	0.73	0.27	0.36	
16859	Cab014612.1	Bo8g118240.	Rs252110	A09_045885127	C08_041690198	R05_023401297		0.03	0.77	0.07	0.09	
16860	Cab014614.1	Bo8g118270.	Rs252730	A09_045903152	C08_041712867	R05_023141472		0.1	0.73	0	0	
16861	Cab014615.2	Bo8g118280.	Rs252740	A09_045907620	C08_041716559	R05_023138797		0.27	0.73	0.1	0.14	
16862	Cab029989.1	Bo8g118290.	Rs000850	A10_000017567	C08_041718705	R01_000198348		0.4	0.73	0.1	0.14	
16863	Cab037725.3	Bo9g001030.	Rs481780	A09_000006809	C09_000014061	R09_024237368		0.43	0	0.7	700	
16864	Cab037721.1	Bo9g001060.	Rs481810	A09_000021078	C09_000026736	R09_024230220		0.73	0	0.47	467	
16865	Cab037717.1	Bo9g001080.	Rs481860	A09_000042770	C09_000041792	R09_024198799		0.47	0	0.53	533	
16866	Cab037716.1	Bo9g001090.	Rs481870	A09_000045455	C09_000042999	R09_024191985		0.57	0.07	0.37	5.42	
16867	Cab037715.2	Bo9g001100.	Rs481880	A09_000049418	C09_000045805	R09_024188847		0.43	0	0.47	467	
16868	Cab037712.1	Bo9g001120.	Rs481900	A09_000061545	C09_000055176	R09_024174247		0.47	0	0.43	433	
16869	Cab037707.1	Bo9g002160.	Rs480630	A09_000119790	C09_000115489	R09_024031814		0.33	0.07	0.43	6.4	
16870	Cab037705.1	Bo9g002170.	Rs480650	A09_000124331	C09_000121347	R09_024027047		0.7	0.03	0.5	14.6	
16871	Cab037704.2	Bo9g002180.	Rs480660	A09_000128273	C09_000123371	R09_024022859		0.4	0	0.47	467	
16872	Cab037703.1	Bo9g002190.	Rs480670	A09_000131536	C09_000128343	R09_024018674		0.47	0	0.57	567	
16873	Cab037702.1	Bo9g002200.	Rs480690	A09_000136462	C09_000131570	R09_024001387		0.27	0.17	0.4	2.39	
16874	Cab037700.1	Bo9g002210.	Rs480700	A09_000146986	C09_000135060	R09_023995770		0.43	0.17	0.57	3.38	
16875	Cab037699.1	Bo9g002220.	Rs480710	A09_000152423	C09_000141294	R09_023989114		0.4	0.07	0.37	5.42	
16876	Cab037698.1	Bo9g002230.	Rs480720	A09_000156748	C09_000146497	R09_023982867		0.33	0.07	0.43	6.4	
16877	Cab037697.2	Bo9g002240.	Rs480730	A09_000161138	C09_000149177	R09_023978801		0.5	0.33	0.47	1.4	
16878	Cab037696.1	Bo9g002250.	Rs480740	A09_000165816	C09_000152517	R09_023972869		0.37	0.33	0.5	1.5	
16879	Cab037695.1	Bo9g002260.	Rs480750	A09_000168790	C09_000156601	R09_023969912		0.6	0.2	0.57	2.82	
16880	Cab037694.1	Bo9g002280.	Rs480770	A09_000175592	C09_000166307	R09_023954339		0.47	0.03	0.53	15.5	
					~							
18073	Cab040595.1	Bo9g078340.	Rs478090	A09_017828874	C09_023996611	R09_019929597		0.2	0.1	0.43	4.29	
18074	Cab040596.1	Bo9g078350.	Rs478080	A09_017830691	C09_023998301	R09_019931250		0.37	0.03	0.43	12.6	
18075	Cab040598.1	Bo9g078360.	Rs478070	A09_017838283	C09_024007982	R09_019937875		0.37	0	0.43	433	
18076	Cab040600.1	Bo9g078380.	Rs478050	A09_017845775	C09_024022786	R09_019956890		0.3	0.07	0.4	5.91	
18077	Cab040601.1	Bo9g078390.	Rs478040	A09_017847399	C09_024024344	R09_019958575		0.3	0.1	0.4	3.96	
18078	Cab040610.1	Bo9g078420.	Rs478020	A09_017885828	C09_024062940	R09_019978476		0.33	0.23	0.6	2.56	
18079	Cab047791.1	Bo9g078450.	Rs477990	A09_017916938	C09_024077995	R09_020028930		0.27	0	0.53	533	
18081	Cab040615.1	Bo9g078470.	Rs477940	A09_017932739	C09_024110184	R09_020058770		0.2	0.27	0.47	1.74	
18082	Cab040614.1	Bo9g078480.	Rs576150	A09_017926827	C09_024112174	R09_019563181		0.3	0.1	0.4	3.96	
18083	Cab006207.1	Bo9g078520.	Rs446220	A06_018376540	C09_024150659	R08_025766370		1.47	0.1	0.1	0.99	
18084	Cab040618.2	Bo9g078550.	Rs477900	A09_017945197	C09_024175752	R09_020093128		1.1	0.1	0.43	4.29	
18085	Cab040620.1	Bo9g079590.	Rs477880	A09_018007178	C09_024246749	R09_020103392		0.37	0.07	0.1	1.48	
18086	Cab040623.1	Bo9g079650.	Rs466520	A09_018065217	C09_024344631	R09_020133606		0.57	0.1	0.47	4.62	
18087	Cab040626.1	Bo9g079660.	Rs469910	A09_018093854	C09_024385347	R09_020186208		0.43	0	0.57	567	
18088	Cab040628.1	Bo9g079670.	Rs469890	A09_018108961	C09_024395133	R09_020218837		0.3	0.2	0.53	2.65	
18089	Cab040629.1	Bo9g079680.	Rs469950	A09_018112650	C09_024397109	R09_019537142		0.3	0.1	0.1	0.99	
18090	Cab005464.4	Bo9g080720.	Rs466380	A09_020138195	C09_024454011	R09_016051056		0.33	0.4	0.07	0.17	
18091	Cab030172.2	Bo9g081800.	Rs250580	A05_015628801	C09_024777900	R05_020487346		7.17	9.1	0.47	0.05	
18092	Cab047671.1	Bo9g081860.	Rs466630	A09_020794685	C09_025047842	R09_016184376		0.2	1.27	0.1	0.08	
18093	Cab005561.1	Bo9g081890.	Rs367050	A09_021109739	C09_025277688	R07_026447884		0.4	0.4	0	0	
18094	Cab005565.2	Bo9g082030.	Rs466240	A09_021151395	C09_025513097	R09_015985586		1.07	0.37	0.07	0.18	
18095	Cab005566.1	Bo9g082070.	Rs466230	A09_021160306	C09_025540713	R09_015996423		8.73	0.93	0.03	0.04	
18096	Cab040961.1	Bo9g083130.	Rs473120	A01_018192218	C09_025627268	R09_012193486		0.27	0.4	0.07	0.17	
18097	BnaA10g0742C	Bo9g083150.	Rs020780	A10_008031737	C09_025660363	R01_011545731		0.4	0.27	0.03	0.12	



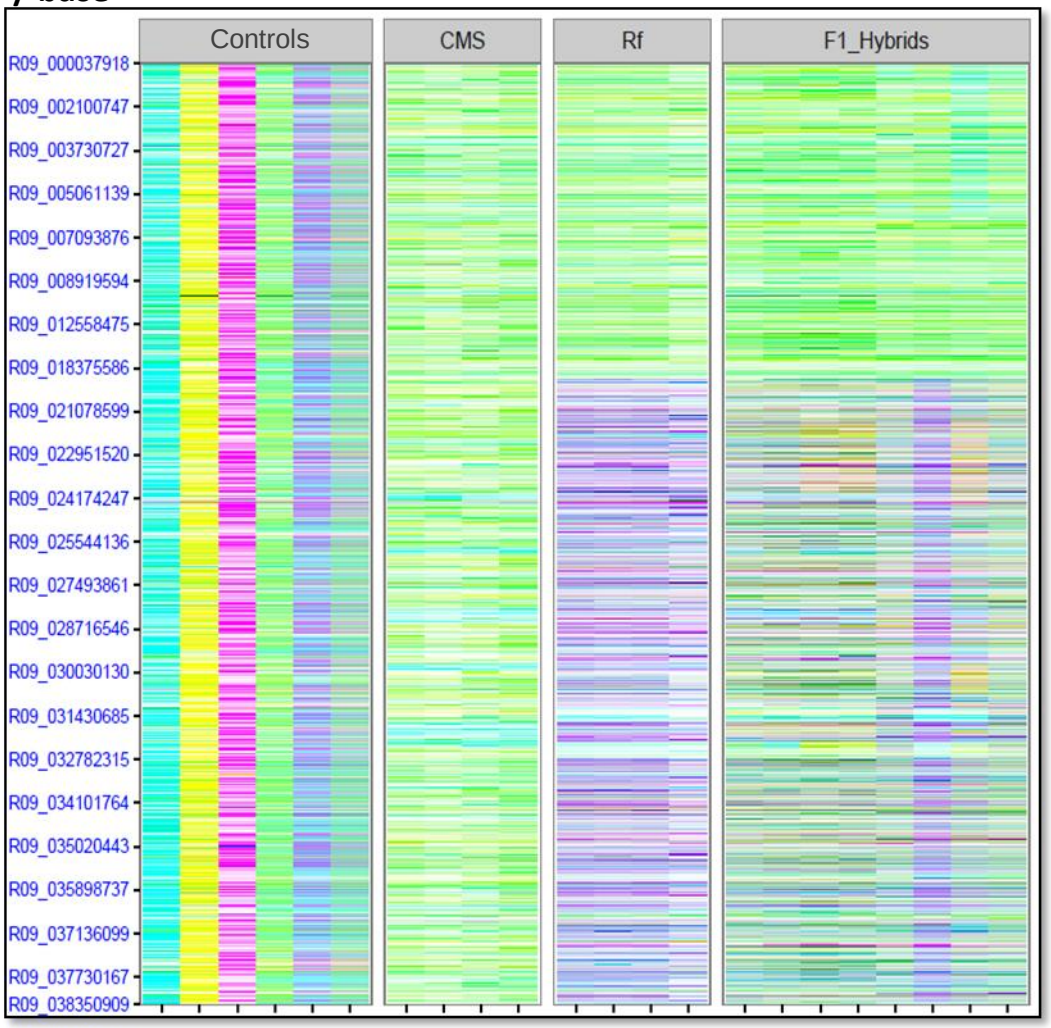
The introgression corresponds to the top half of chromosome C9

Visualization of a radish introgression in *B. napus*



Coordinate / base

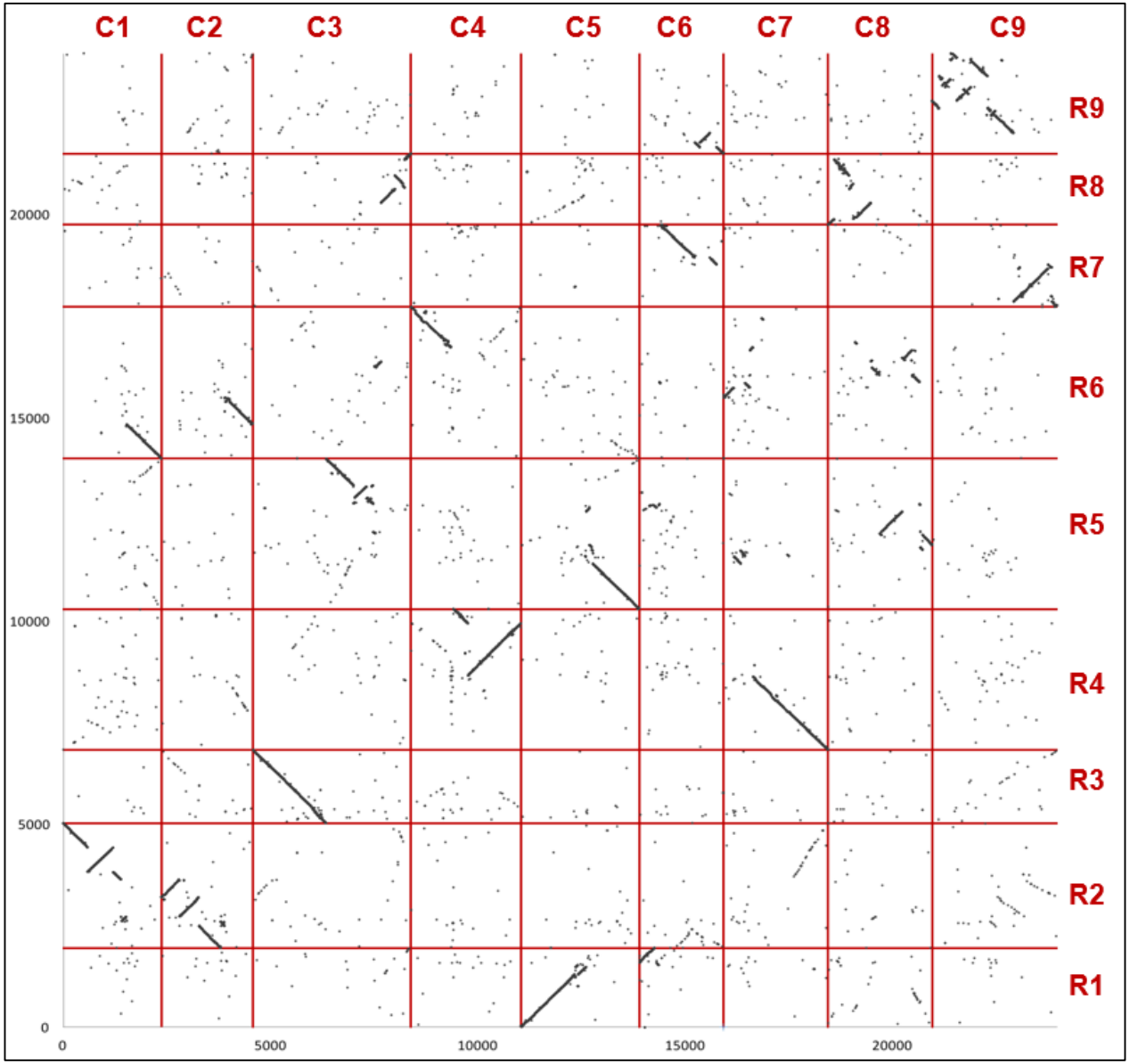
R9



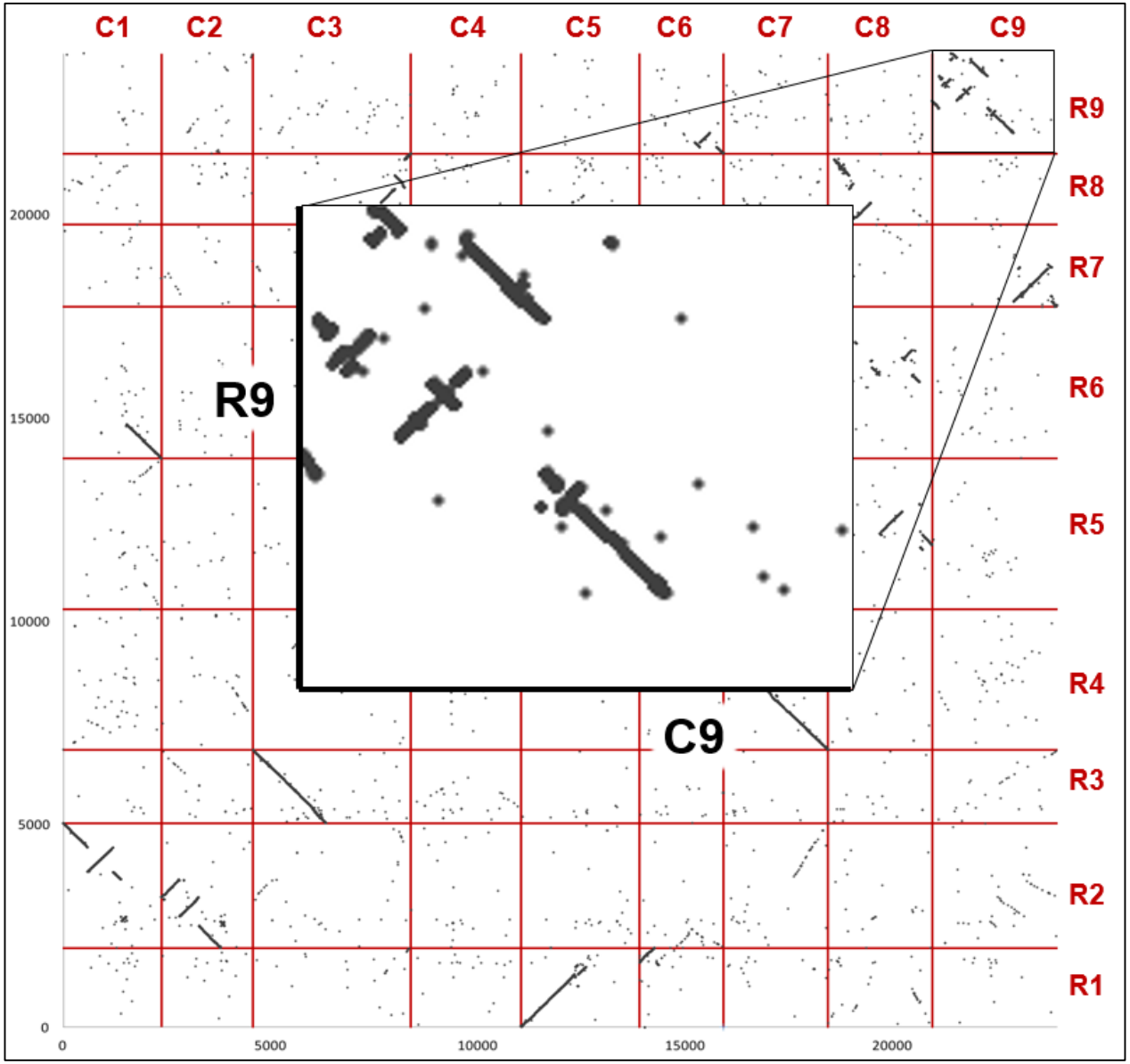
The introgression corresponds to the bottom half of chromosome R9

Some regions missing from the introgressions; difference between material from different breeders

Homoeology relationships between C and R genomes



Homoeology relationships between C and R genomes



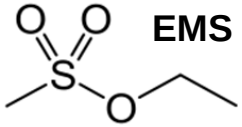
Genomics-led "traditional" approaches to crop genome engineering

Ian Bancroft

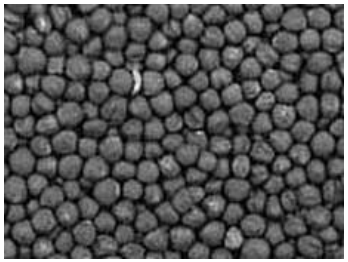
22nd May, 2019

- Introduction to rapeseed genomics
- Genome structural variation by homoeologous exchange
- Genome structural variation by alien introgression
- **Genome sequence and structural variation by gamma irradiation**

Genomics-led predictive mutation breeding



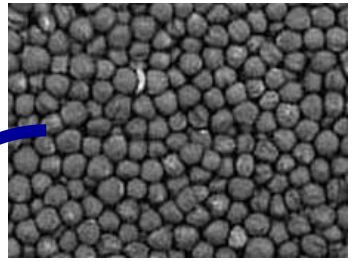
Treat seeds



Sow at high density

Grow and self plants from treatments with ~50% kill

Collect seed from M_1 plants



Grow

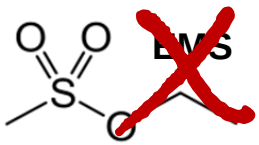
Make DNA from a leaf of each M_2 plant from which seeds will be collected

Panel of C/G to T/A mutations

Collect selfed seed from one M_2 plant per M_1 individual

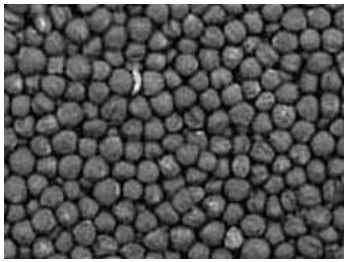
Seed stocks

Genomics-led predictive mutation breeding



γ Radiation

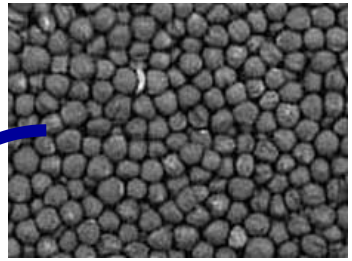
Treat seeds



Sow at high density

Grow and self plants from treatments with ~50% kill

Collect seed from M_1 plants



Grow

Make DNA from a leaf of each M_2 plant from which seeds will be collected

Collect selfed seed from one M_2 plant per M_1 individual

Point, frame-shift and deletion mutations

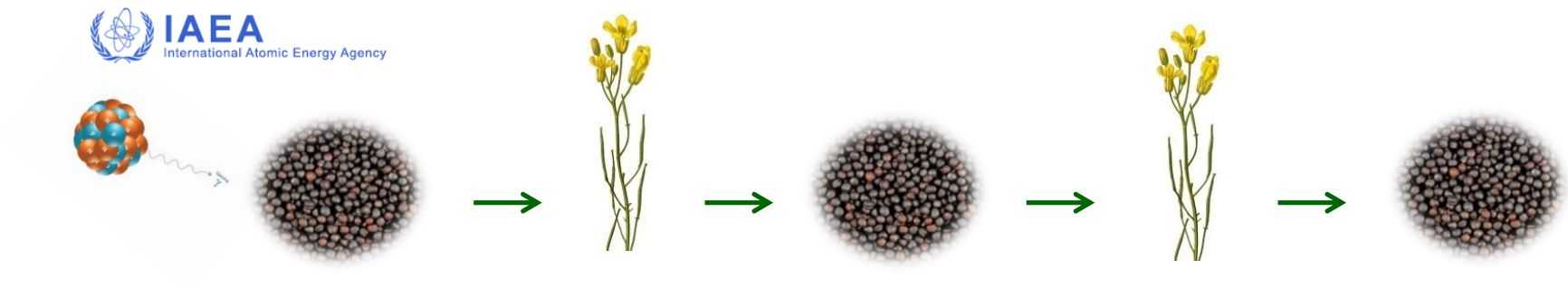
Panel of C/G to T/A mutations

Seed stocks

Genomics-led predictive mutation breeding

RIPR genome re-sequenced radiation mutagenesis panel

Control	250	500	750	1000	1250	1500	1750	2000



M_1
Gamma irradiated seeds

→

M_1
Grow plants

→

M_2
Collect selfed seed

→

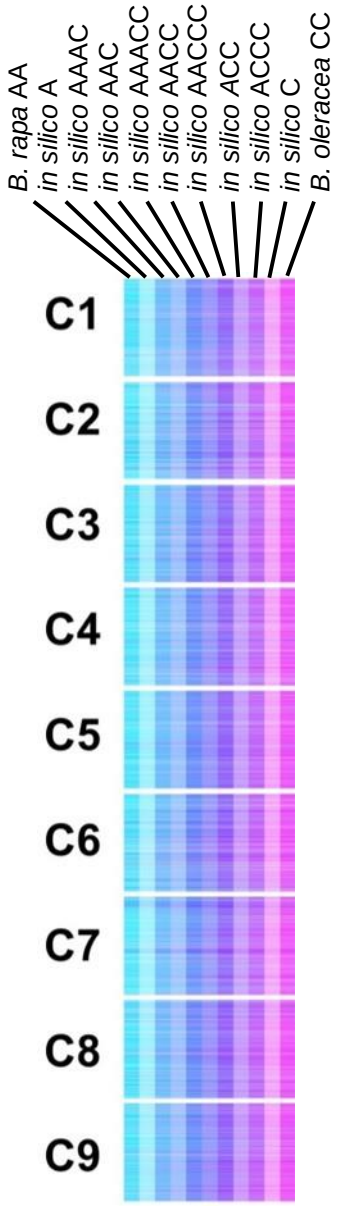
M_2
~800 selfed M_2 plants (one per M_1 individual)

Freeze leaves

→

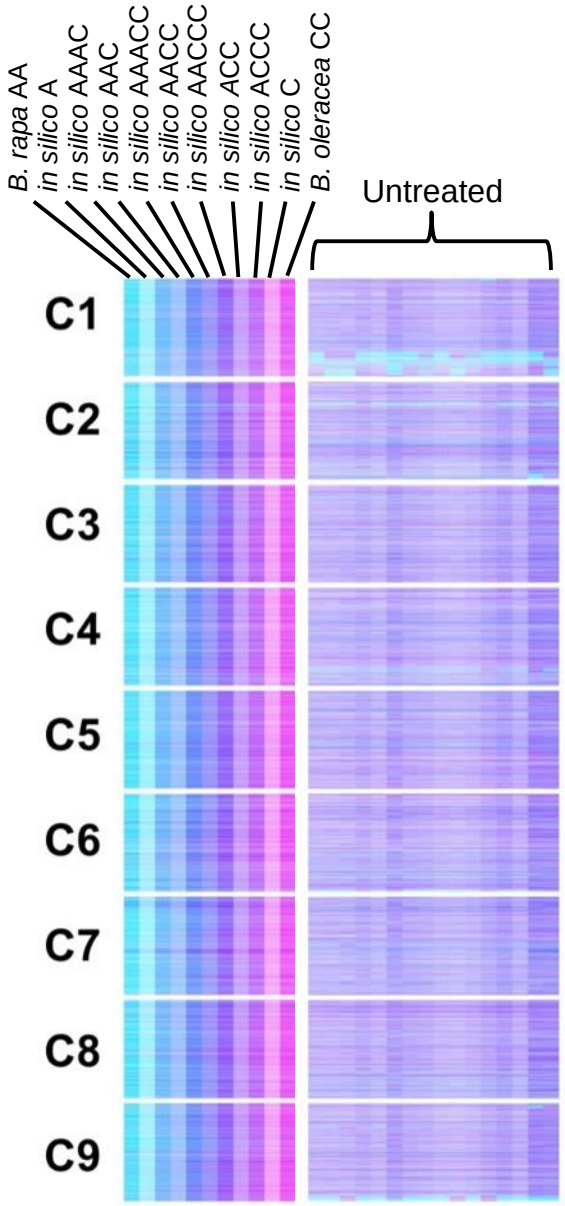
M_3
Seed stock

Genomics-led radiation mutagenesis



TDTP visualization based on Illumina mRNAseq data

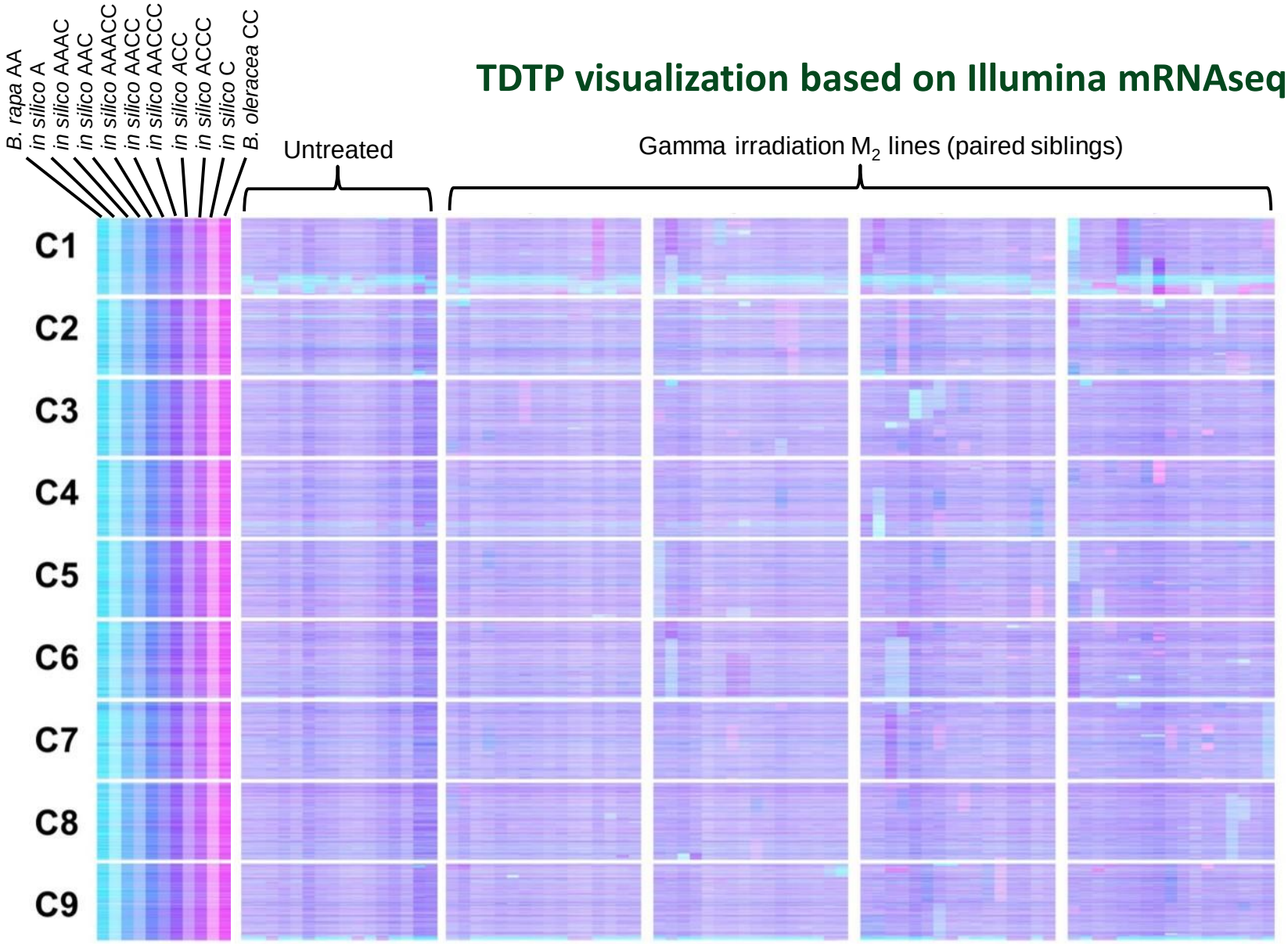
Genomics-led radiation mutagenesis



TDTP visualization based on Illumina mRNAseq data

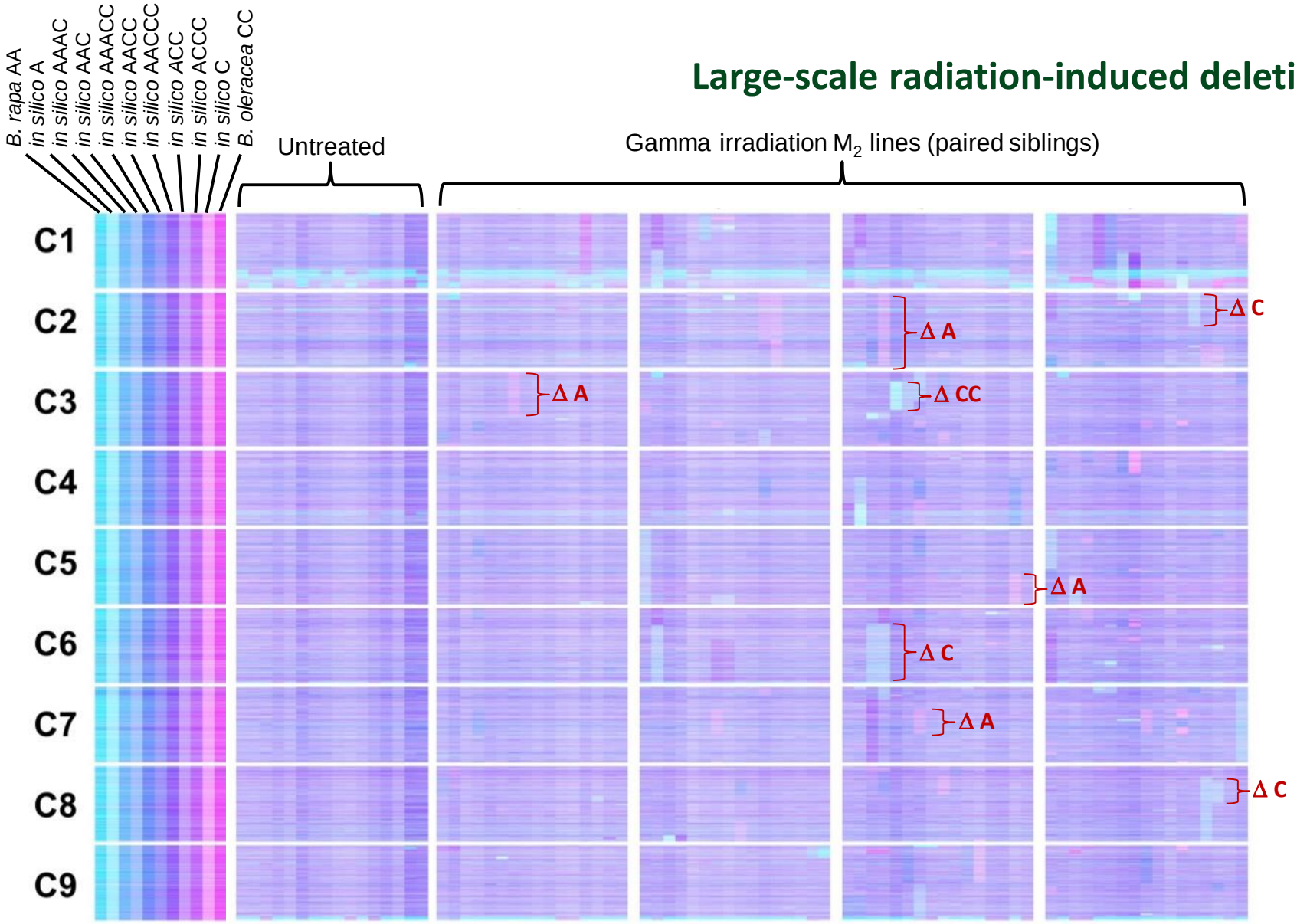
Genomics-led radiation mutagenesis

TDTP visualization based on Illumina mRNAseq data



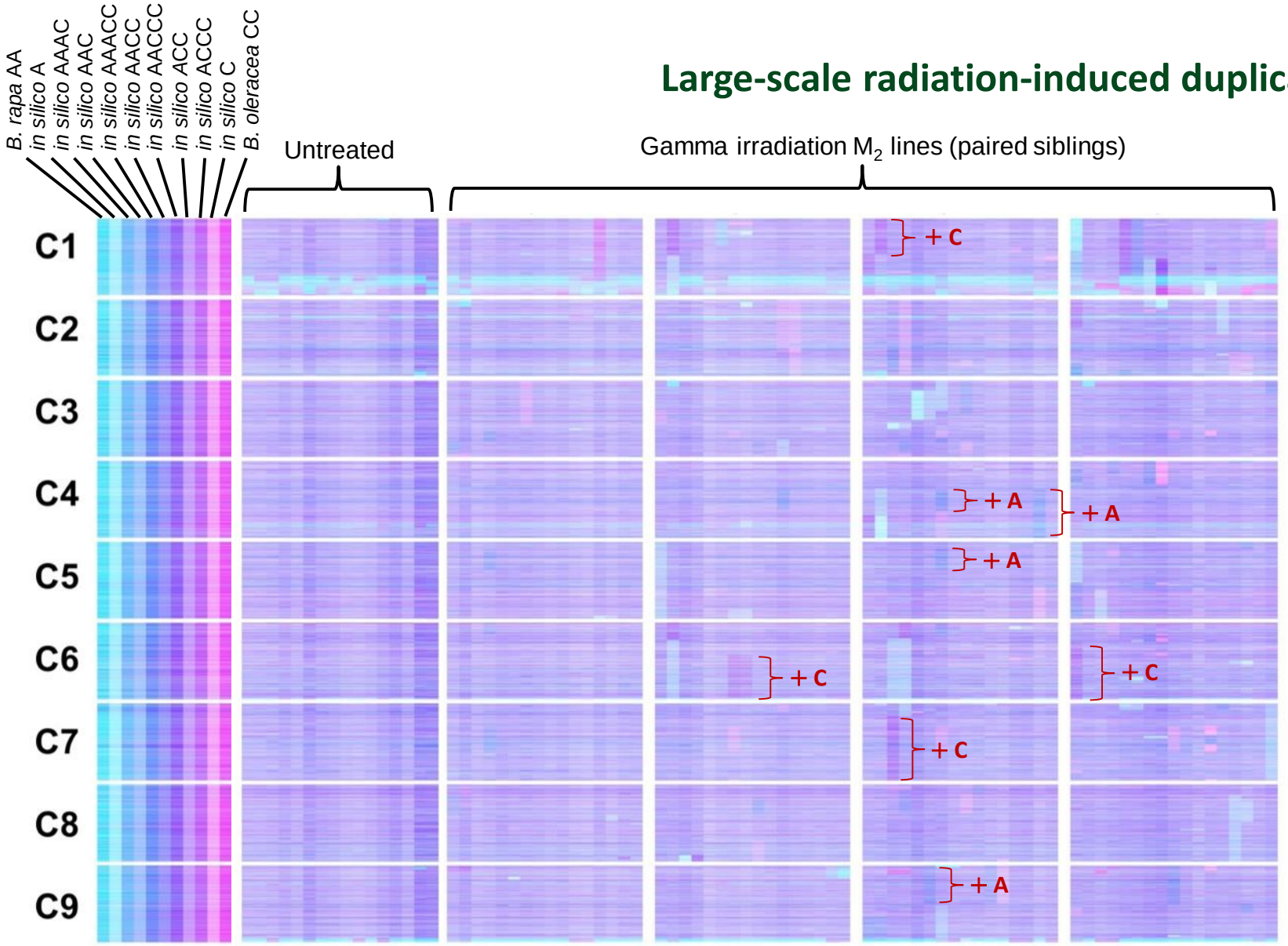
Genomics-led radiation mutagenesis

Large-scale radiation-induced deletions



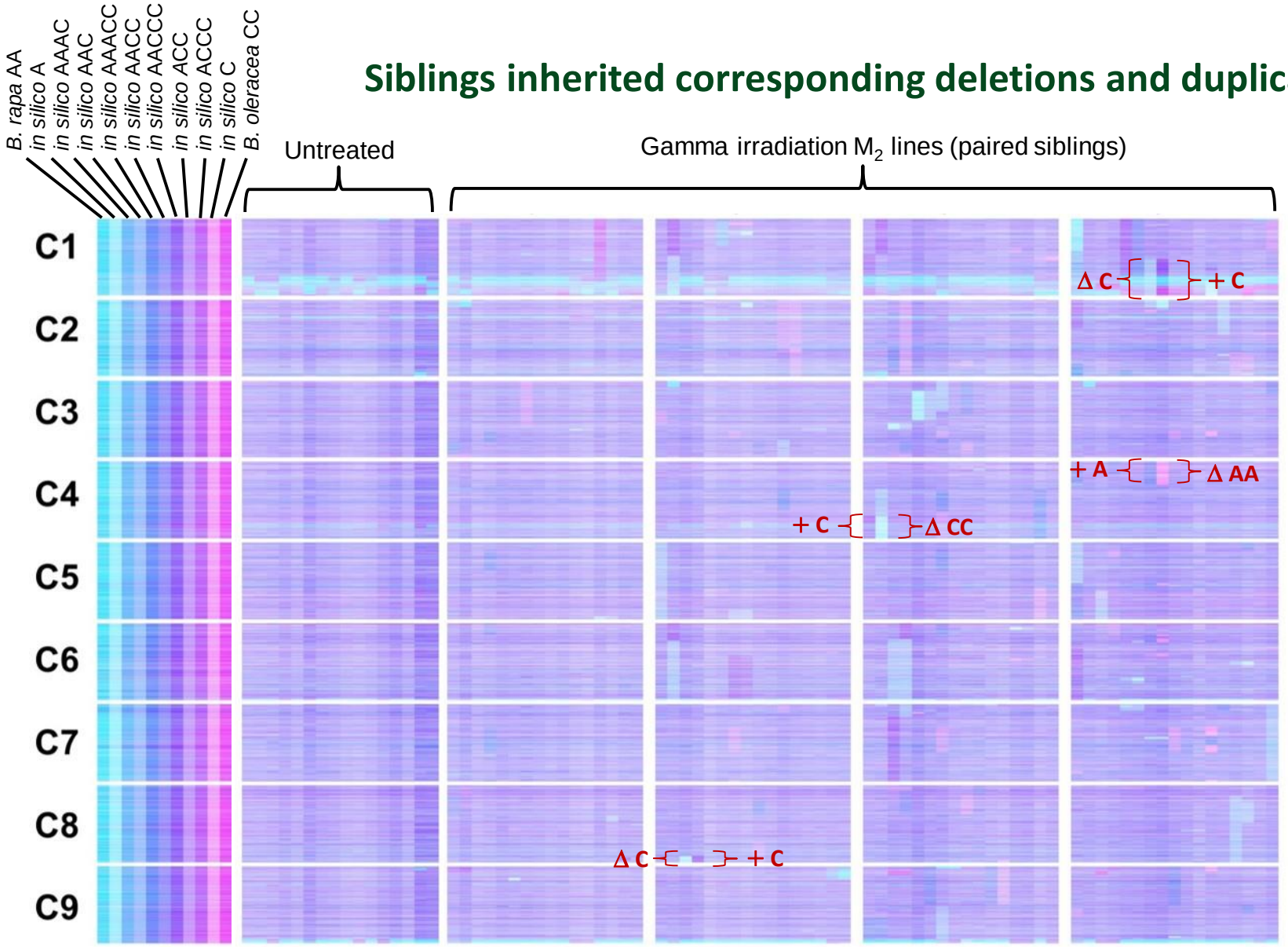
Genomics-led radiation mutagenesis

Large-scale radiation-induced duplications



Genomics-led radiation mutagenesis

Siblings inherited corresponding deletions and duplications



Genomics-led radiation mutagenesis

First 64 lines: Illumina mRNAseq with variants detected using VarScan

~24M reads gives at least 7x coverage for 35,000 genes

~9,000 transcriptome SNPs per line:

A -> C/M	12,574
A -> G/R	36,510
A -> T/W	15,250
C -> A/M	13,530
C -> G/S	11,592
C -> T/Y	46,986
G -> A/R	42,079
G -> C/S	13,836
G -> T/K	15,366
T -> A/W	13,707
T -> C/Y	49,035
T -> G/K	13,680

~360 transcriptome frame-shift InDels per line

Genomics-led radiation mutagenesis

Genome re-sequenced radiation mutagenesis panel

Background genotype: Maplus

~2000 M_2 lines available, mostly gamma radiation but some FN.

Genome re-sequencing (~12x) completed; variant scoring protocol under development:

- **16 individual untreated Maplus plants**
- **584 M_2 lines genome re-sequenced at BGI**

GDTP visualization based on genome re-sequencing

Analysed for large-scale structural variation affecting genes of interest using GDTPs

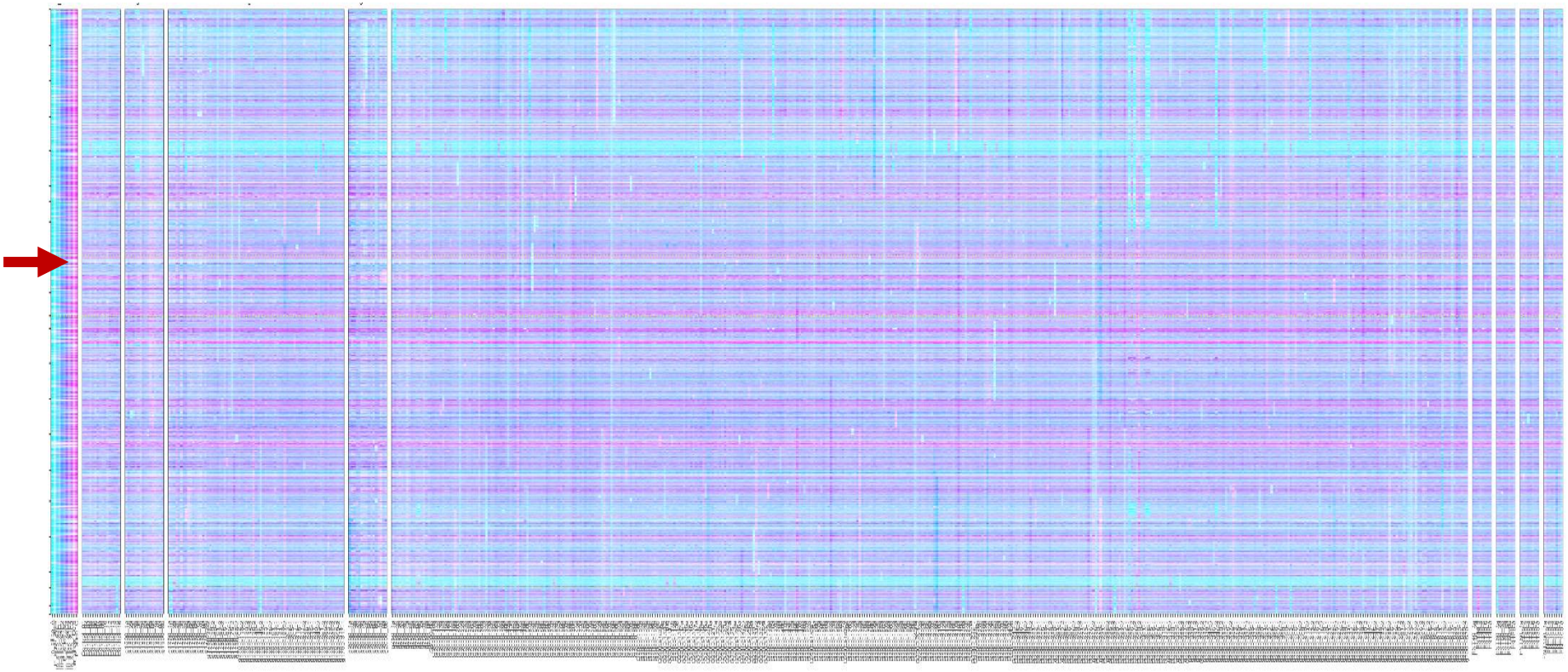
Analysed for frame-shift InDels, nonsense and mis-sense mutations using VarScan

Genomics-led radiation mutagenesis

Identifying gene dosage variants using the RIPR radiation mutagenesis panel

***e.g. VTE4.C2a/VTE4.C2b* deletion of potential tandem gene pair**

Chromosome C2

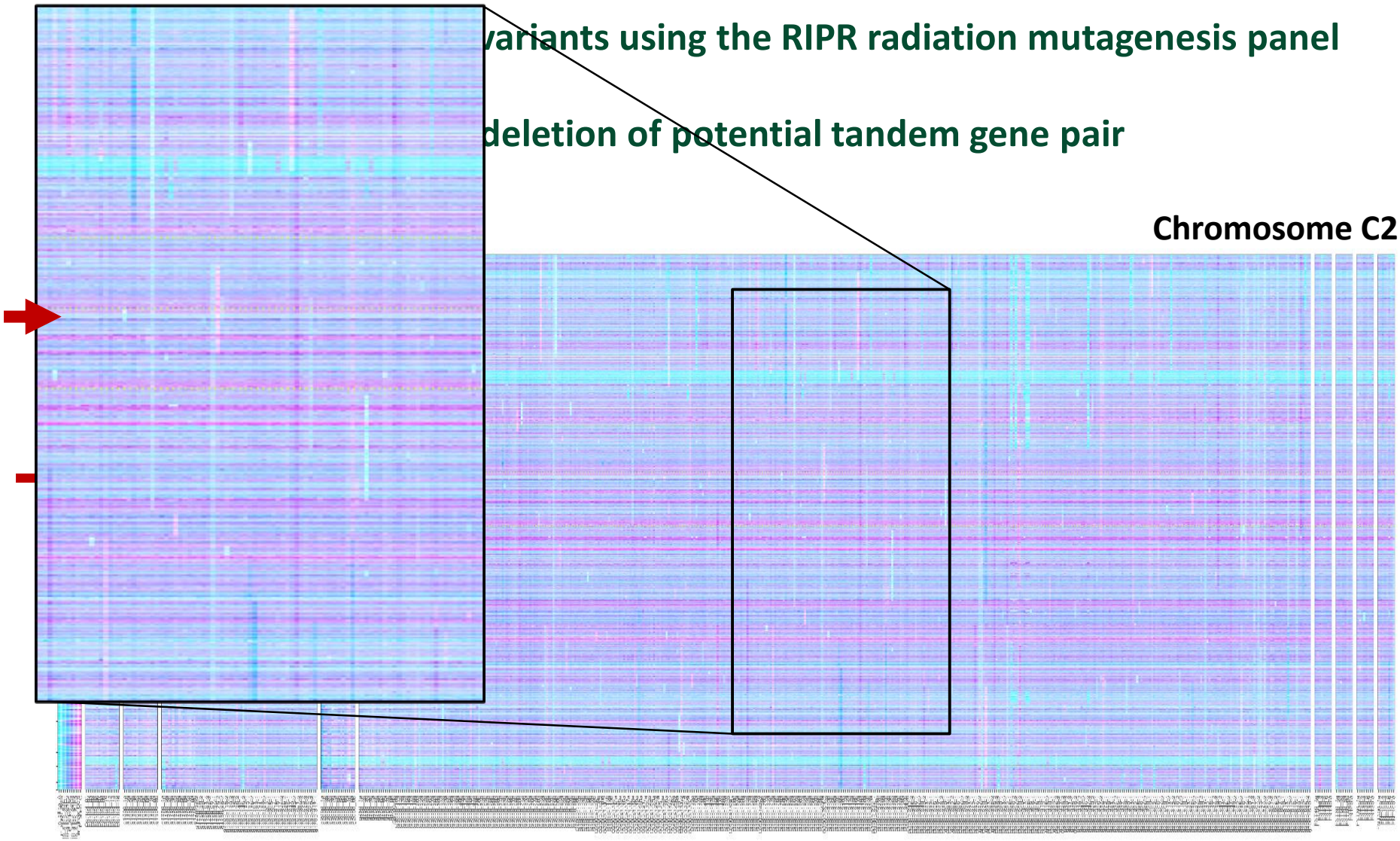


All 600 TDTP visualization (C02) based on 12x Illumina genome re-sequencing

Genomics-led radiation mutagenesis

variants using the RIPR radiation mutagenesis panel

deletion of potential tandem gene pair

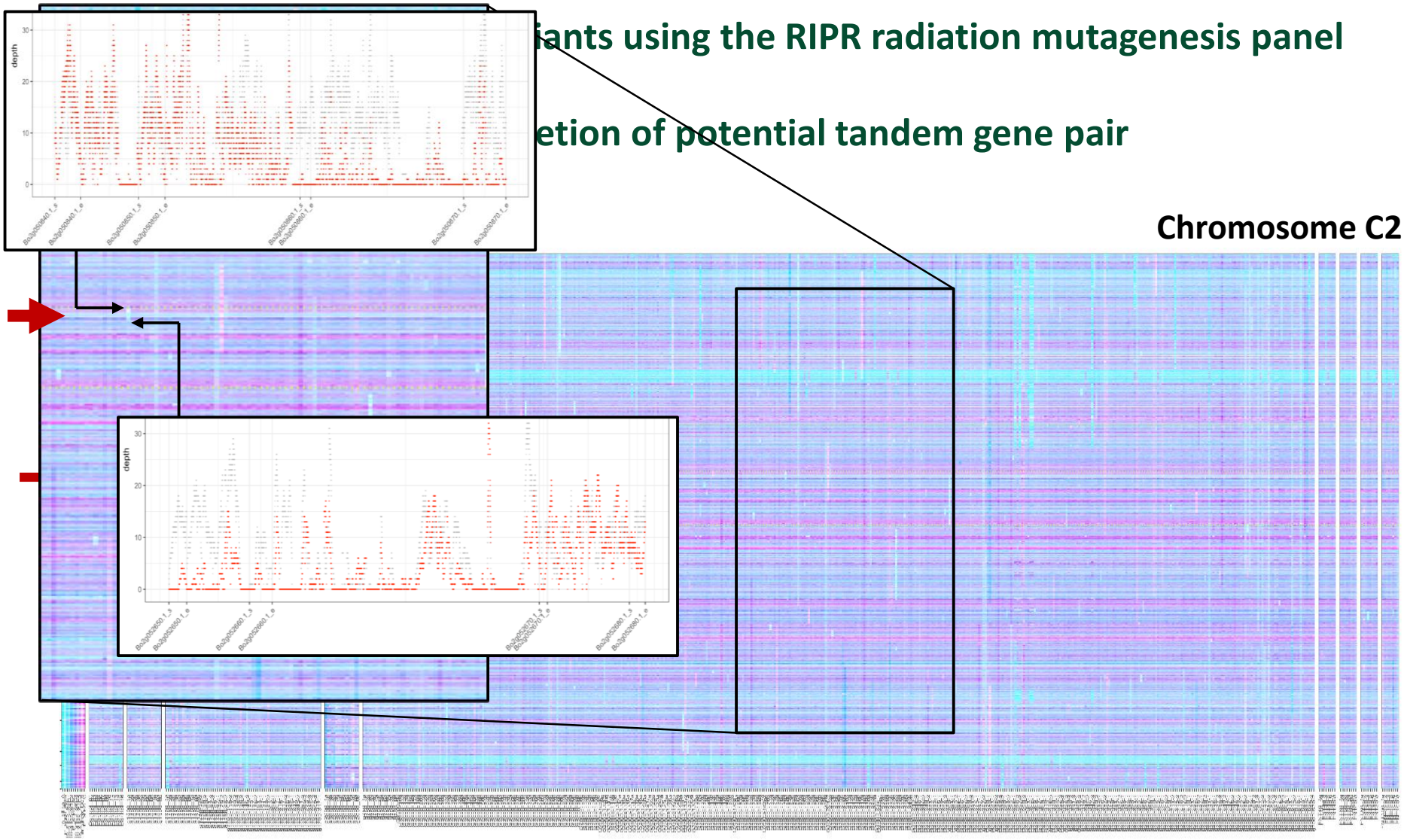


All 600 TDTP visualization (C02) based on 12x Illumina genome re-sequencing

Genomics-led radiation mutagenesis

iants using the RIPR radiation mutagenesis panel

etion of potential tandem gene pair



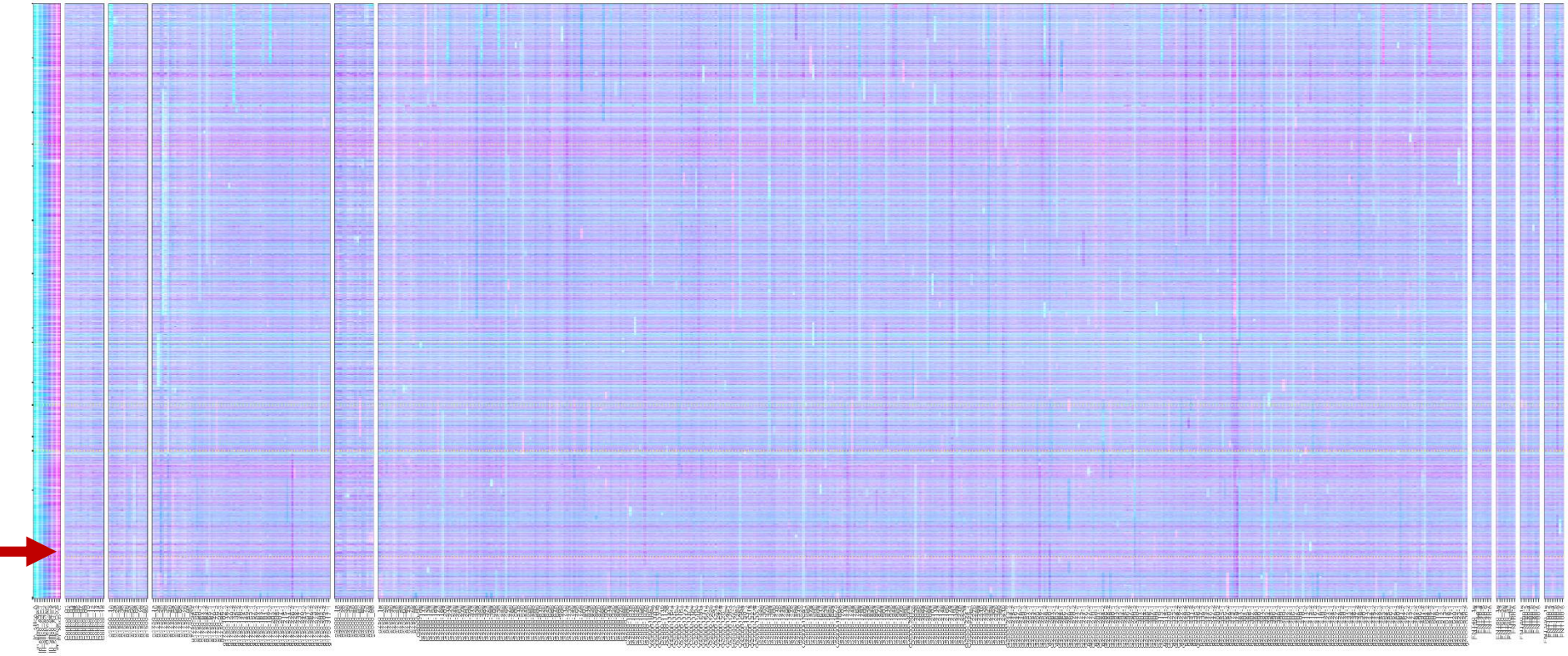
All 600 TDTP visualization (C02) based on 12x Illumina genome re-sequencing

Genomics-led radiation mutagenesis

Identifying gene dosage variants using the RIPR radiation mutagenesis panel

e.g. FAE1.C3 deletion

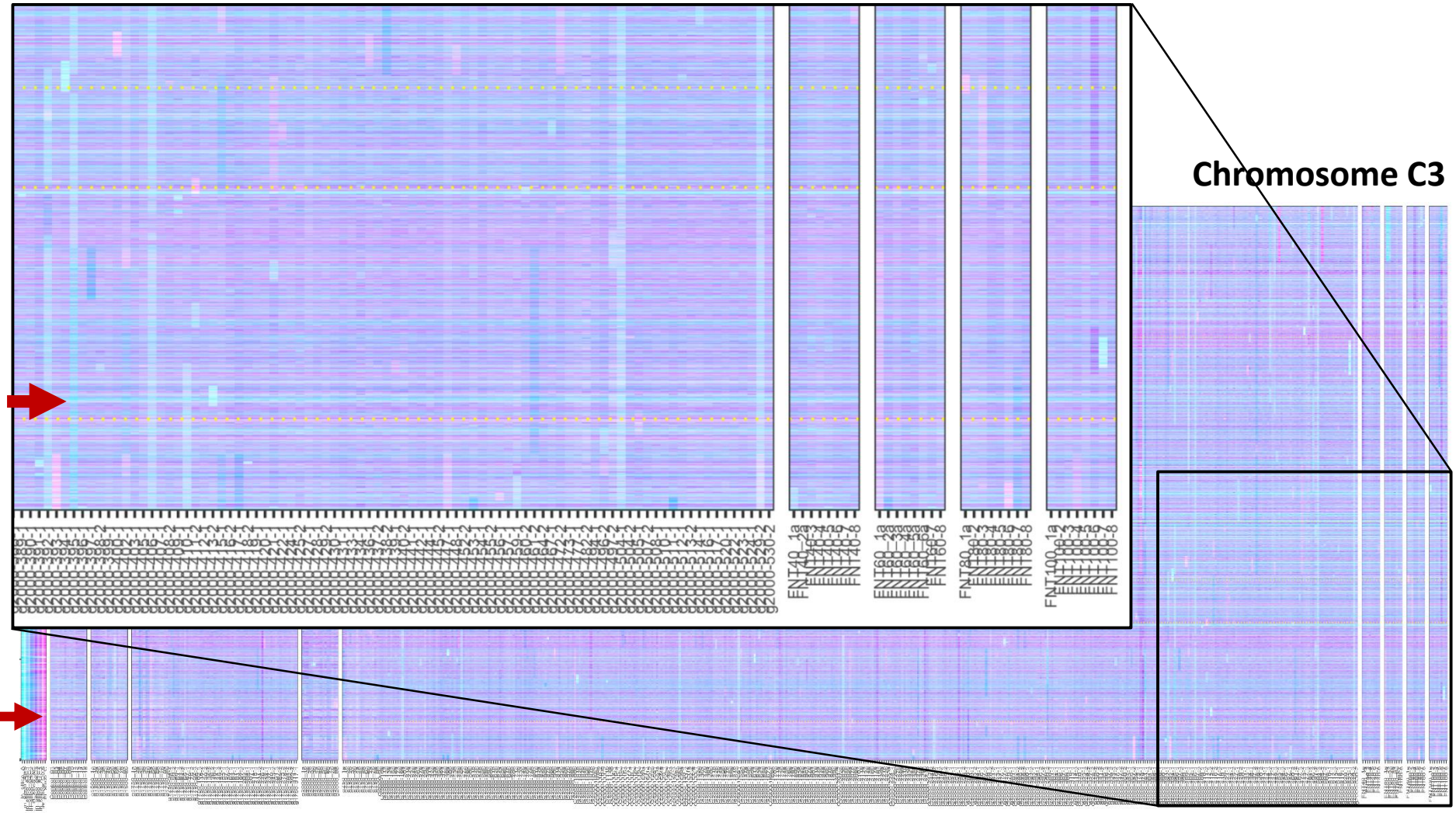
Chromosome C3



All 600 TDTP visualization (C03) based on 12x Illumina genome re-sequencing

Genomics-led radiation mutagenesis

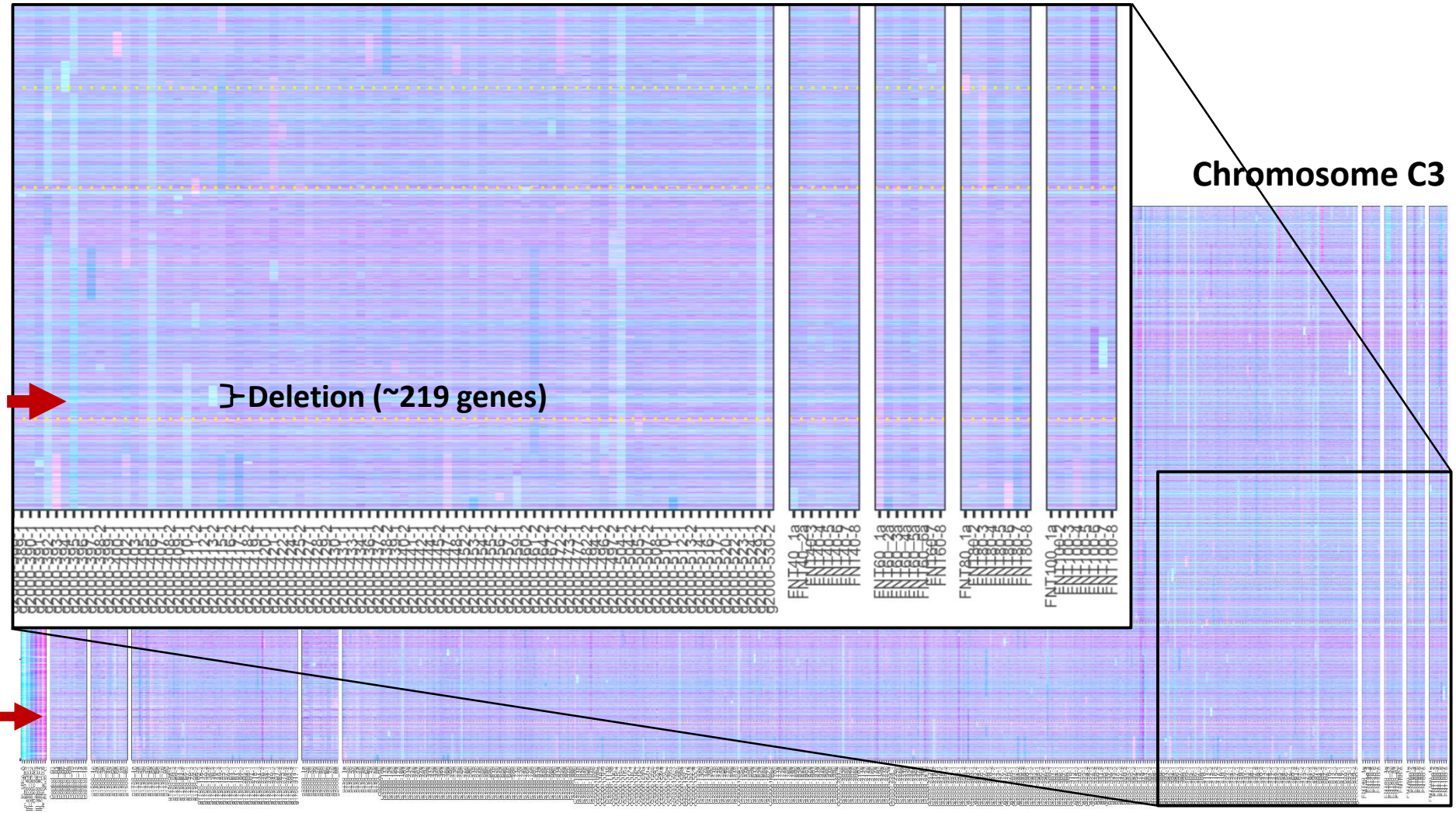
Identifying gene dosage variants using the RIPR radiation mutagenesis panel



All 600 TDTP visualization (C03) based on 12x Illumina genome re-sequencing

Genomics-led radiation mutagenesis

Identifying gene dosage variants using the RIPR radiation mutagenesis panel



All 600 TDTP visualization (C03) based on 12x Illumina genome re-sequencing

Genomics-led radiation mutagenesis

Identification of differential representation in spreadsheets, e.g. *FAE1.C3*

RPKM600_C_IB.xlsx - Excel

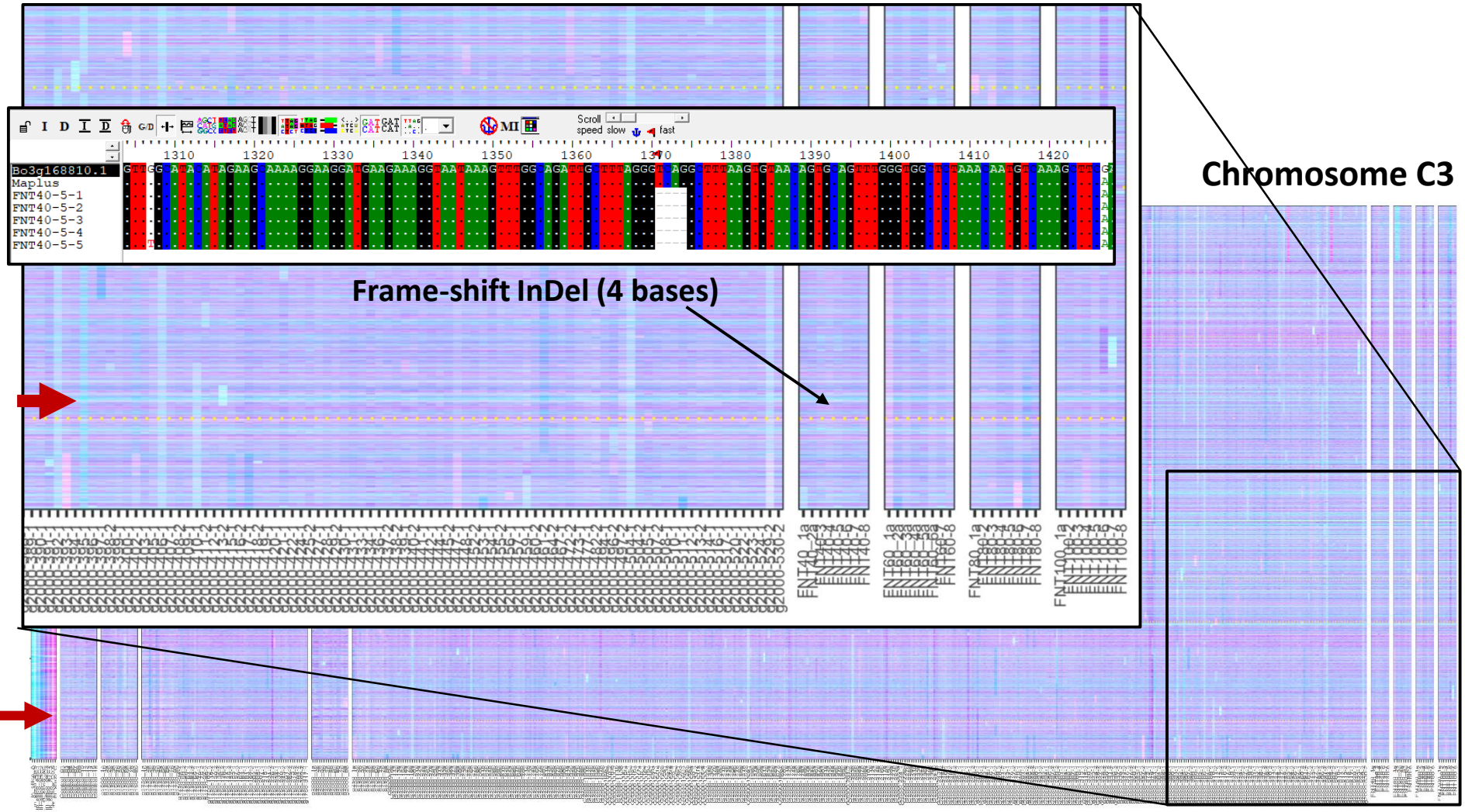
	cds_name		interpolate_loc	AT_id	g2000-413	g2000-454
22506	Bo3g168610.1		C03_064262308_064262802		0	0.63
22507	Bo3g168620.1		C03_064262886_064266637		1.93	3.11
22508	Bo3g168630.1		C03_064267404_064267601		0.02	0.11
22509	Bo3g168640.1		C03_064270371_064271453		0.04	4.73
22510	Bo3g168650.1		C03_064272106_064272690		0.03	1.73
22511	Bo3g168660.1		C03_064273522_064274677	AT4G34700.1	0	0.45
22512	Bo3g168670.1		C03_064276015_064276880	AT4G34680.2	0	0.86
22514	Bo3g168690.1		C03_064290679_064292077	AT4G34670.1	0	1.1
22515	Bo3g168700.1		C03_064295392_064298223	AT4G34660.3	0.03	0.57
22516	Bo3g168710.1		C03_064298911_064302093	AT4G34640.1	1.28	3.86
22517	Bo3g168720.1		C03_064304162_064304807		4.15	10.92
22518	Bo3g168730.1		C03_064310376_064311536		0.75	0.65
22519	Bo3g168740.1		C03_064311740_064313379		0.94	1.23
22520	Bo3g168750.1		C03_064321277_064321950		1.5	0.93
22523	Bo3g168770.1		C03_064335926_064339338	AT4G34580.1	10.24	28.46
22524	Bo3g168780.1		C03_064343360_064343563		0	0.19
22525	Bo3g168790.1		C03_064343594_064344782		1.29	8.43
22526	Bo3g168800.1		C03_064348567_064349575	AT4G34530.1	0.89	2.88
22527	Bo3g168810.1	FAE1.C3	C03_064359517_064361037	AT4G34520.1	0	1.33
22533	Bo3g168820.1		C03_064363039_064363509		0.01	0.15
22534	Bo3g168830.1		C03_064364583_064365190		12.24	26.8
22[illegible]					2.88	12.93
22[illegible]					2.04	3.6
22[illegible]					1.86	2.62
22539	Bo3g168880.1		C03_064393069_064396563	AT4G34440.1	1.1	1.2
22540	Bo3g168890.1		C03_064400607_064402433		1.69	1.85
22541	Bo3g168900.1		C03_064403378_064407368	AT4G34430.2	1.75	1.54
22542	BnaC03g66120D		C03_064411295_064410915	AT4G34419.1	0.73	1.36
22543	Bo3g168910.1		C03_064425838_064426626	AT4G34410.1	7.48	6.2
22544	Bo3g168920.1		C03_064450719_064455768	AT4G34390.1	5.37	4.7
22545	Bo3g168930.1		C03_064463165_064463425	AT4G34380.1	2.26	2.73
22546	Bo3g168940.1		C03_064471476_064473823	AT4G34350.1	2.88	3.03
22548	Bo3g168950.1		C03_064484015_064485707	AT4G34320.1	4.54	4.6
22550	Bo3g168960.1		C03_064506811_064507056	AT4G34300.1	1.99	4.35

Putative single gene duplication

Bottom of 219-gene deletion

Genomics-led radiation mutagenesis

Identifying frame-shift mutations using the RIPR radiation mutagenesis panel



All 600 TDTP visualization (C03) based on 12x Illumina genome re-sequencing

Acknowledgements

BBSRC BB/L027844/1 EvoGenapus; BBSRC BB/L002124/1 RIPR; BBSRC BB/L011751/1 CGAT

National Capability in Genomics (BB/J010375/1) Platforms and Pipelines Group at The Genome Analysis Centre

High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (Wellcome Trust 090532/Z/09/Z)