Understanding and exploiting the dynamic *Brassica napus* genome

Rod Snowdon
Department of Plant Breeding, Justus Liebig University, Giessen, Germany

Applying genomics for rapeseed breeding

Experimental populations

Phenotype data

Controlled climate phenotyping

Multi-environment field evaluations

Large-scale genotype datasets

Markers, QTL, genes

Multidimensional *omics* data

Breeding populations

Trait proxies, models for gene pool development

Improved hybrids

New diversity

Genomic prediction

Yield Gain

Wilhelm Johanssen

"No breeding success without genetically determined diversity"

Johanssen (1903)

Om arvelighed i samfund og i rene linier
– *On heredity in populations and pure lines*

The *Brassica napus* diversity paradox?

Diploid progenitor species

Strong allopolyploidisation bottleneck

***De novo* allopolyploid**

Few species founders

Seed quality selection

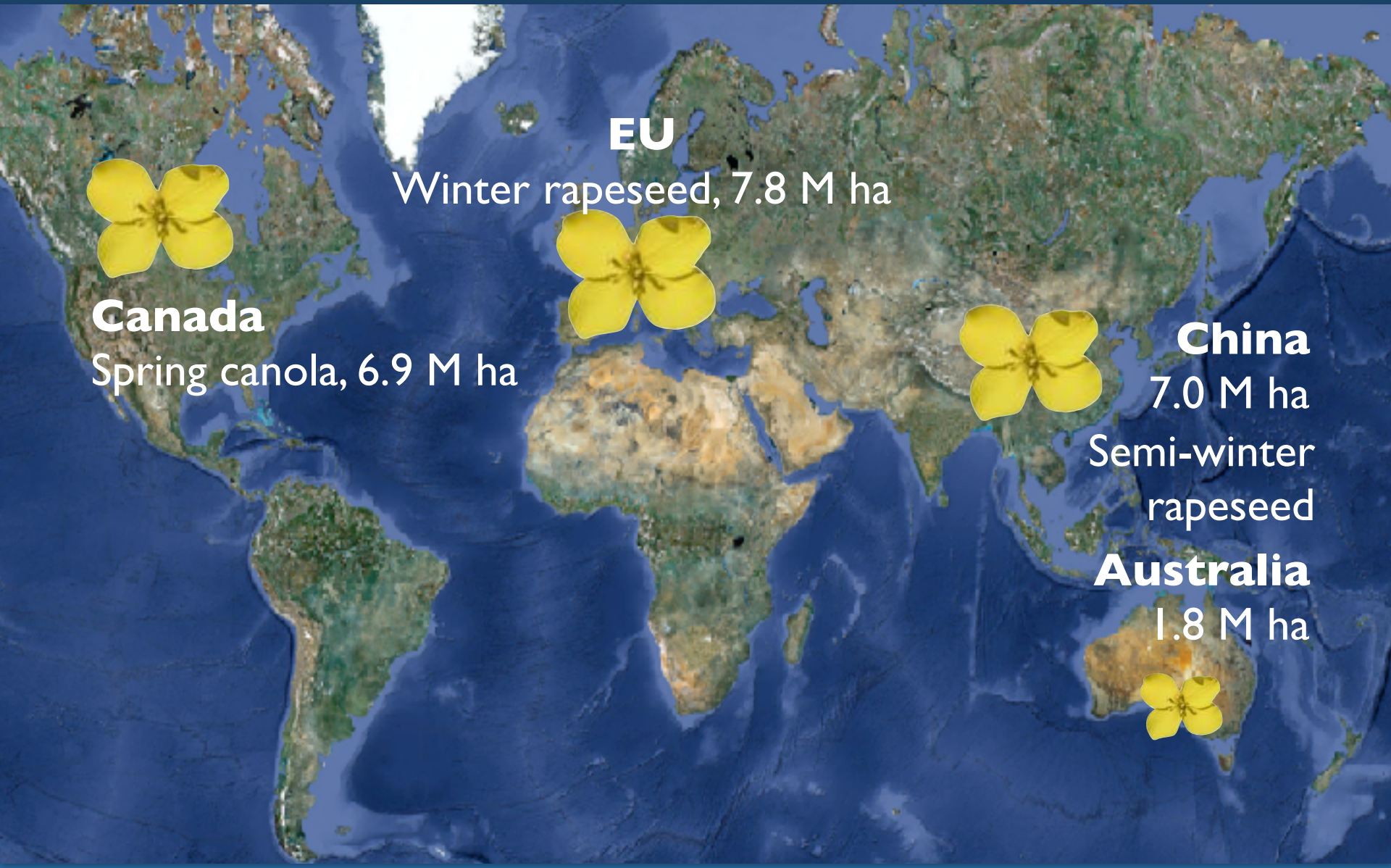
Extreme breeding bottlenecks

Extreme erosion of genetic diversity essential for breeding

However: Unexpected **adaptive capacity** and **breeding success**

Rod Snowdon, JLU Giessen

Brassica napus: Rapid ecogeographic diversification



Rags to riches in a few decades: A “Cinderella” crop

Today the world’s No. 2 oilseed crop

Healthiest vegetable oil for human nutrition, equally suitable as a biofuel

High-quality, protein-rich extraction meal for livestock feed

Vital component in cereal crop rotations

Images: UFOP, Canola Council, NPZ Lembke

How did adaptive diversity arise in *Brassica napus*?

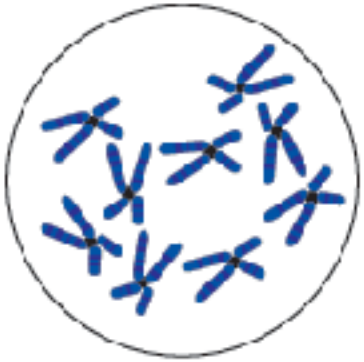
Lessons from *B. napus* (genome) evolution

- How did different forms achieve sufficient *de novo* diversity (despite extreme bottlenecks) to adapt to completely new environments and become a hugely successful global crop?

Breeding applications

- Can we **generate, identify and exploit new diversity** for adaptation to new challenges (e.g. climate change, disease, constraints on fertiliser or chemical inputs)?

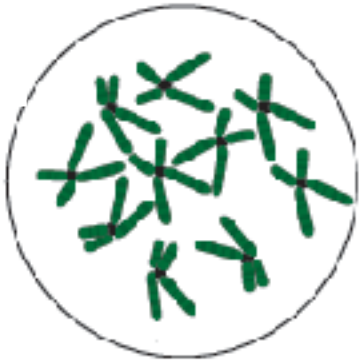
Tracing the origins of *Brassica napus* diversity



Brassica oleracea
2n = 2x = 18
Genome CC

Mediterranean cabbages and kales

Brassica napus
2n = 4x = 38
Genome AACC



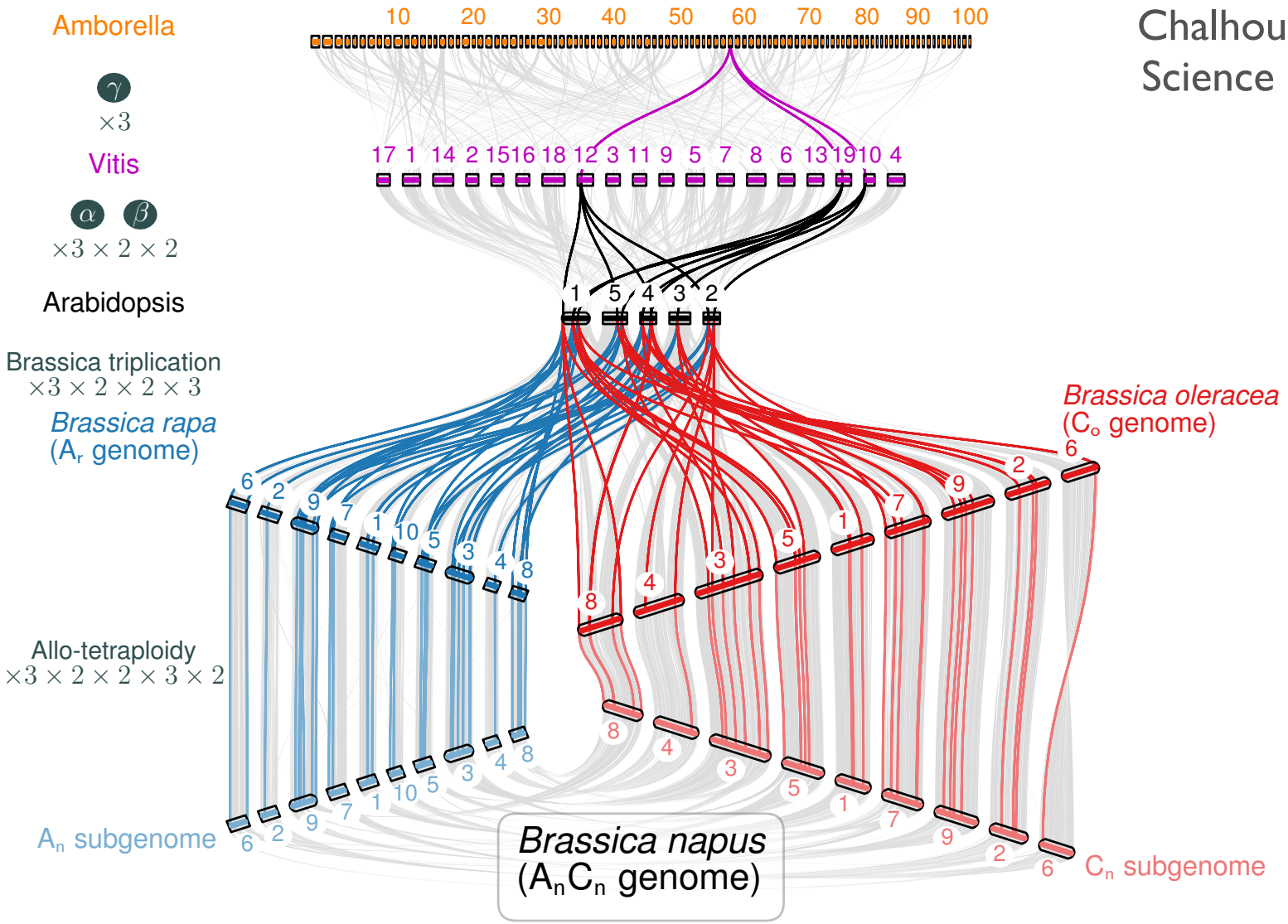
Brassica rapa
2n = 2x = 20
Genome AA

Asian cabbages, Sarson and turnips

Canola/rapeseed, kale & swede/rutabaga

Originated under cultivation, just a few thousand years ago, from few founding hybridisation events – wild forms unknown

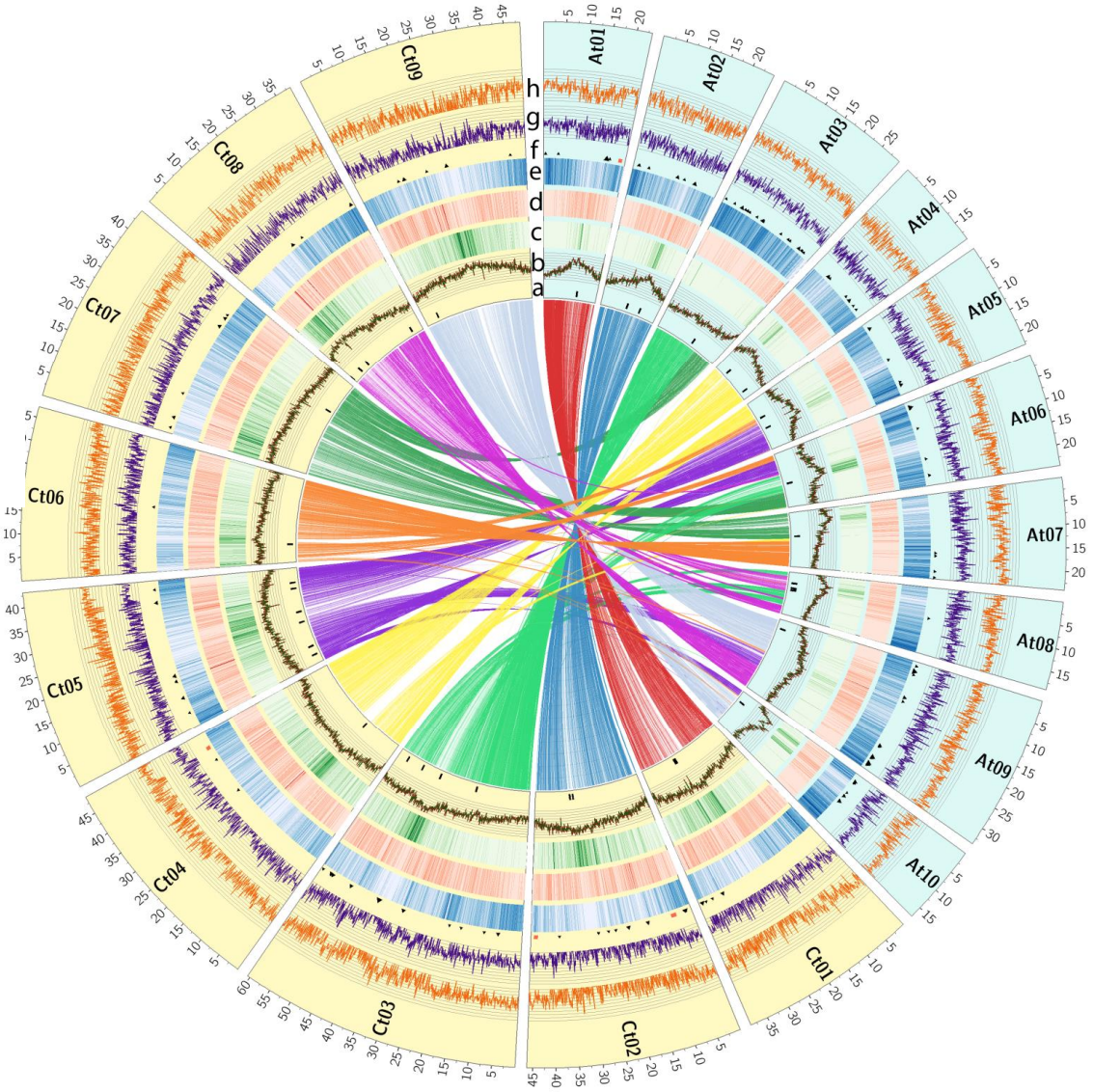
Reunification of 2 similar, but reorganised genomes



Chalhoub et al.
Science (2014)

Extensive inter-subgenomic "homoeology"

Subgenome C
n = 9
540 Mbp

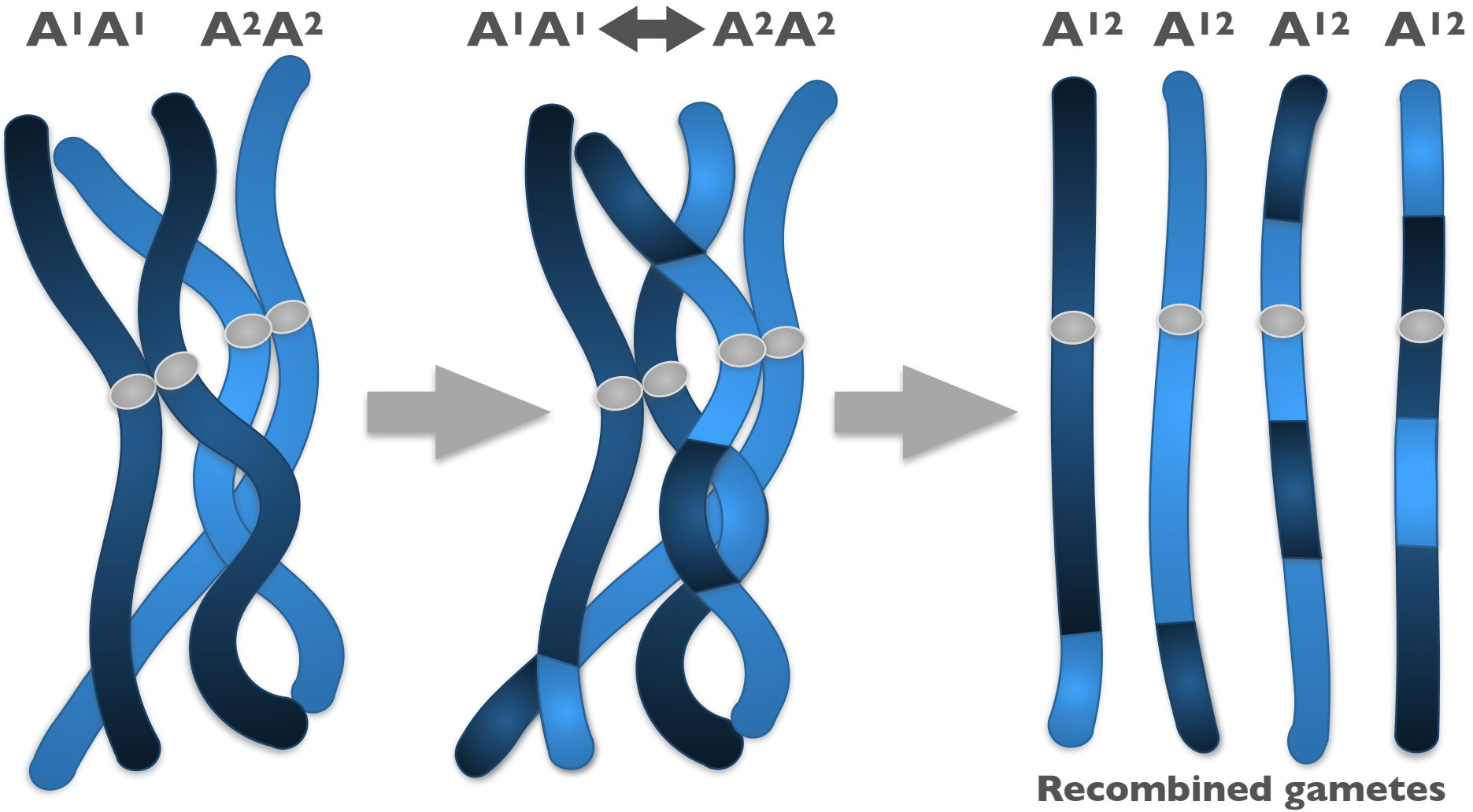


Subgenome A
n = 10
314 Mbp

Chalhoub et al.
Science (2014)

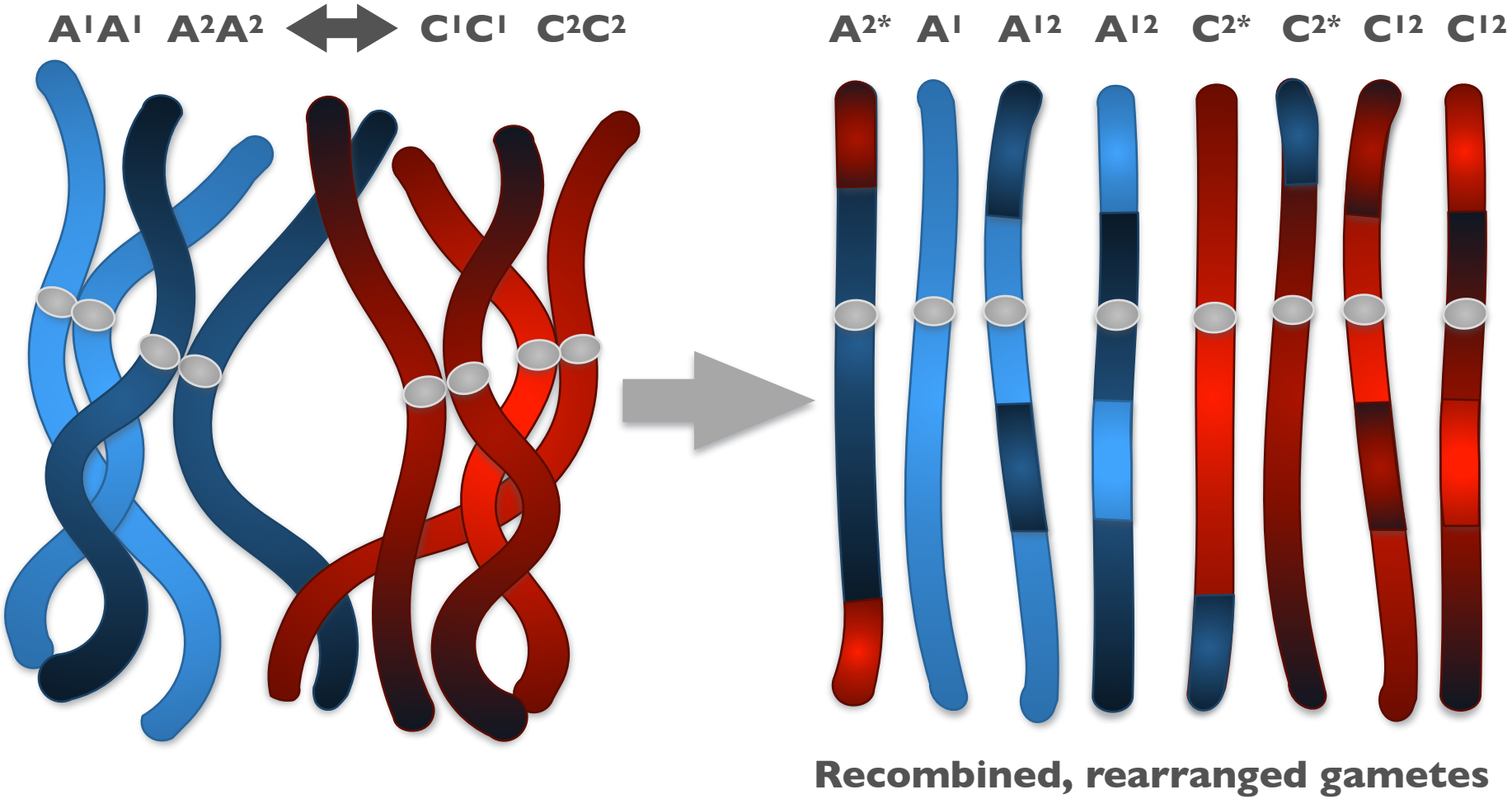
Chromosome exchanges in a simple diploid genome

Normal pairing and crossovers between homologous chromosomes



Meiotic exchanges in a recent allopolyploid

Illicit pairing and crossovers between homoeologous chromosomes



Post-polyploidisation genome evolution

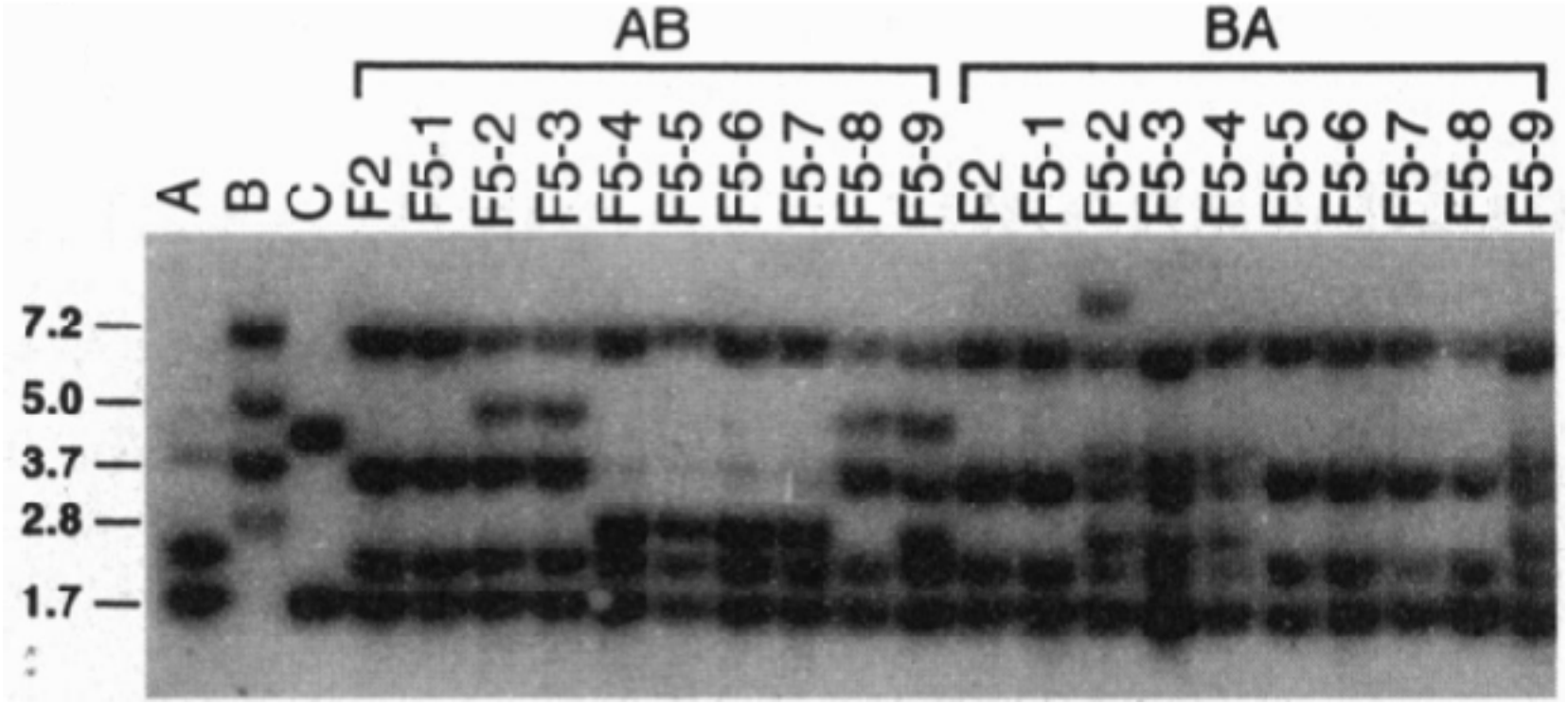
Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution

(hybridization/restriction fragment length polymorphism/molecular evolution/cytoplasmic–nuclear interaction)

KEMING SONG*, PING LU, KELIANG TANG*, AND THOMAS C. OSBORN†

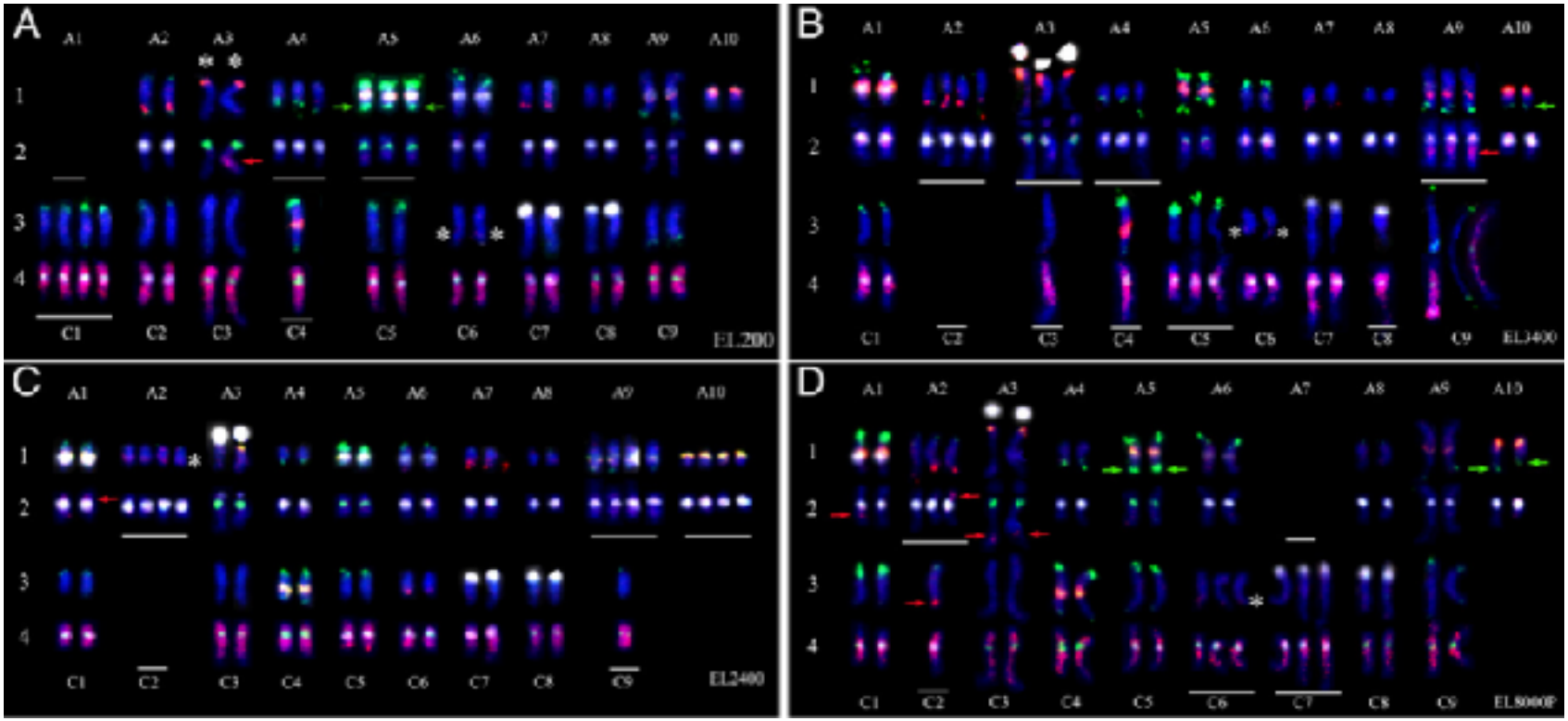
Department of Agronomy, University of Wisconsin, Madison, WI 53706-1597

Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 7719–7723, August 1995



"Genome shuffling" in synthetic *B. napus*

- *De novo* allopolyploidisation in *B. napus* induces massive genome rearrangements, particularly homoeologous non-reciprocal translocations (HNRT)
- e.g. Gaeta et al. *Plant Cell* (2007), Nicolas et al. *Plant Cell* (2009), Szadkowski et al. *New Phytol* (2010), Grandont et al. *Plant Cell* (2014)

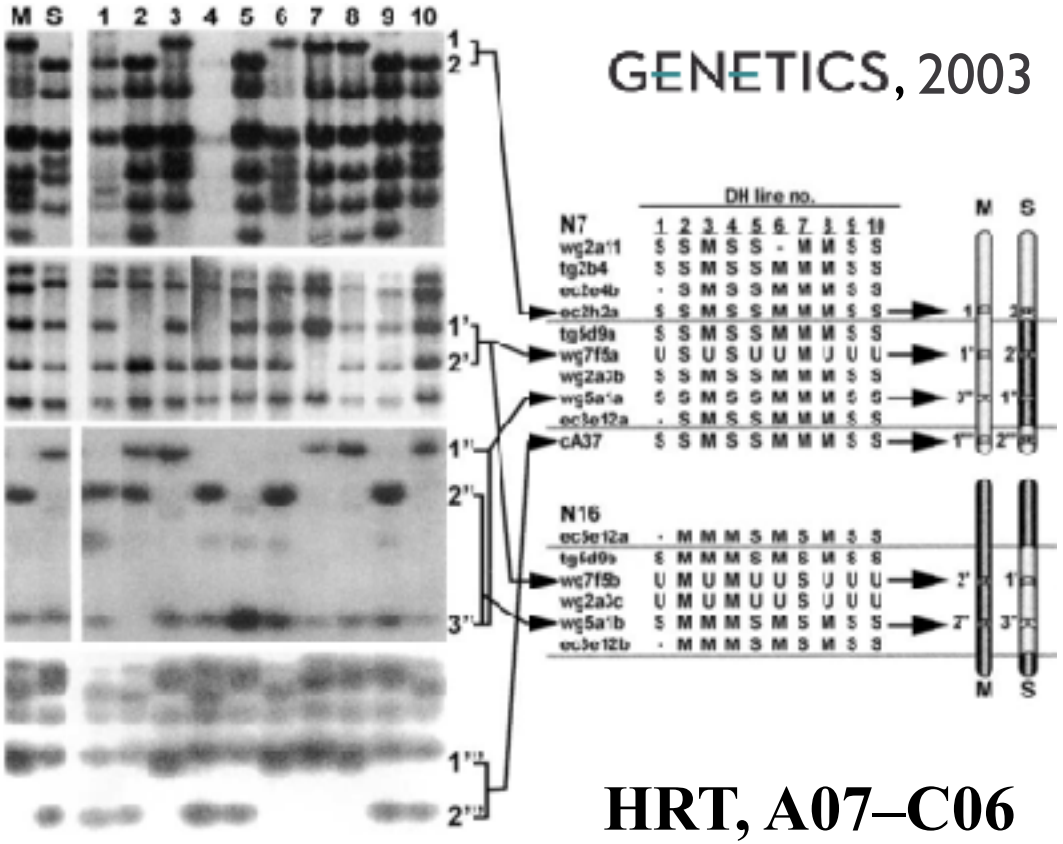


Xiong et al. PNAS 2009

Homoeologous exchanges in natural *B. napus*

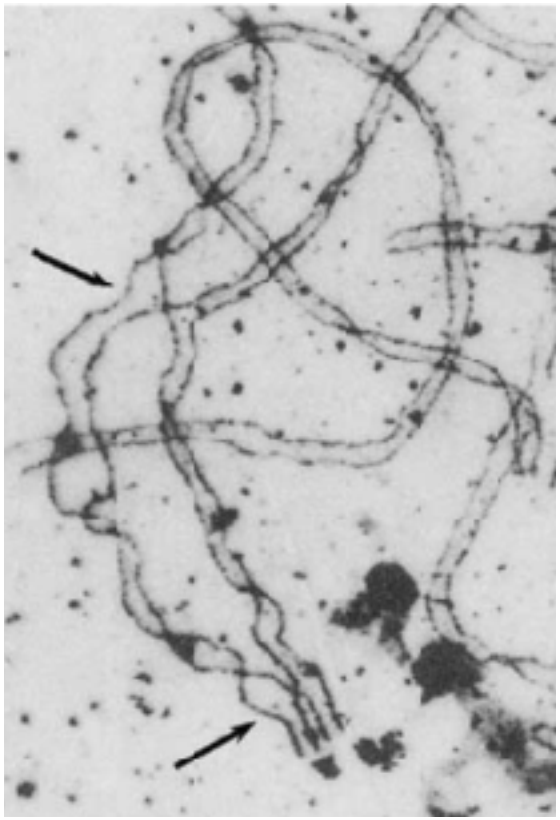
Detection and Effects of a Homeologous Reciprocal Transposition in *Brassica napus*

Thomas C. Osborn,*[,1] David V. Butrulle,*[,2] Andrew G. Sharpe,[†] Kathryn J. Pickering,[‡]
Isobel A. P. Parkin,[†] John S. Parker[§] and Derek J. Lydiate[†]



GENETICS, 2003

HRT, A07–C06

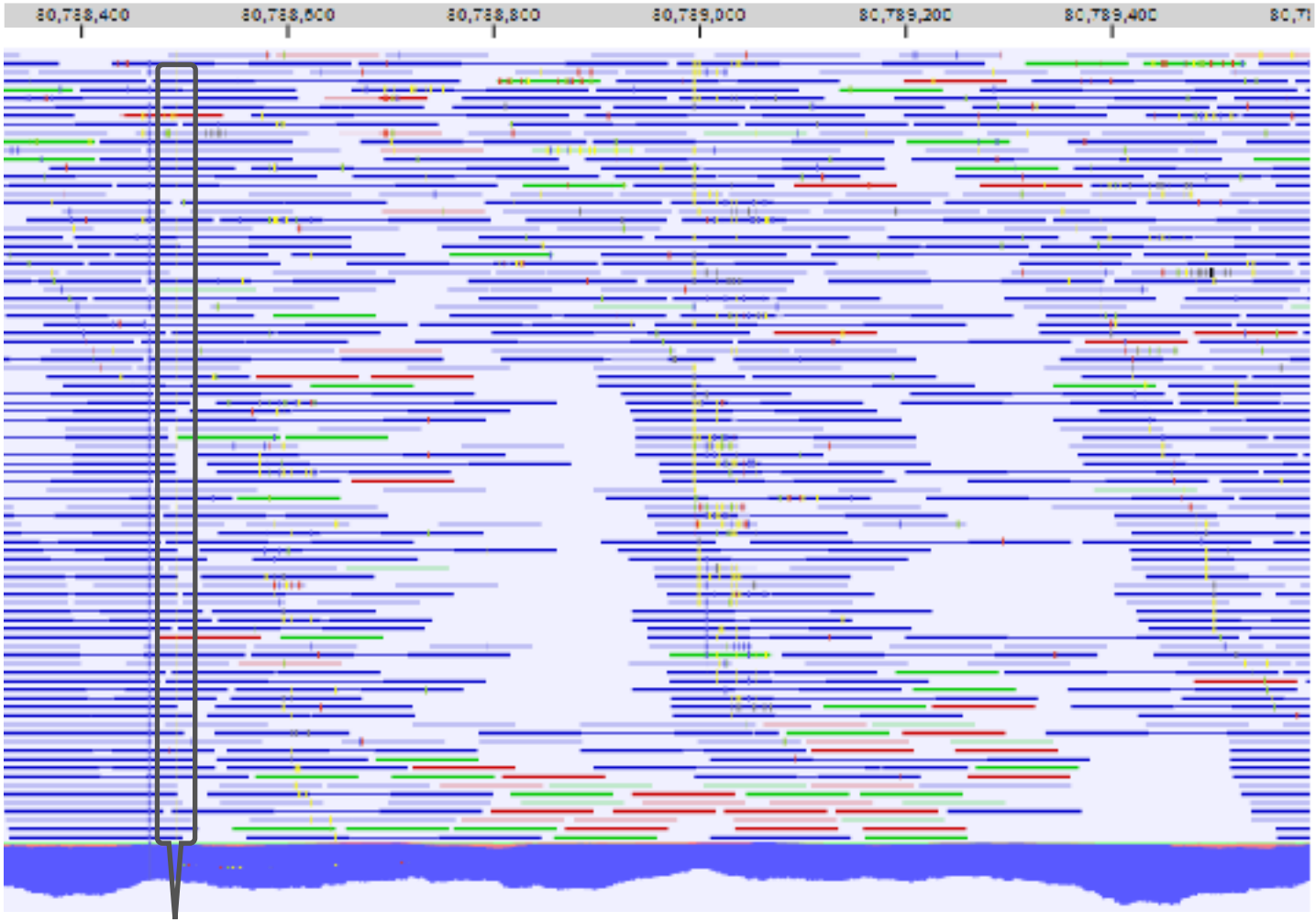


What a difference 20 years makes…

1999

2019

“Genome resequencing”



“Reference” genome sequence (e.g. Darmor 4.1)

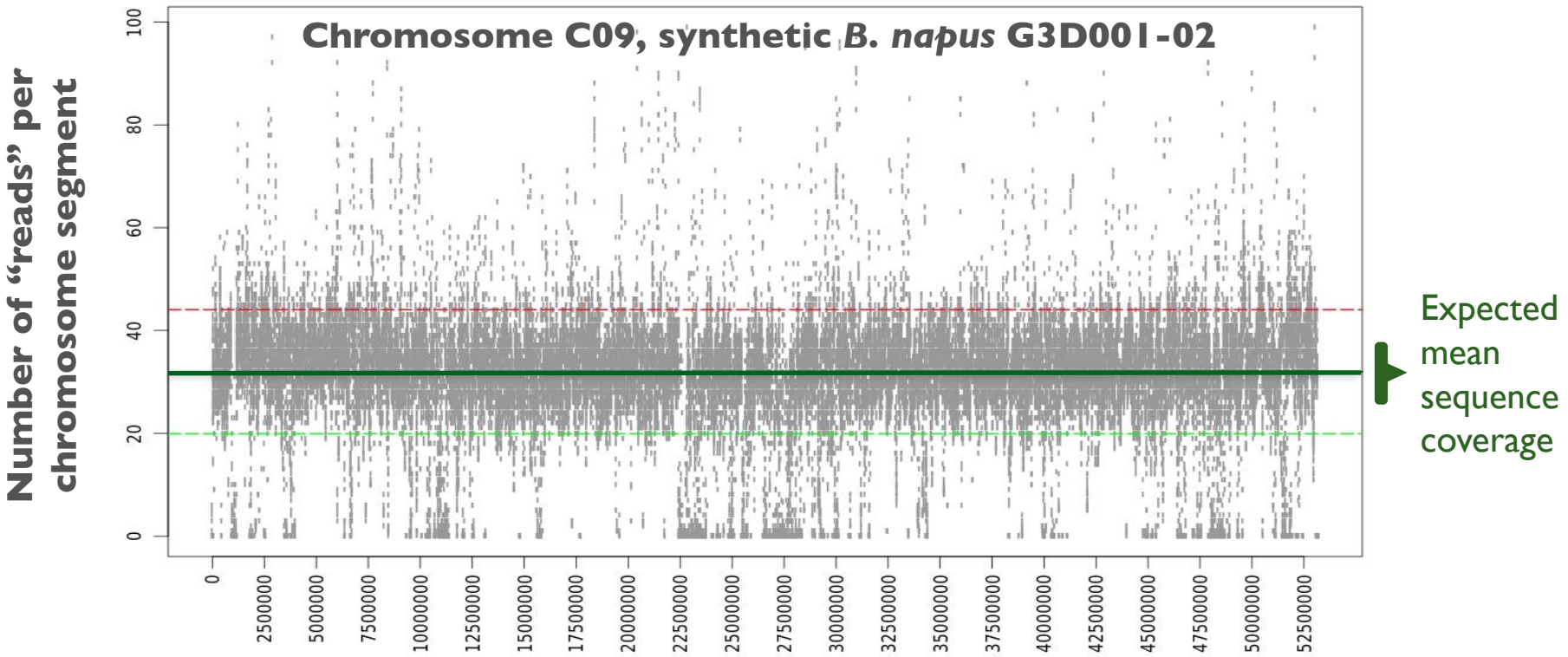
Billions of random, **short DNA sequence fragments** aligned to reference

Mean **“coverage”** with sequenced fragments

Detection of genome-wide sequence variants
(e.g. “single-nucleotide” or “Insertion-Deletion polymorphisms – SNPs, InDels)

“Cytogenetics by sequencing”

Expectation: Random DNA fragments should be distributed evenly across the genome

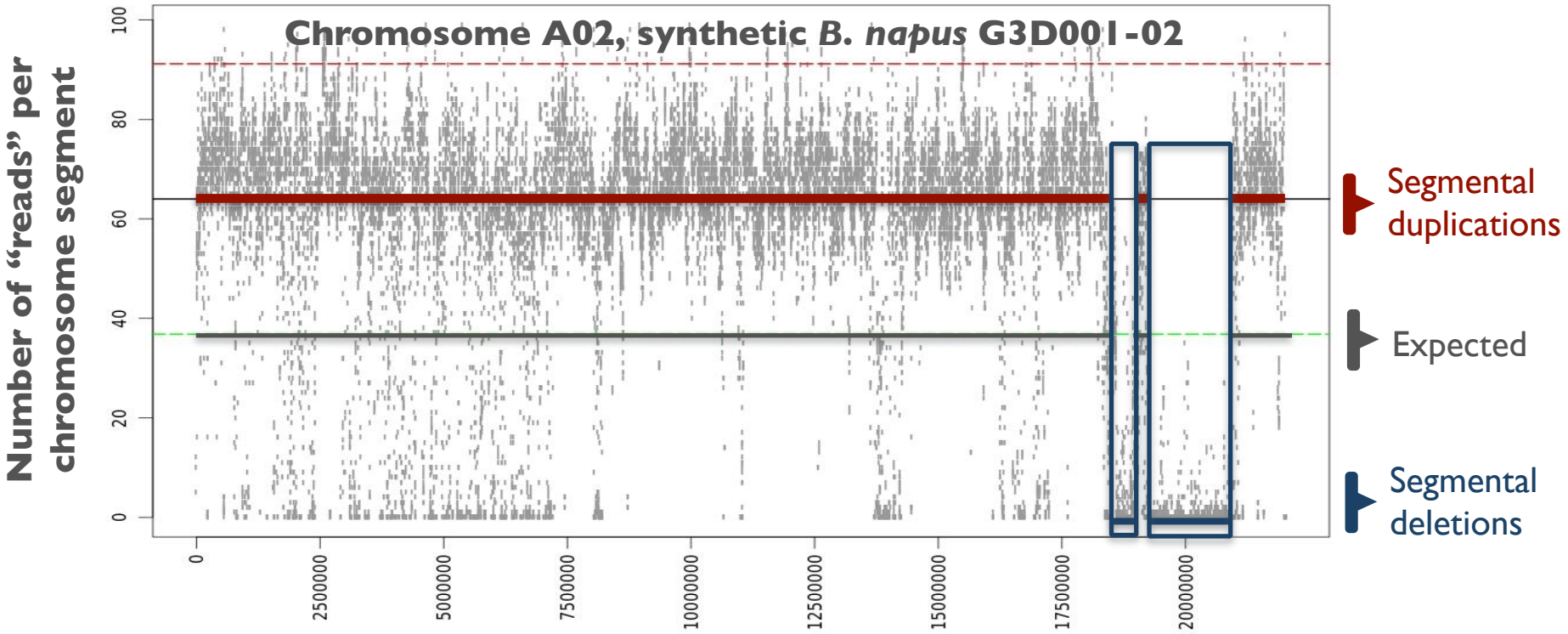


Mauricio Orantes Bonilla, Jun Zou

Rod Snowdon, JLU Giessen

“Cytogenetics by sequencing”

Consequently: Major chromosome rearrangements cause signifi cant deviations from the expected, even read distributions

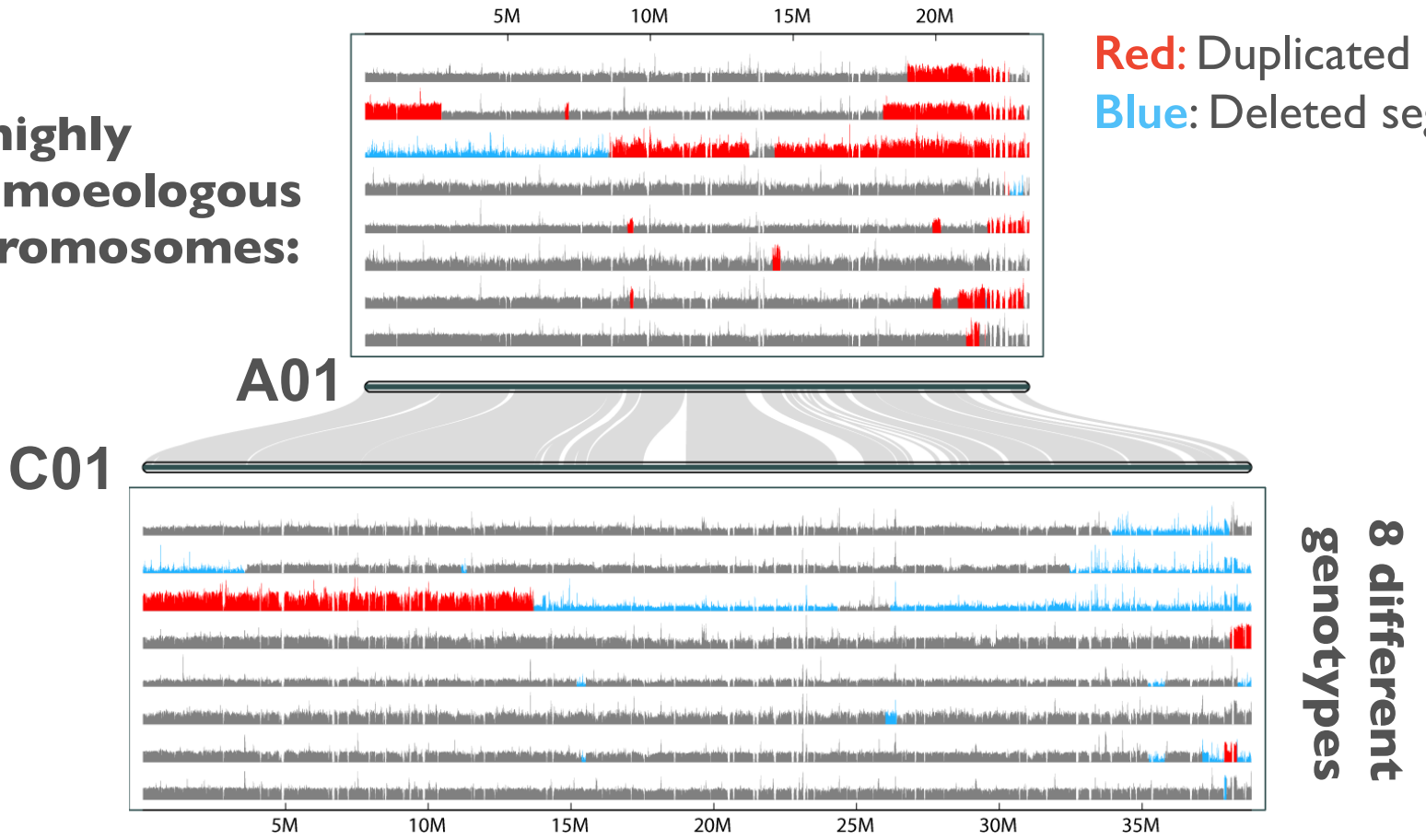


Mauricio Orantes Bonilla, Jun Zou

Rod Snowdon, JLU Giessen

Dynamic genome rearrangements: Examples

2 highly homoeologous chromosomes:

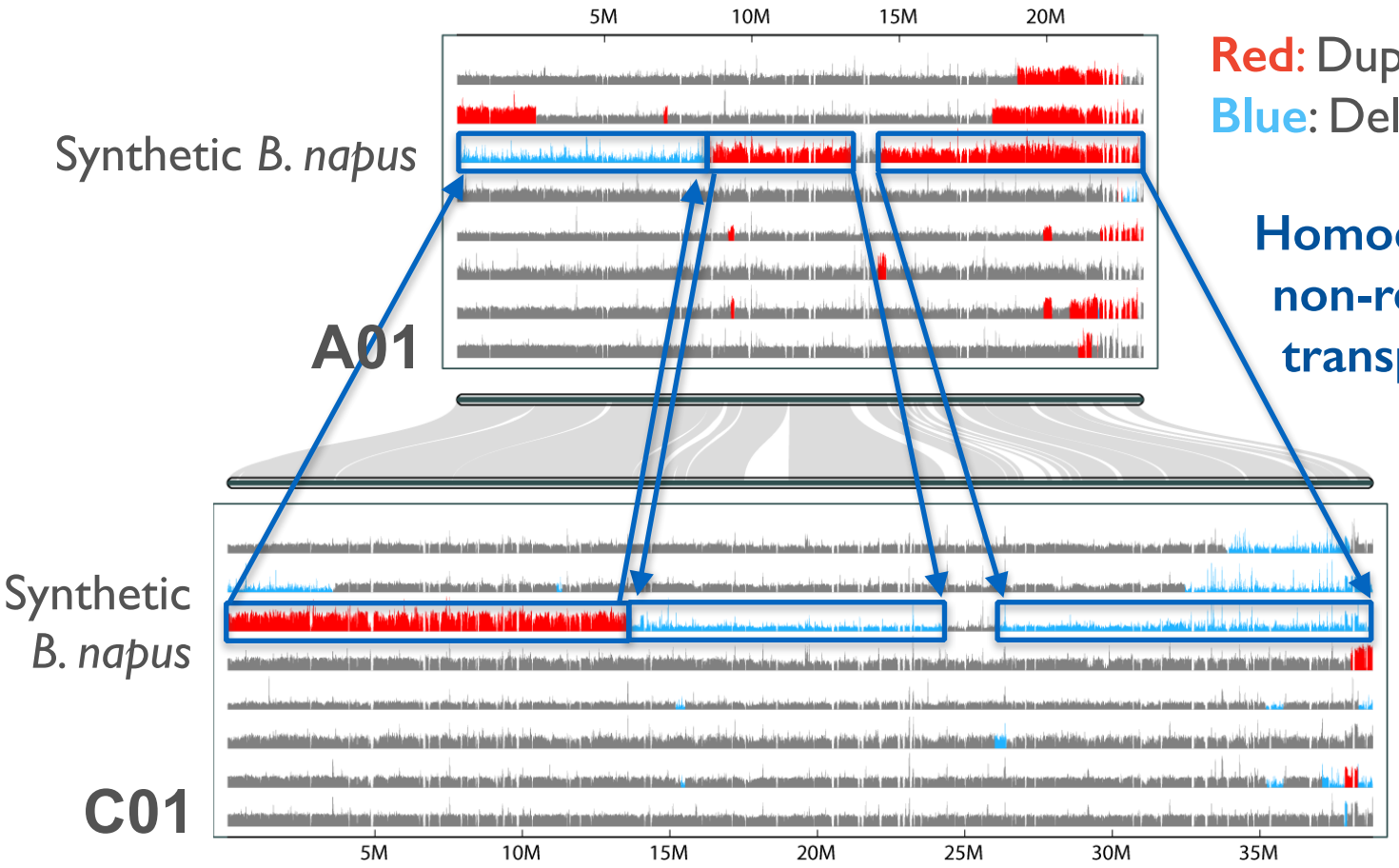


Red: Duplicated segments
Blue: Deleted segments

8 different genotypes

Chalhoub et al., *Science* (2014)

B. napus: Widespread structural genome variation



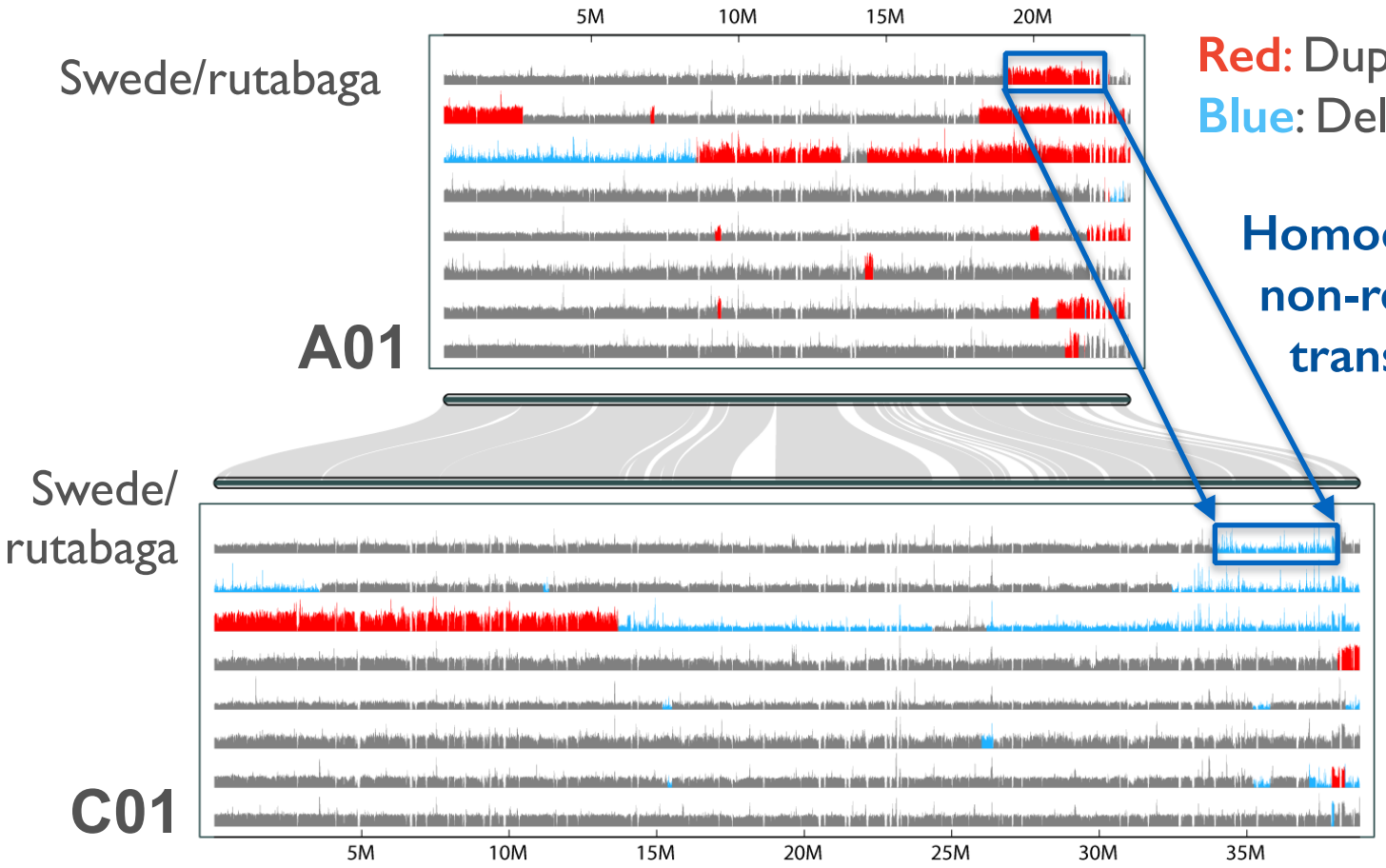
Red: Duplicated segments

Blue: Deleted segments

Homoeologous non-reciprocal transpositions

Chalhoub et al., *Science* (2014)

B. napus: Widespread structural genome variation



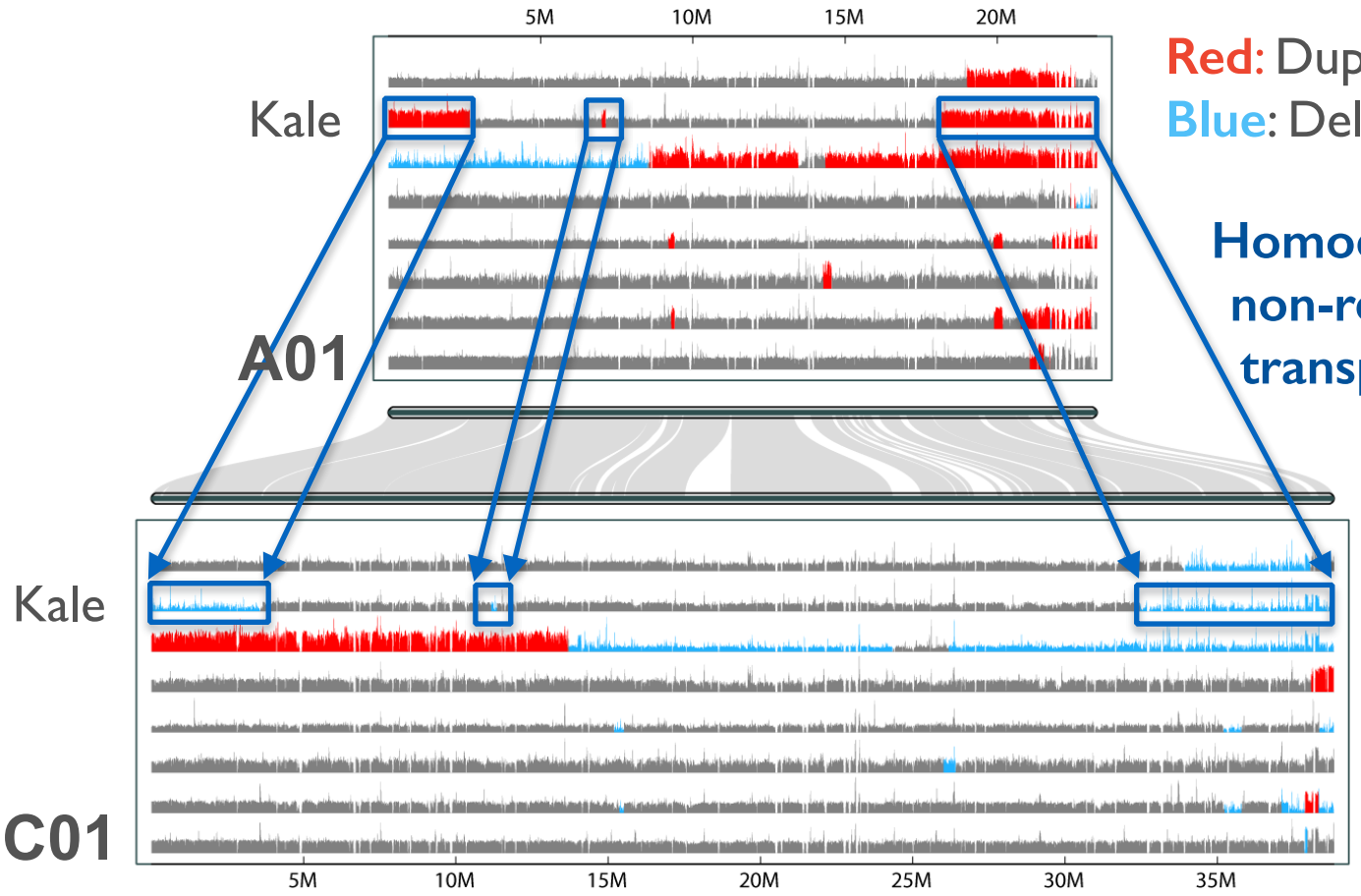
Red: Duplicated segments

Blue: Deleted segments

Homoeologous non-reciprocal transposition

Chalhoub et al., *Science* (2014)

B. napus: Widespread structural genome variation



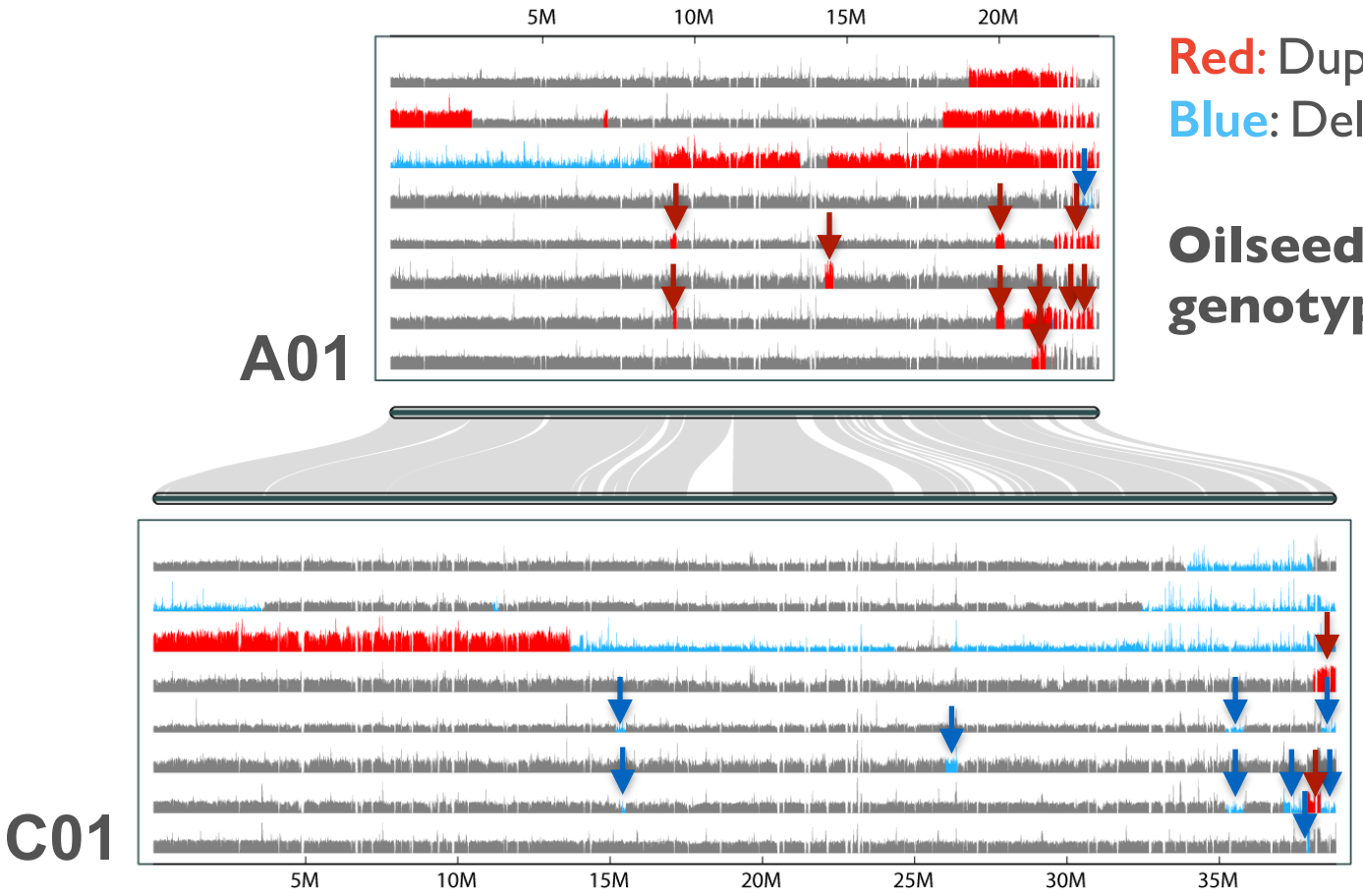
Red: Duplicated segments

Blue: Deleted segments

Homoeologous non-reciprocal transpositions

Chalhoub et al., *Science* (2014)

B. napus: Widespread structural genome variation



Red: Duplicated segments

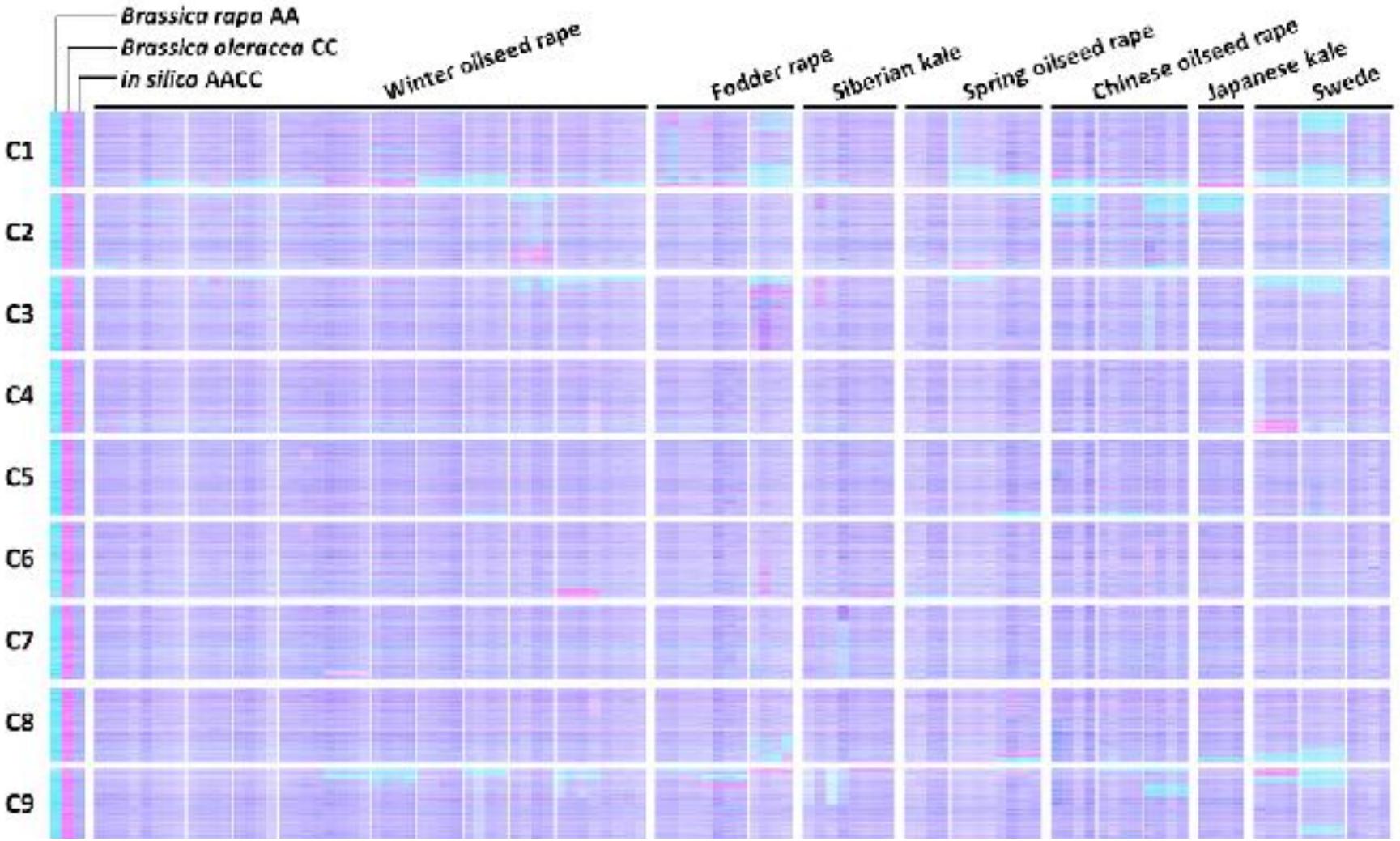
Blue: Deleted segments

Oilseed genotypes

Oilseed genotypes

Chalhoub et al., *Science* (2014)

Independent validation with transcriptome data



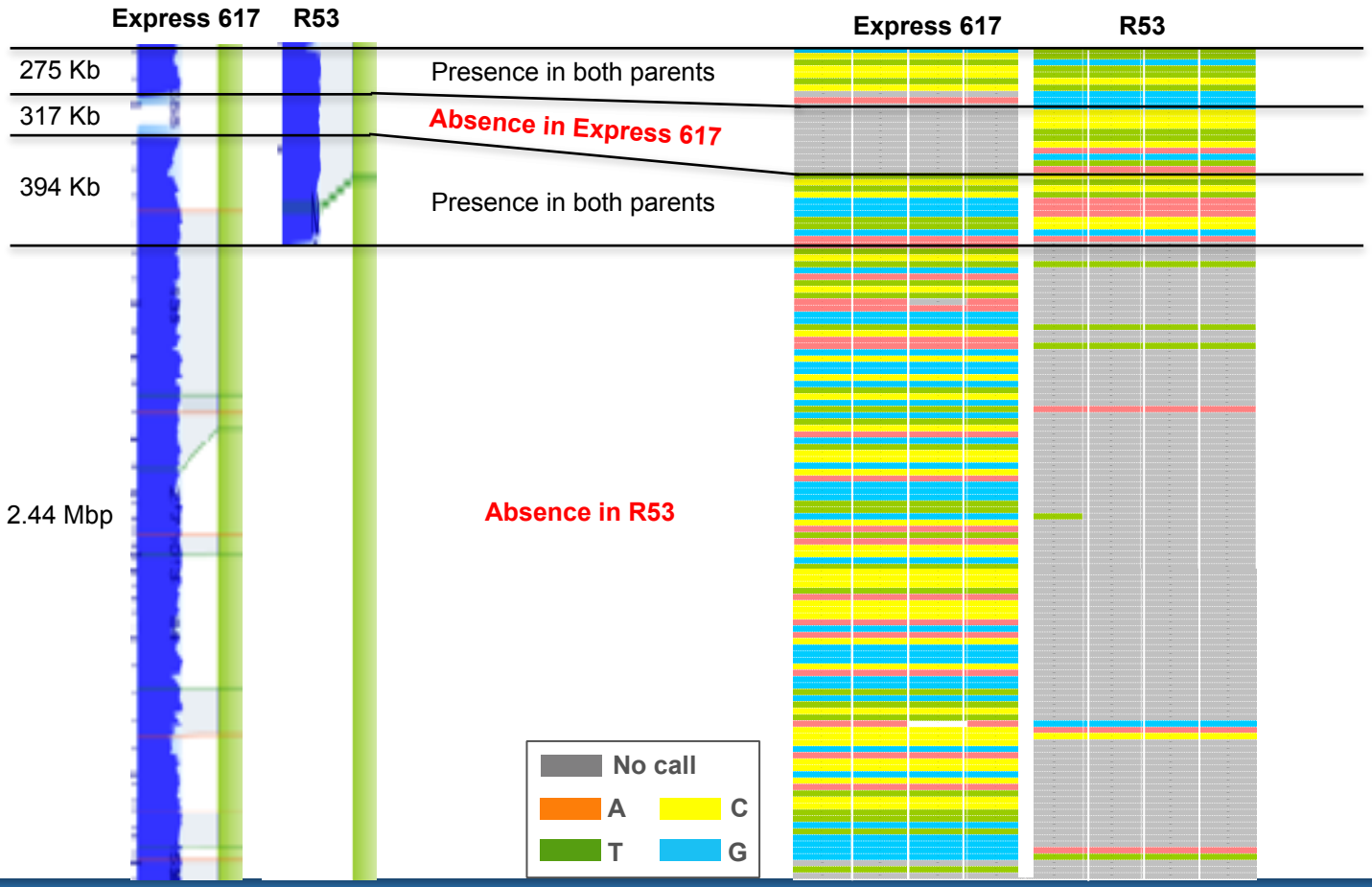
Zhesi He, Ian Bancroft et al. 2015, 2016

UNIVERSITY of York

Independent validation with genetic/genomic data

Independent validation: Example confirming predicted large-scale deletions in natural (Express 617) and synthetic (R53) parents of a DH mapping population

1) Reference-anchored Optical Mapping contigs **2) Reference-ordered SNP markers**

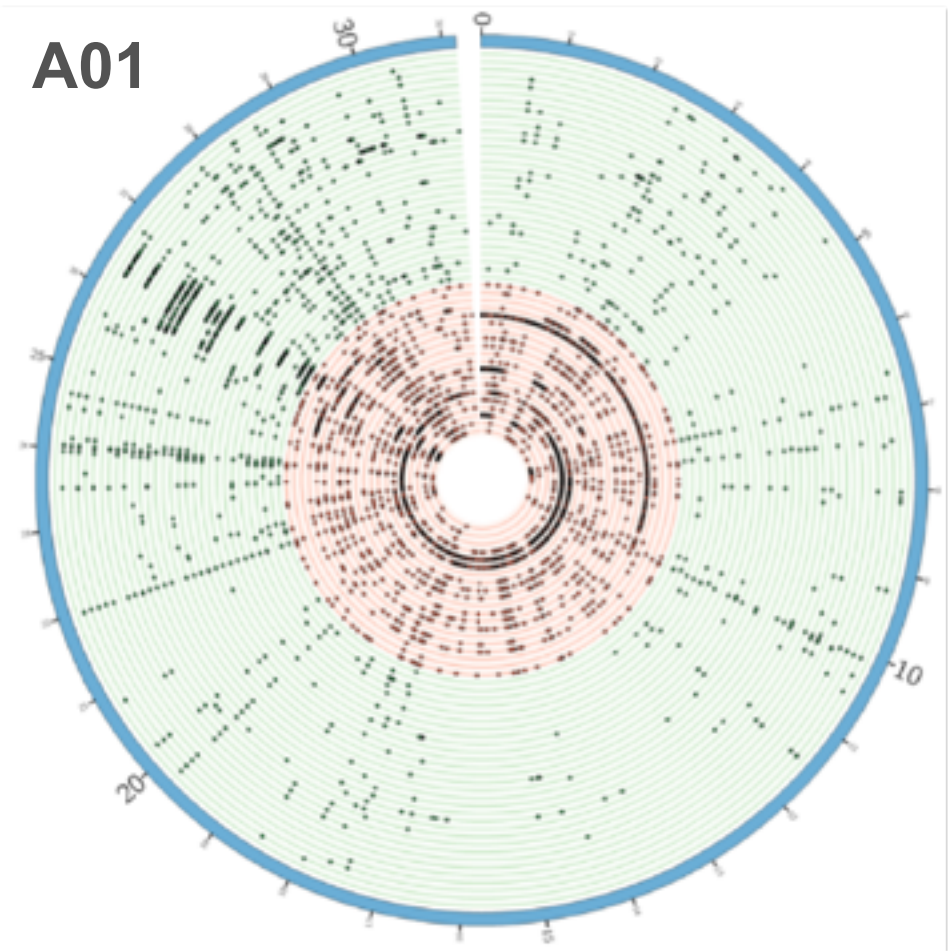


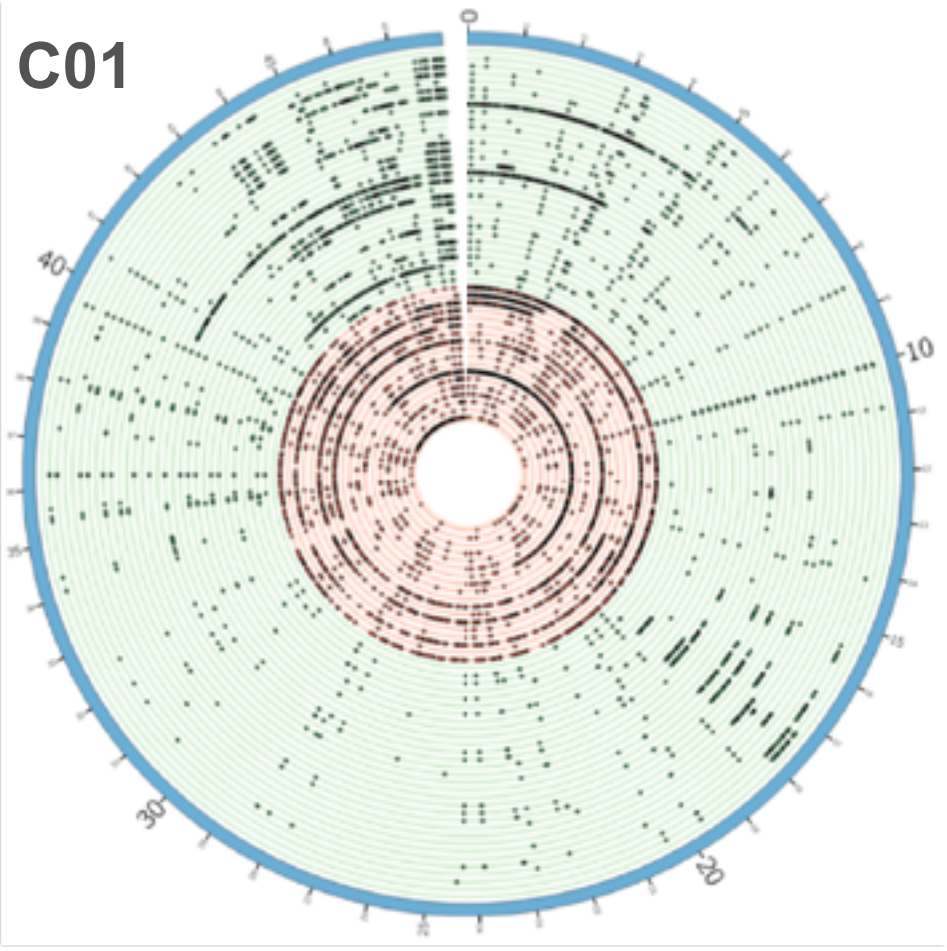
Harmeet Singh Chawla, Christian Obermeier

Rod Snowdon, JLU Giessen

Large-scale SV: Extensive gene loss

Example: Patterns of segmental gene deletions on chromosomes A01/C01





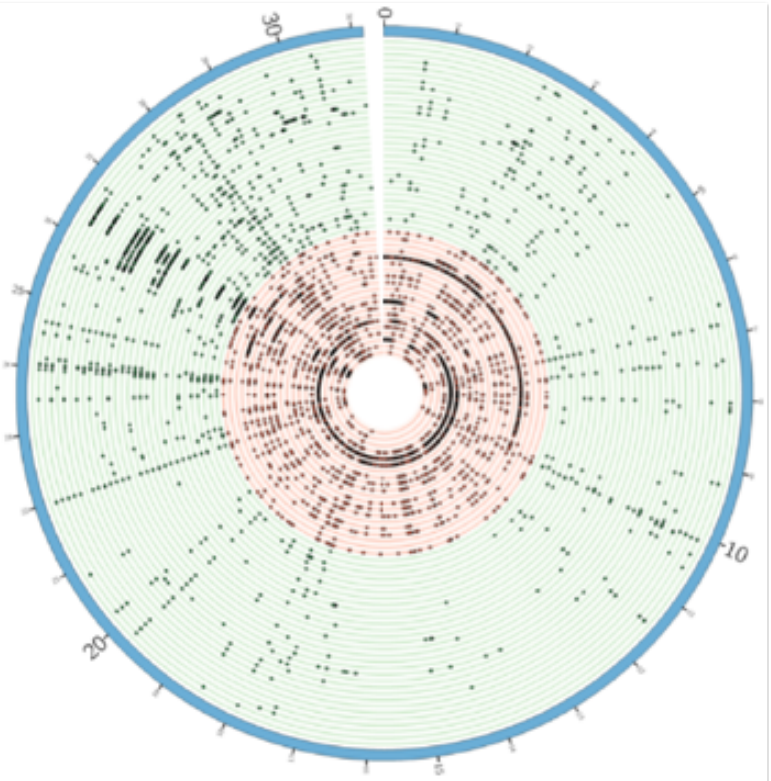
Hurgobin *et al. Plant Biotech J* (2017), Samans et al., *Plant Genome* (2017)
Data from 52 *B. napus* genomes, Schmutzer et al., *Scientific Data* (2015)

Bhavna Hurgobin, Dave Edwards (University of Western Australia)
Rod Snowdon, JLU Giessen

Deletions may be a key to polyploid crop adaptation

Significantly enriched GO terms among genes affected by PAV

Font size proportional to –log(p)

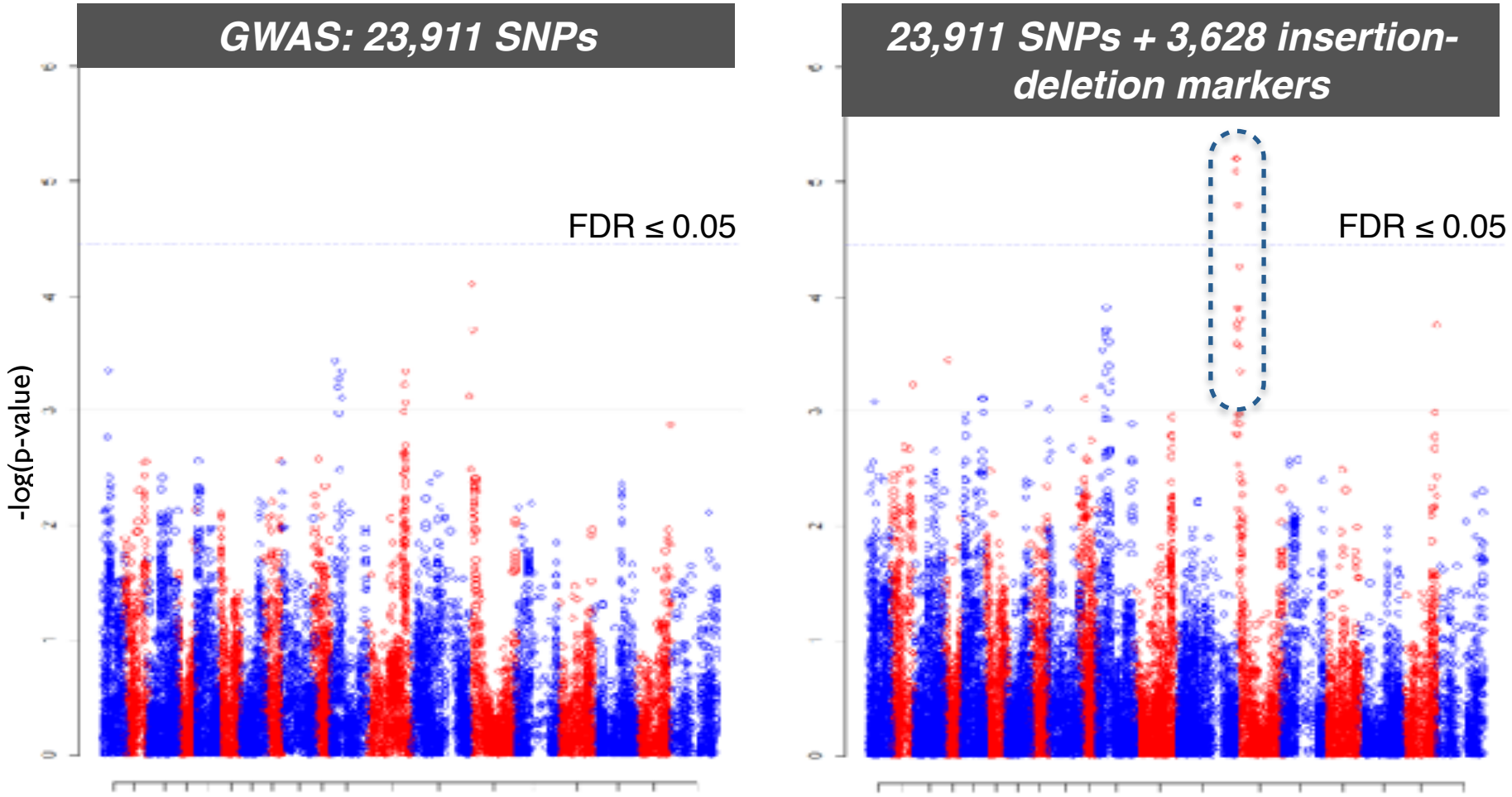


Hurgobin *et al. Plant Biotech J* (2017)

Bhavna Hurgobin, Dave Edwards (University of Western Australia)

Rod Snowdon, JLU Giessen

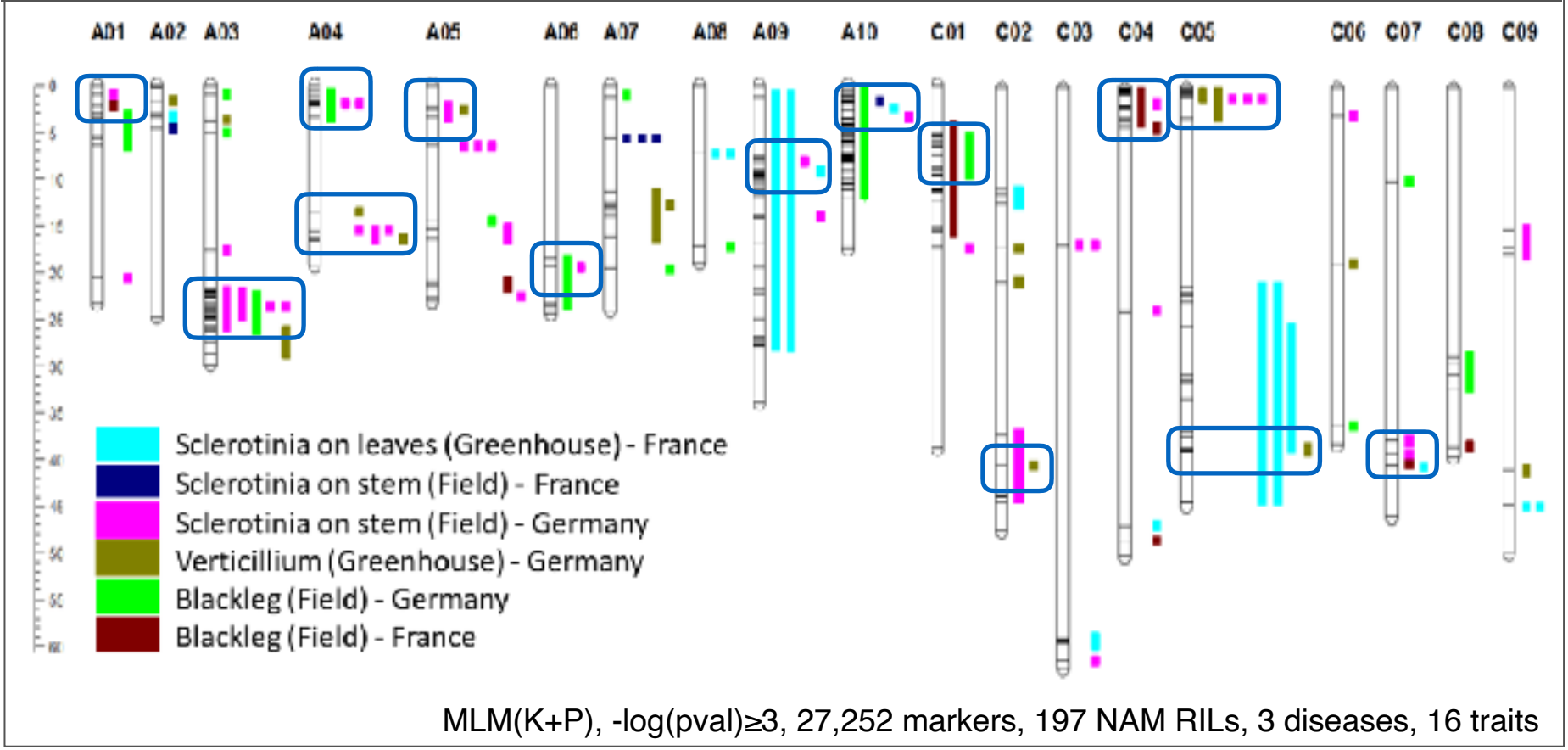
“SNP absence” associates with resistance QTL



Example: Increased detection power and resolution of **QTL for adult plant blackleg resistance** on chromosome C04 by implementation of SNAP markers

Finding "invisible" resistance QTL with missing data

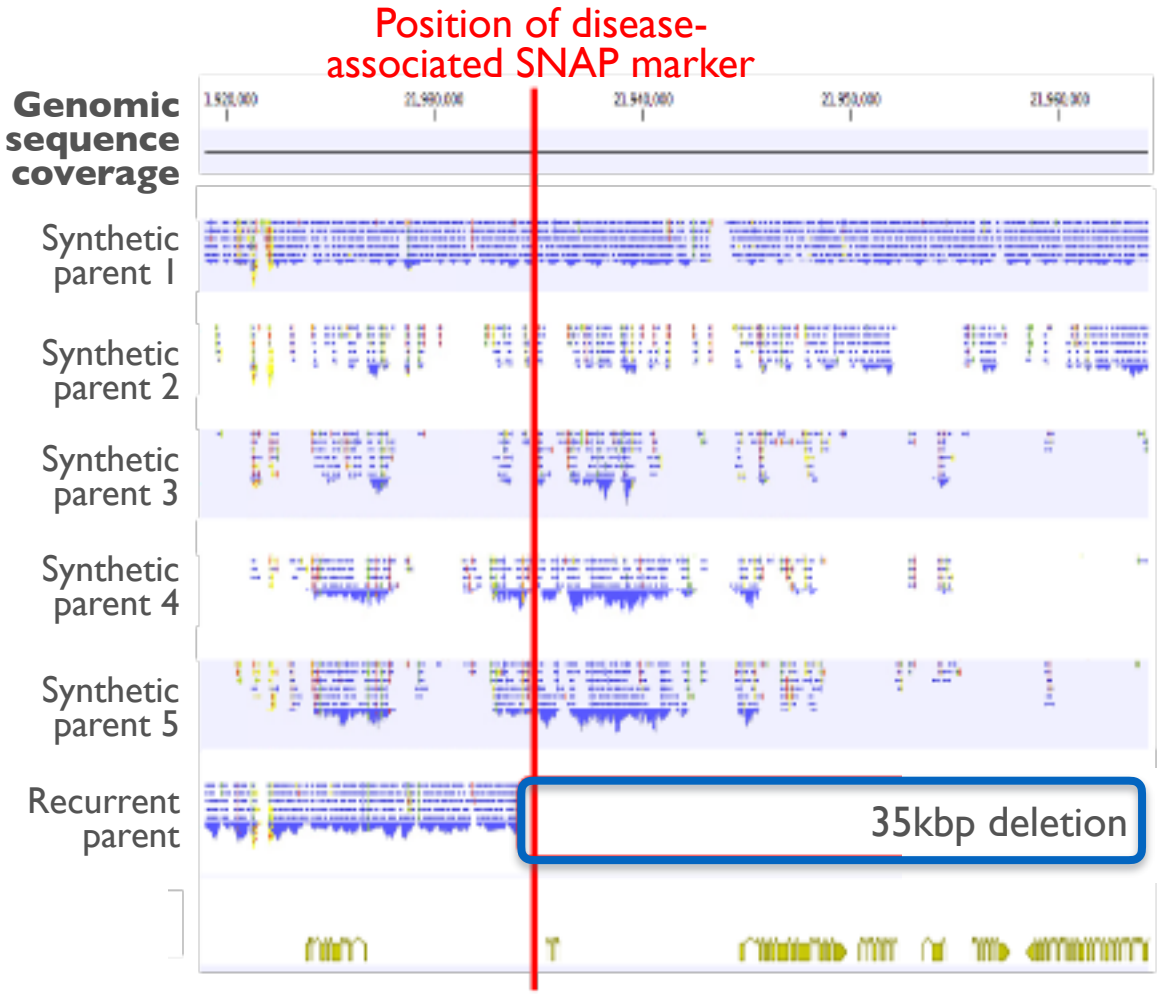
GWAS in 5 NAM-subfamilies: 23,911 SNPs + 3,628 SNAPs

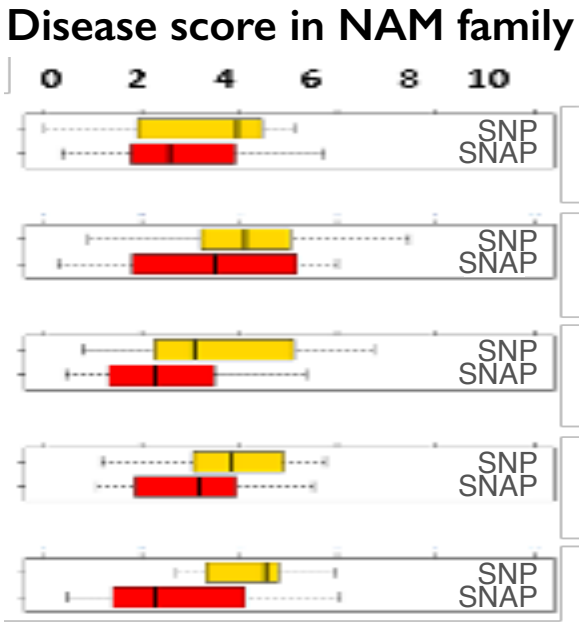


MLM(K+P), -log(pval)≥3, 27,252 markers, 197 NAM RILs, 3 diseases, 16 traits

- **Numerous overlapping loci influencing quantitative resistance to one or more diseases associate with SNAP markers**

Sequence data associates gene absence to QTL





Absence **decreases** disease score

- **Example: Sclerotinia** stem necrosis QTL, **resistance associated with a segmental deletion** spanning **five genes** in the Darmor-*bzh* reference

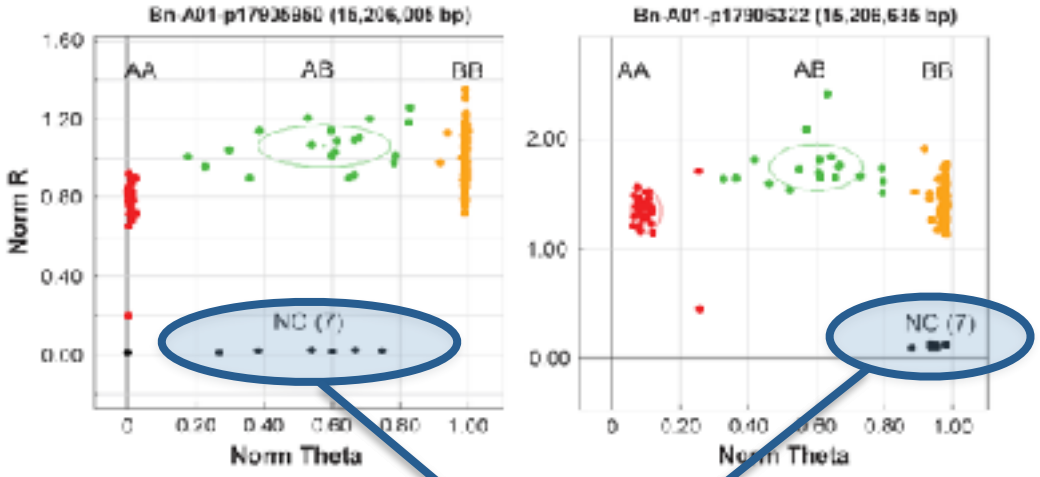
The closer we look, the more we see

- High-throughput "sequence capture" experiments suggested that small-scale structural variants may be much more frequent than large-scale SV

Questions

- How much does SV impact gene presence/absence or function?
- How can we accurately detect and assay smaller-scale SV
- How does small-scale SV arise?

Identifying structural variants from SNP array data



gsrc: An R package for genome structure rearrangement calling

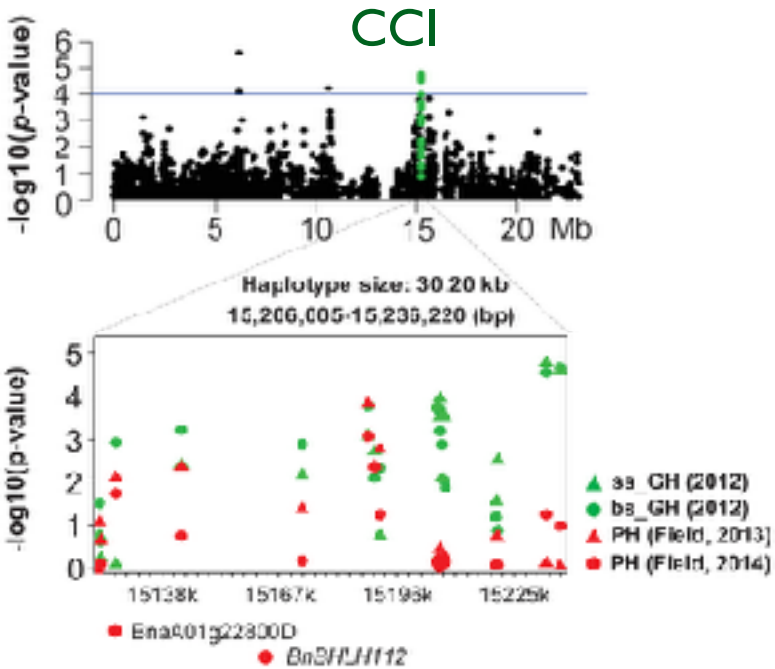
Grandke et al. *Bioinformatics* (2016)

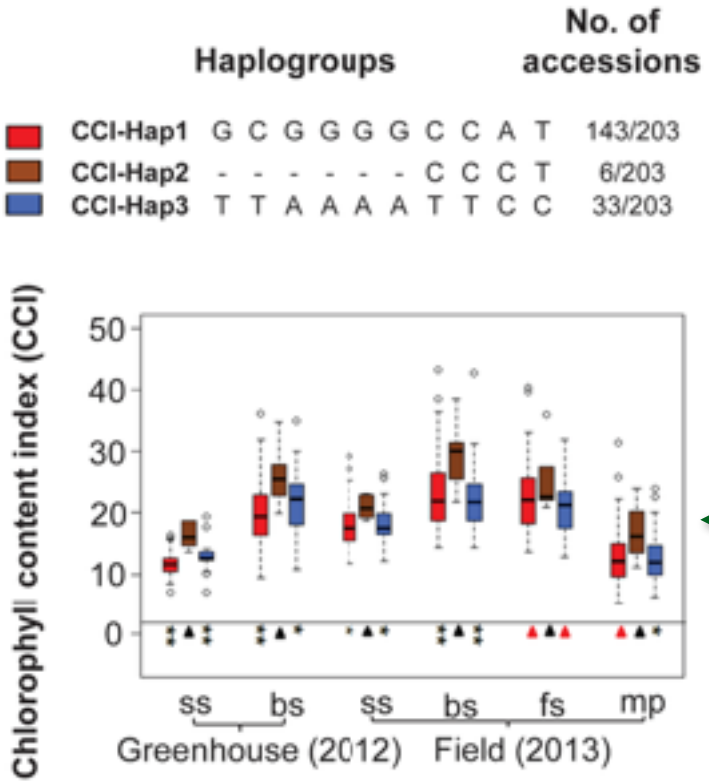
Physically adjacent SNP markers which consistently "fail" in the same genotypes represent deletions in those genotypes and can be confidently called as **deletions**

~ 5-20% of assays on *Brassica* 60k SNP array

Deletions associate with selection patterns

Example: Strongly selected haplotype block on A01 associated with chlorophyll content (CCI), plant height (PH) and oil content in Chinese rapeseed



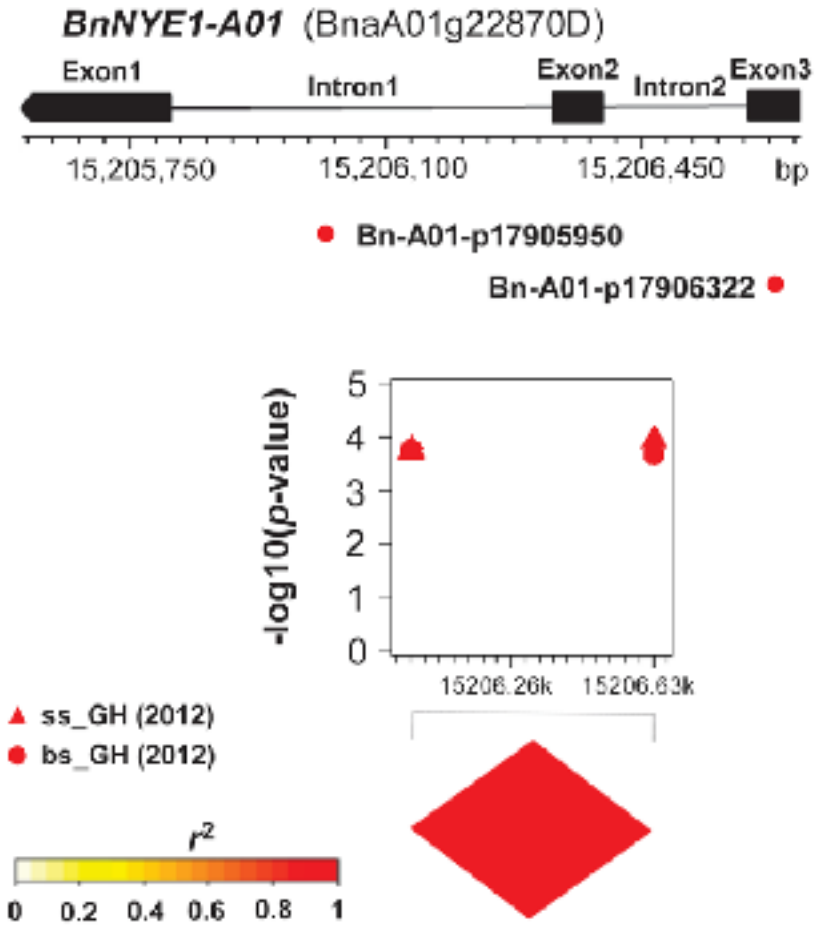


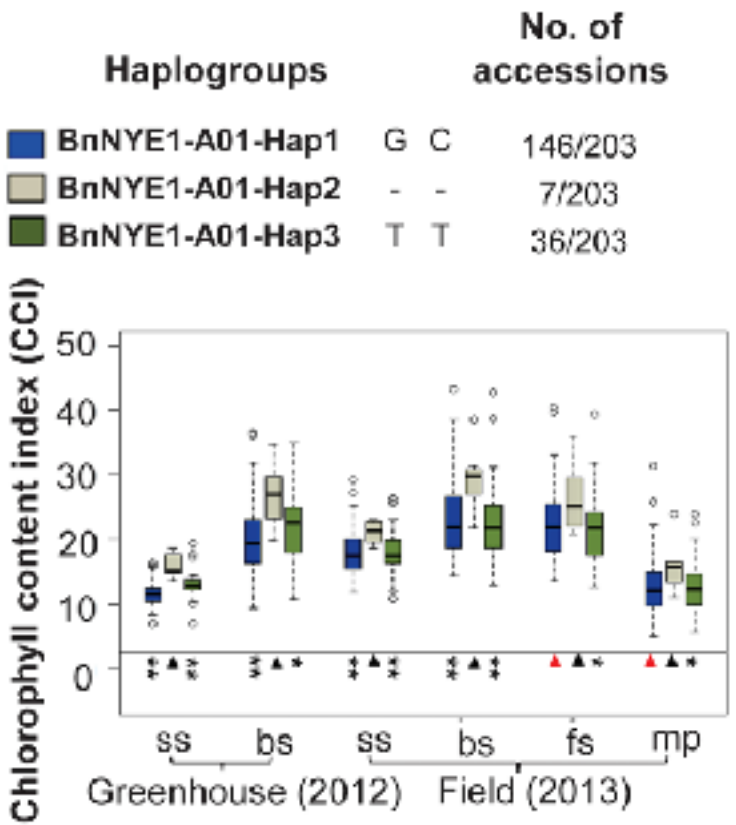
3 haplogroups

Qian et al., Molecular Plant (2016)

Deletions associate with quantitative trait variation

Phenotypic variance explained completely by haploytpes of two SNPs within *BnNYE1-A01*





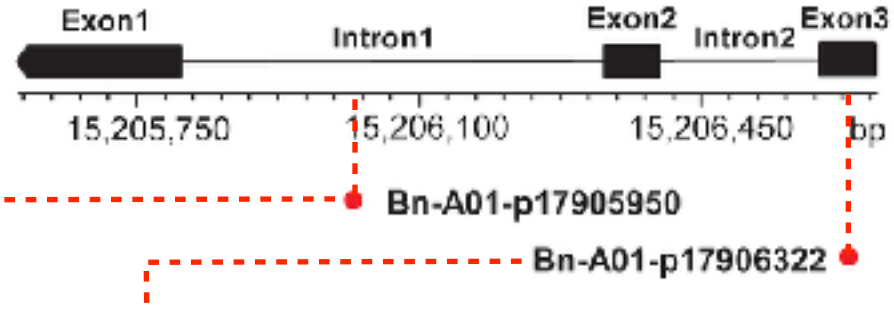
Qian et al., Molecular Plant (2016)

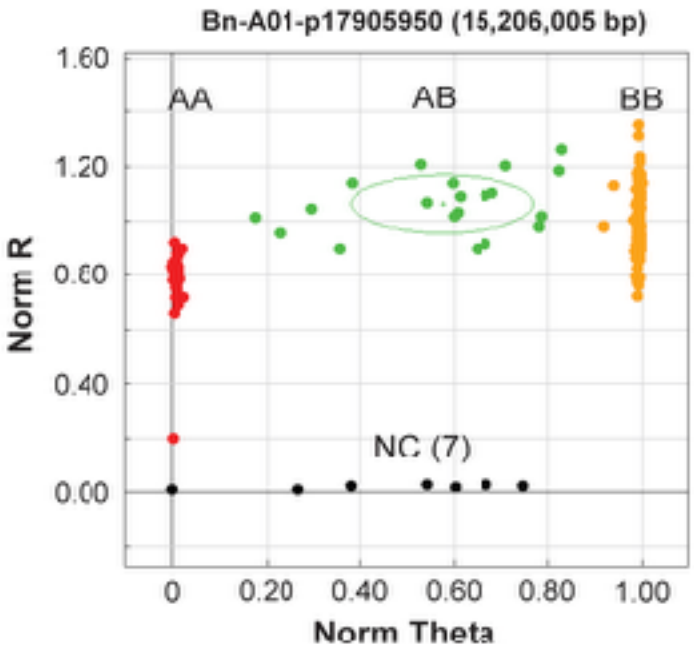
Deletions can impact at single-gene level

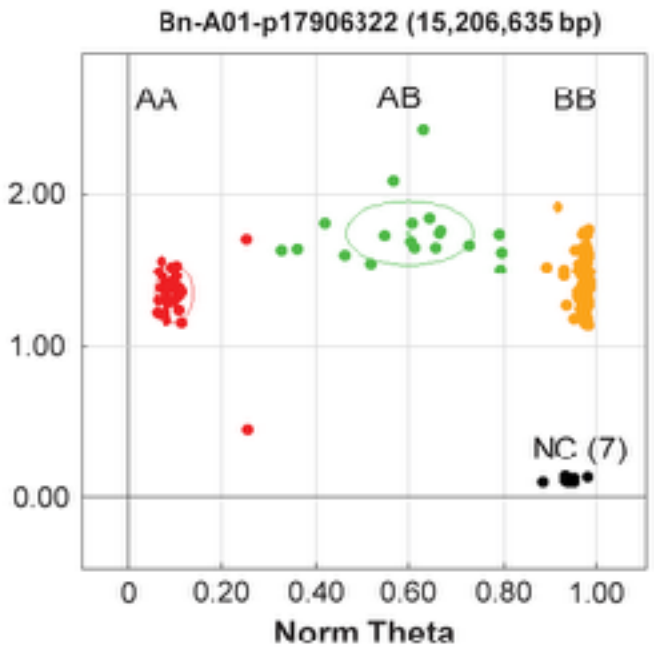
A novel, small-scale intergenic deletion

BnNYE1-A01

B. napus orthologue of *STAYGREEN 1*
(Mendel's Green Cotyledon gene)







Both chlorophyll-associated SNPs **fail** in the seven accessions with elevated chlorophyll

Qian et al.
Mol Plant (2016)

Presence-absence associates with all sorts of traits

Chlorophyll content, stress adaptation

- *NYE1* gene deletion
  (Qian et al. *Mol Plant* 2017)

Seed quality

- Glucosinolate content (Harper et al. *Nat Biotech* 2012)
- Seed coat fibre (Stein et al. *Plant Biotech J* 2017)

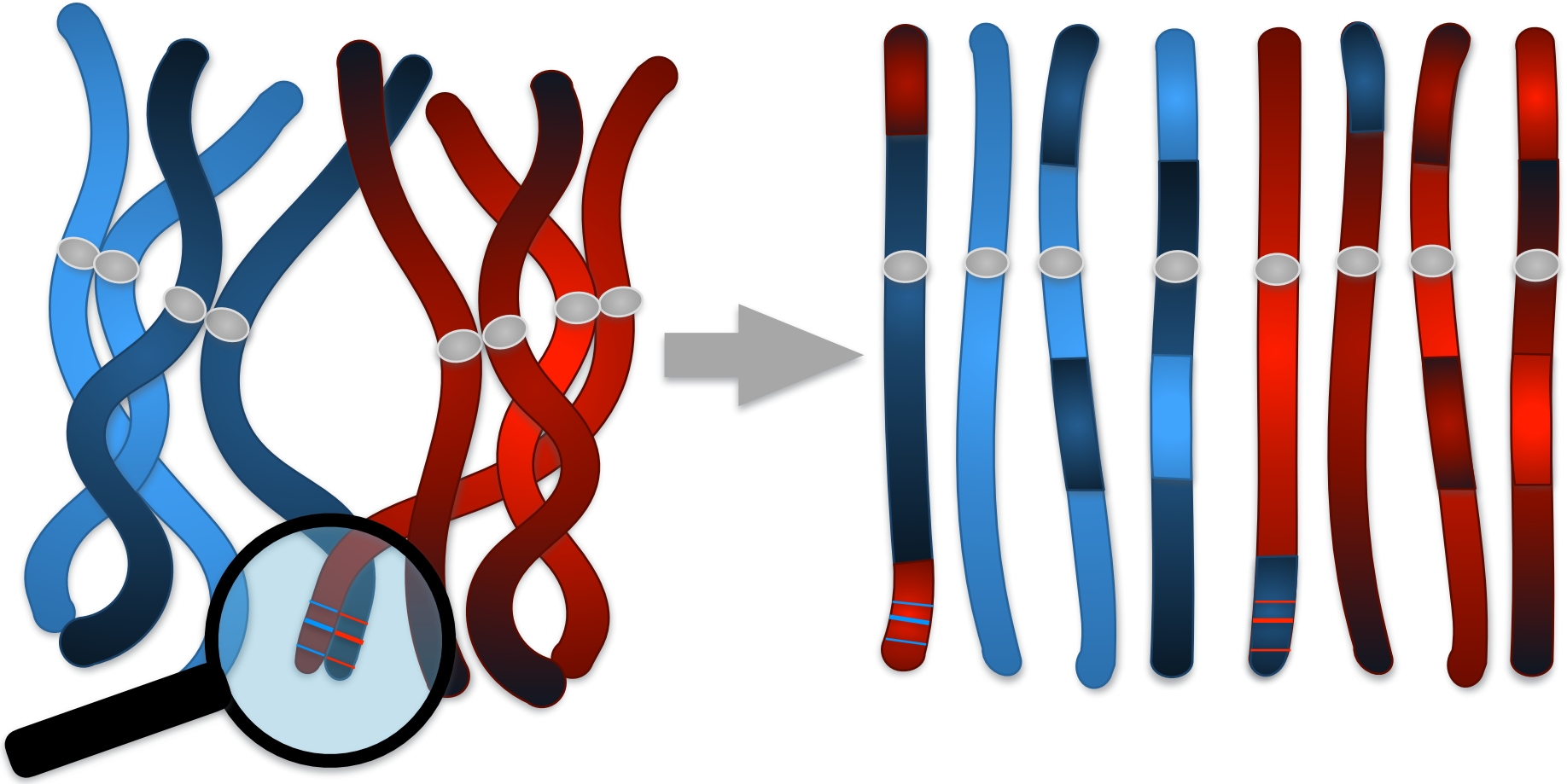
Flowering time, vernalisation, cold tolerance

- Widespread copy-number and presence-absence variation among flowering.time regulatory genes in all *B. napus* gene pools (Schiessl et al. *Scientific Data* 2017, *Scientific Reports* 2017)

And probably many other things too?

How do small-scale homologous exchanges arise?

Is there more weird stuff going on during homoeologous interactions?



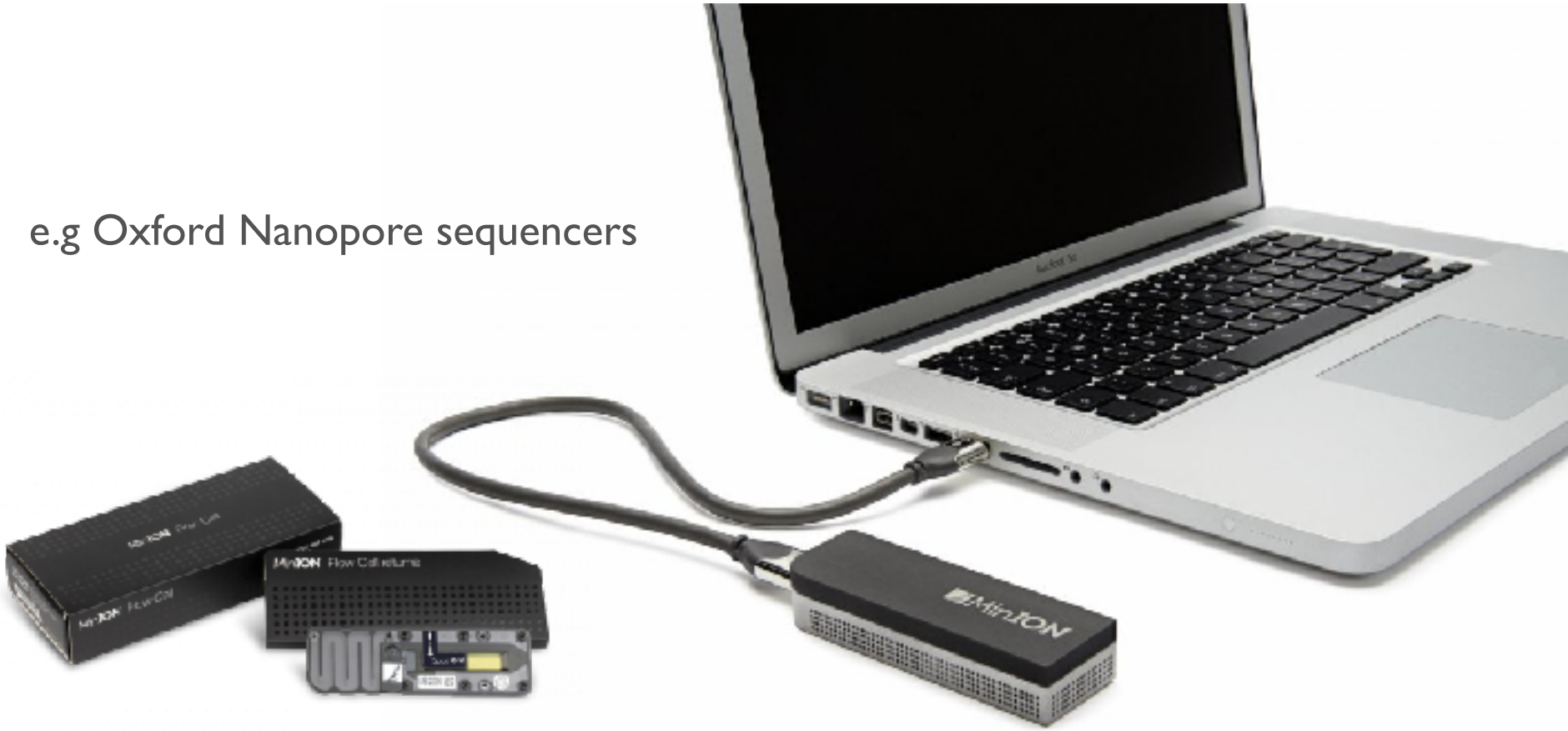
Gene conversions, likely arising from small-scale homoeologous exchanges

Recombined, rearranged gametes with redesigned, “chimeric” genes

How much small-scale SV are we missing?

- **Problem:** Deleted SNPs and Illumina reads can only detect large-scale SV
- **But:** "Long-read" sequencing is astoundingly good at tracing gene-level SV events

e.g Oxford Nanopore sequencers



How much small-scale SV are we missing?

- **Problem:** Deleted SNPs and Illumina reads can only detect large-scale SV
- **But:** "Long-read" sequencing is astoundingly good at tracing gene-level SV events

e.g Oxford Nanopore sequencers

Up to 48 flow cells simultaneously, yields up to 7.6 Tb of data per run

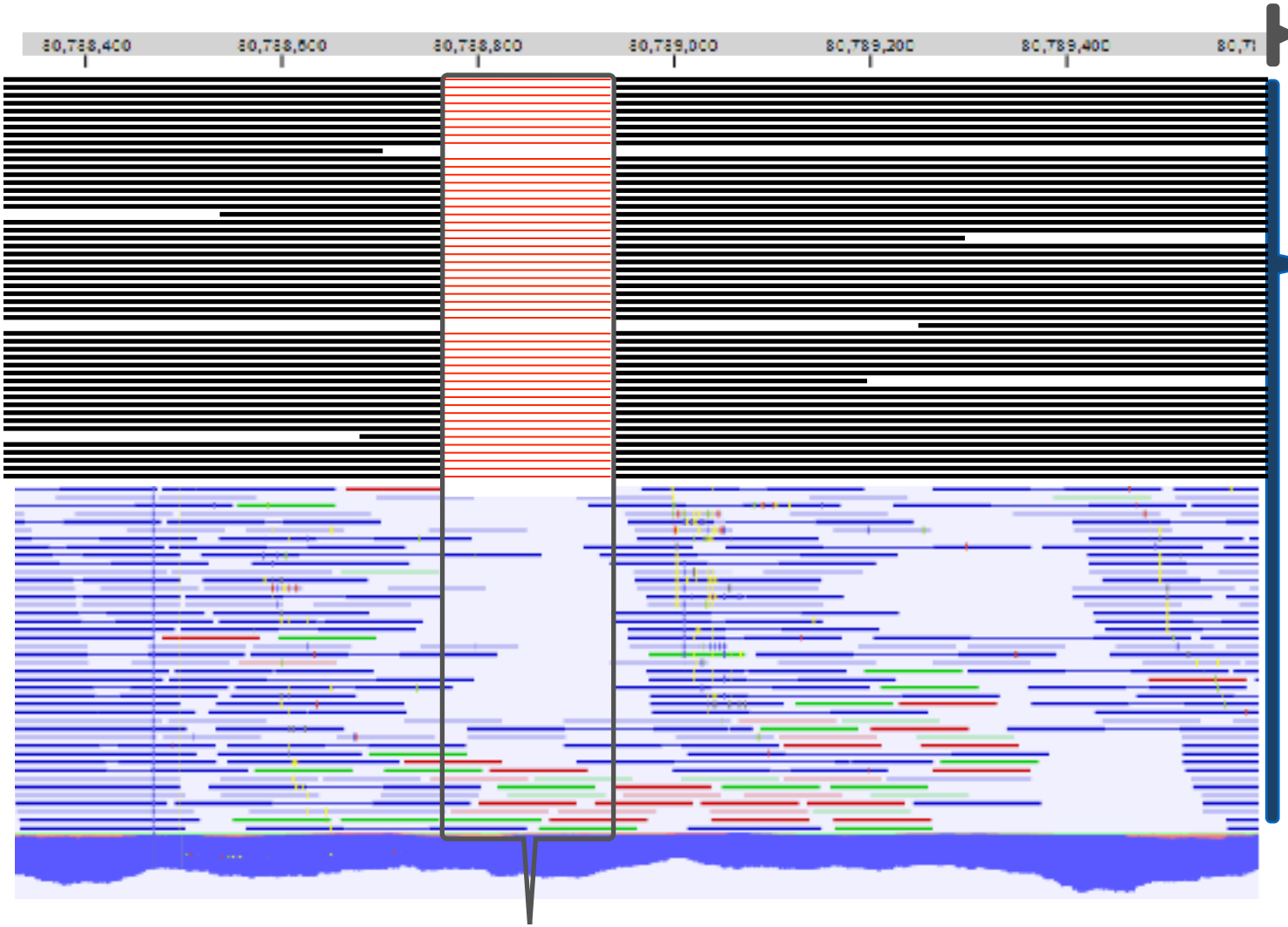
SV detection from Nanopore sequence data

ONT MinIon output, single flow cell

Read statistics	Express 617 (WOSR)
N50	36 kbp
>10 kbp	95 % of raw reads
>20 kbp	85 %
>50 kbp	30 %
>100 kbp	8 %
Total data	21.5 Gbp
Median identity (accuracy)	90.8 %

Sequenced fragments >300 times longer than with Illumina reads

Long-reads can accurately detect small deletions

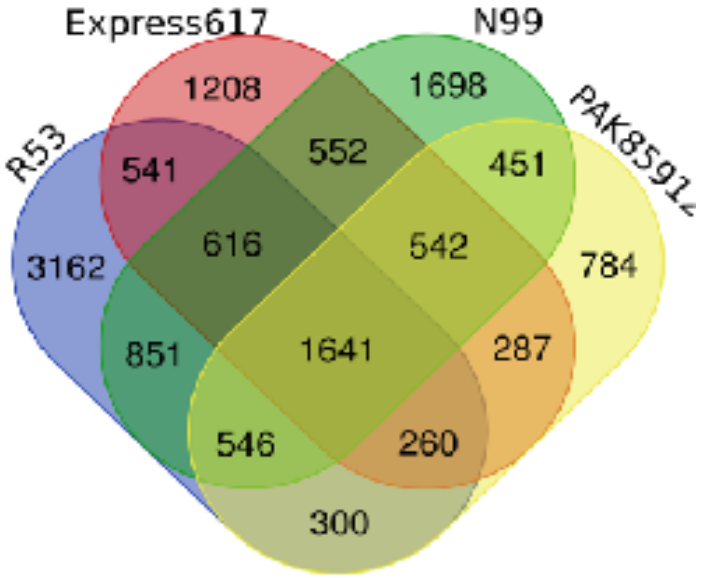


Highly confident detection of small deletions (10~1000 bp)

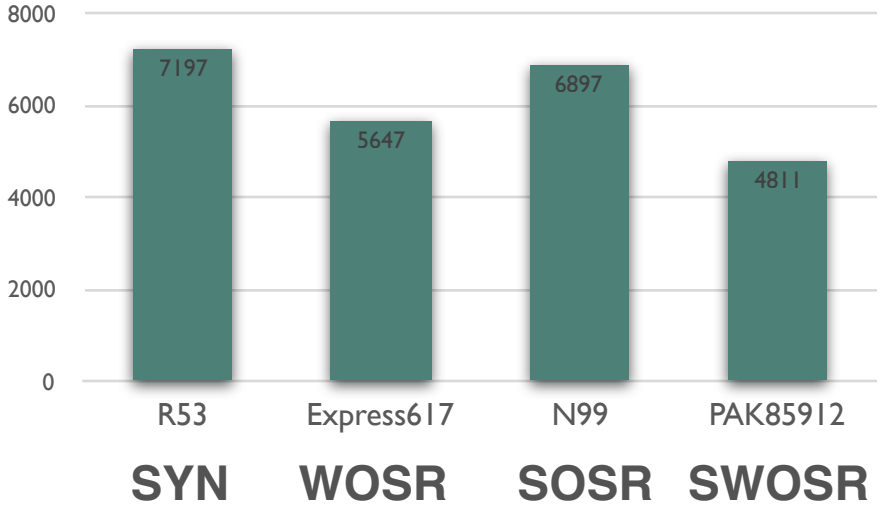
Genome-wide SV size distribution

- Small-scale SV is extremely widespread
- Most SV events range from ~100 to 1000bp
- **5–8% of genes contain small-scale structural variants**
- Also widespread in regulatory regions and other non-coding DNA

Common SV calls



SV frequency



Harmeet Chawla, Isobel Parkin

Rod Snowdon, JLU Giessen

Polyploid “genome collision” causes novel diversity

AA + CC ≠ AACC

What next?

Structural variation is widespread and important in *B. napus*

- SV contributes to adaptive variation and distinguishes major gene pools
- SV drives quantitative trait variation, particularly for stress adaptation, and is therefore highly relevant for breeding

Small-scale SV is extremely widespread

- Long-read sequencing technologies will unlock undiscovered diversity
- Understanding the pan-genomic SV landscape will help breeders develop and understand heterotic pools, introduce and manage novel diversity

Department of Plant Breeding, JLU Giessen

Rapeseed/Brassica Research Team, 2019

Christian Obermeier, Andreas Stahl, Benjamin Wittkop Jenny Lee, Sarah Schiessl-Weidenweber, Annaliese Mason

Iulian Gabur, Harmeet Chawla, Isabelle Deppé, Katharina Tyson, Paul Vollrath, Andreas Eckert, Daniela Quezada, Elizabeth Katche, Elvis Katche, Paula Vasquez

Petra Degen, Birgit Keiner, Annette Plank, Stavros Tzigos, Regina Illgner, Sabine Frei, Liane Renno, Ingrid Schneider-Huth, Juliette Kellermann, Anja Pöltl, Mario Tolksdorf, Horst Schaub, Karl-Heinz Balzer, Lothar Behle-Schalk