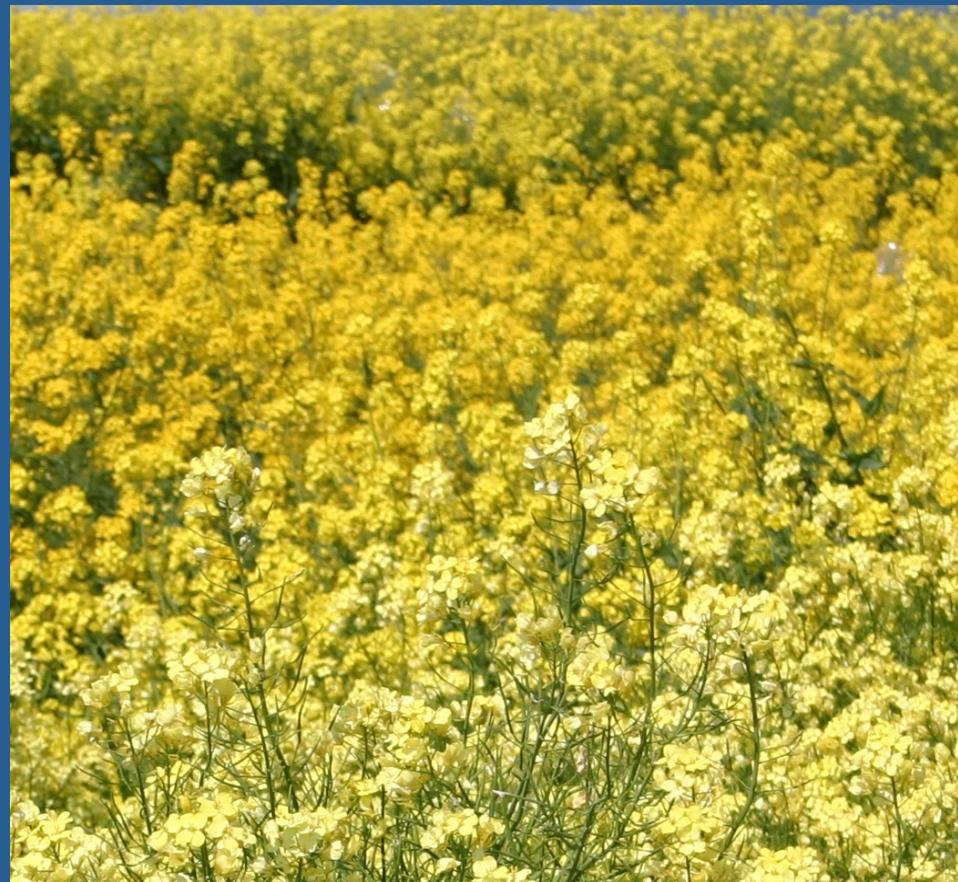
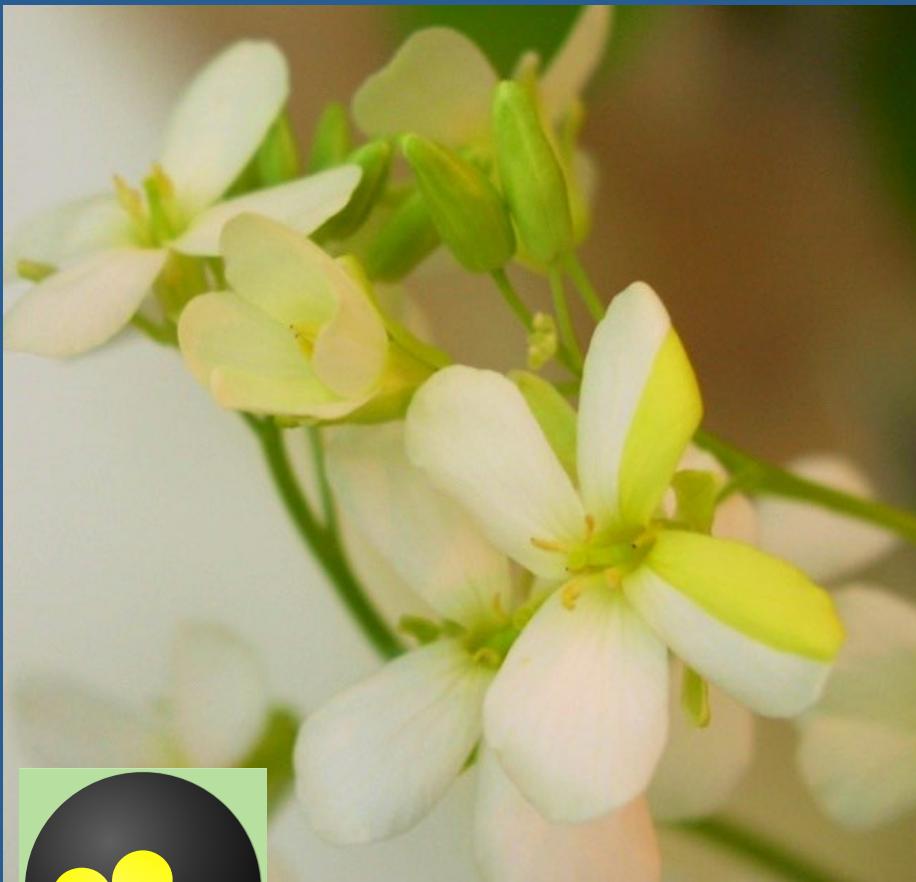


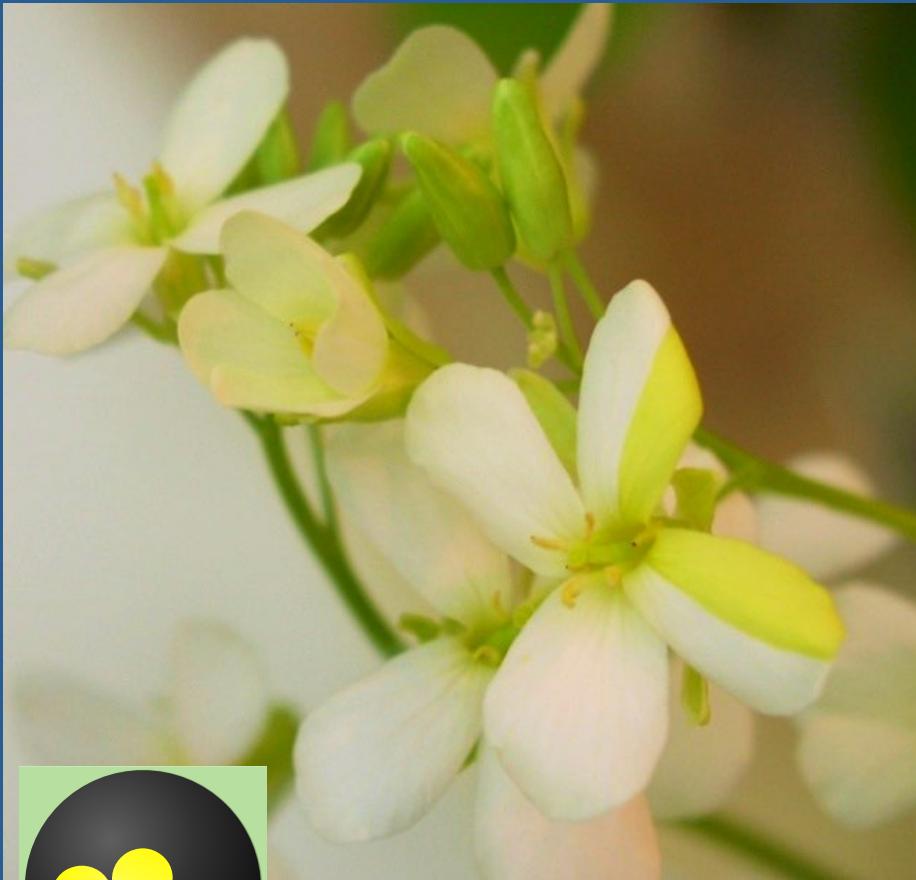
# Remnants of recent polyploidy shape trait variation in *Brassica napus*



Rod Snowdon

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

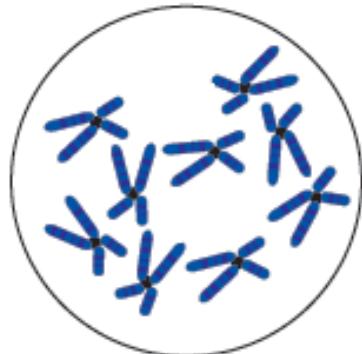
# AA + CC $\neq$ AAC



Rod Snowdon

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

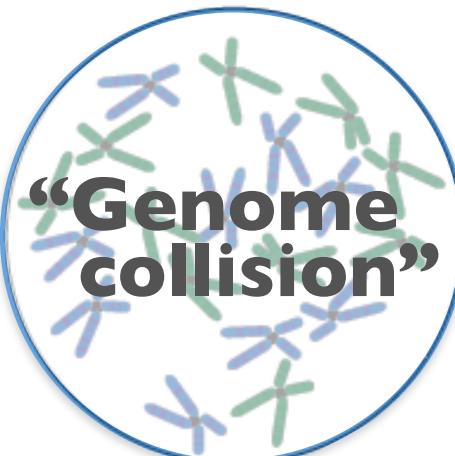
# *Brassica napus*: A recent, anthropogenic allopolyploid



*Brassica oleracea*  
 $2n = 2x = 18$   
Genome CC



Mediterranean  
cabbages and  
kales

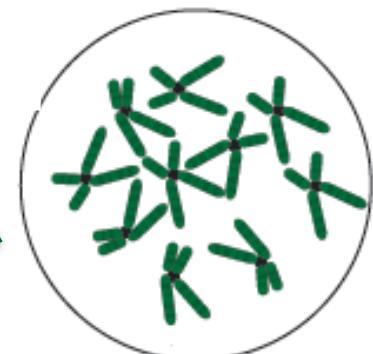


*Brassica napus*  
 $2n = 4x = 38$   
Genome AACCC



**Canola/rapeseed, kale & swede/rutabaga**

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



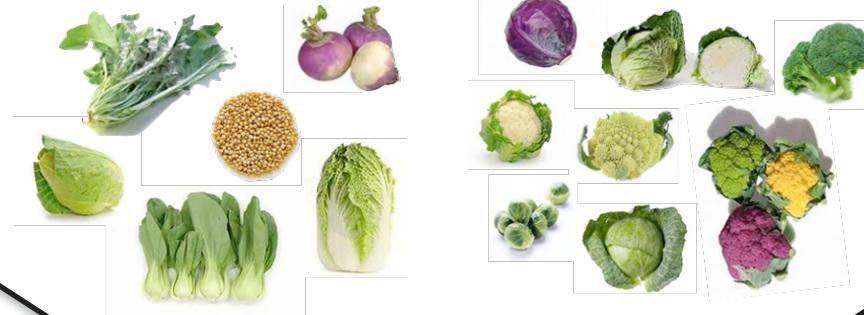
*Brassica rapa*  
 $2n = 2x = 20$   
Genome AA



Asian cabbages,  
Sarson and turnips

# “The Cinderella crop” – a paradoxical fairy-tale?

## Diploid progenitor species



Strong allopolyploidisation bottleneck

***De novo* allopolyploid**

Very few species founders

**Seed quality selection**

Extreme breeding bottlenecks

Extreme erosion of genetic diversity essential for breeding

However: Unexpected adaptive capacity and breeding success

# Learning from the *Brassica napus* genome collision

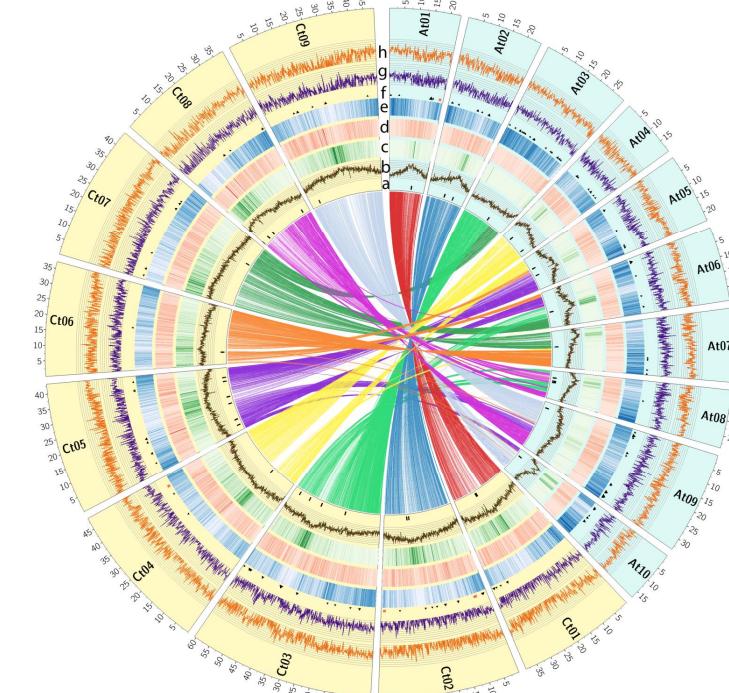
## Retrospective learning

- How did different oilseed *B. napus* forms achieve sufficient *de novo* diversity (despite their enormous polyploidisation and quality selection bottlenecks) to adapt to completely new environments and become a hugely successful global crop?

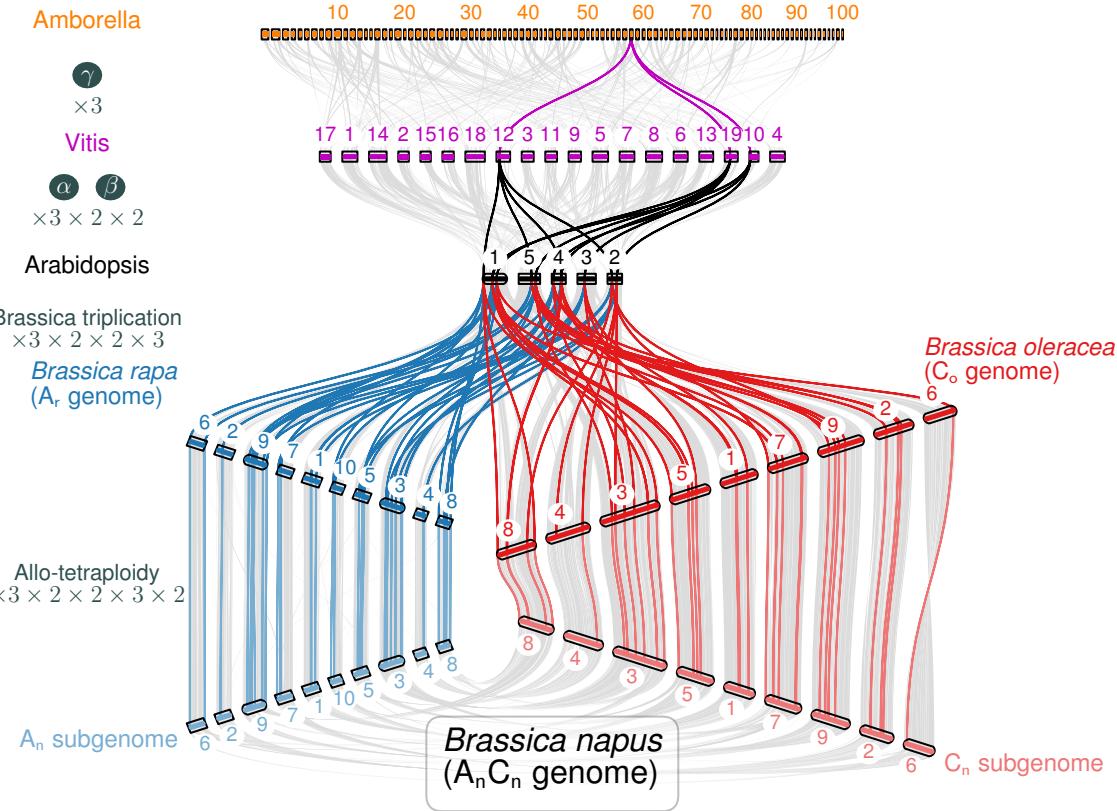
## Breeding applications

- Can the answers to this question help us to more efficiently find and reinstate diversity for adaptation to new challenges (e.g. climate change, disease, constraints on fertiliser or chemical inputs)?

# The complex, allopolyploid genome of *Brassica napus*



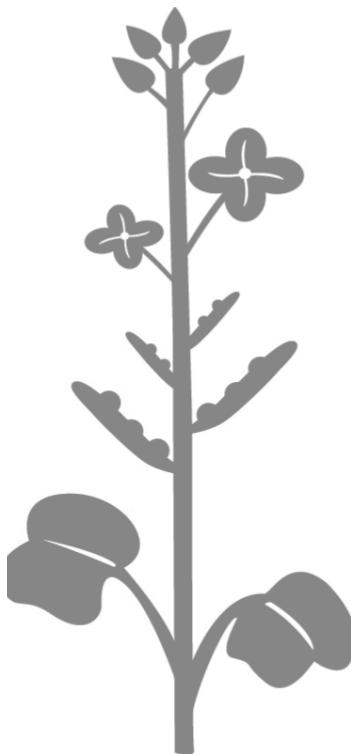
Chalhoub et al., Science (2014)



- Extensive gene duplication and structural genome rearrangement
- Complex patterns of chromosome exchanges

# High-throughput genotyping in a complex polyploid

**Brassica 60k SNP Infinium  
consortium array (led by Isobel  
Parkin, released 2012)**



Clarke et al., TAG (2016)

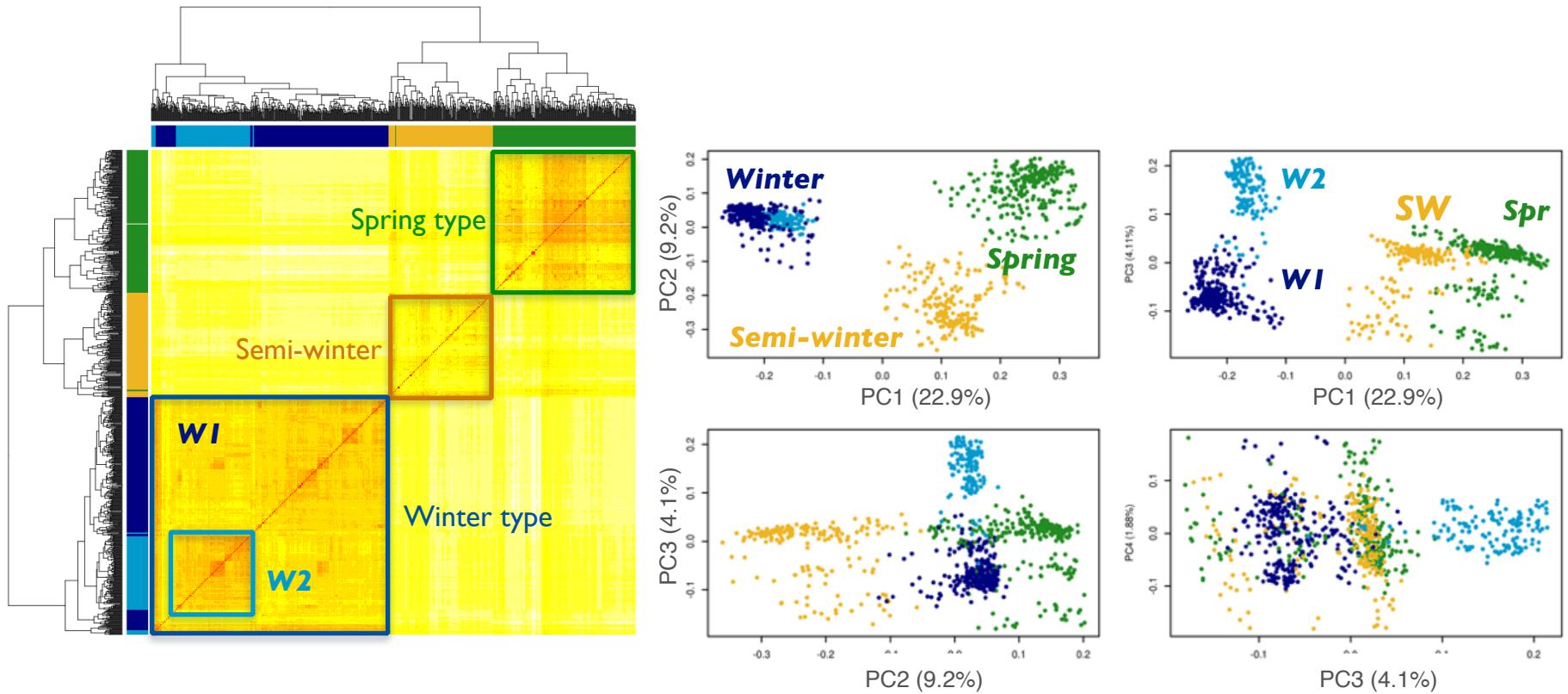
- Highly successful community-designed SNP genotyping array
- Widely used in rapeseed/canola genetics and breeding

## However:

- ~40% of the Infinium assays address multiple loci (hemi-SNPs)
- Hence, most downstream analysis (mapping, GWAS, GS etc.) works best with a filtered panel of ~25,000 single-locus SNPs

# Disturbingly low SNP diversity in major breeding pools?

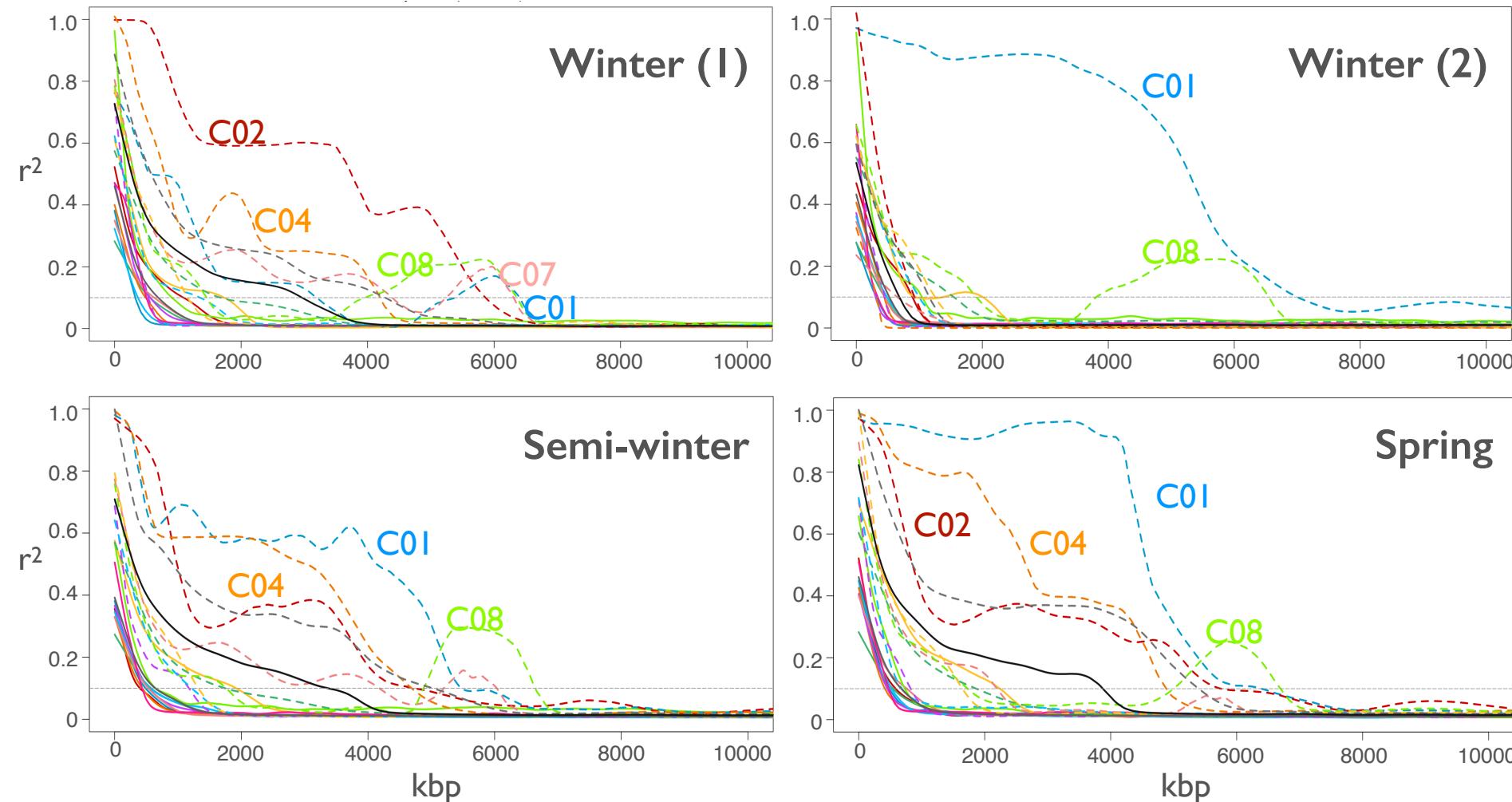
Genetic diversity analysis in a species-wide germplasm collection (n~850) based on ~20,000 single-copy, polymorphic SNPs



- SNP diversity is relatively high between ecogeographic pools, but appears to be very low within the major gene pools

# Chromosomal and genomic bias in allelic diversity

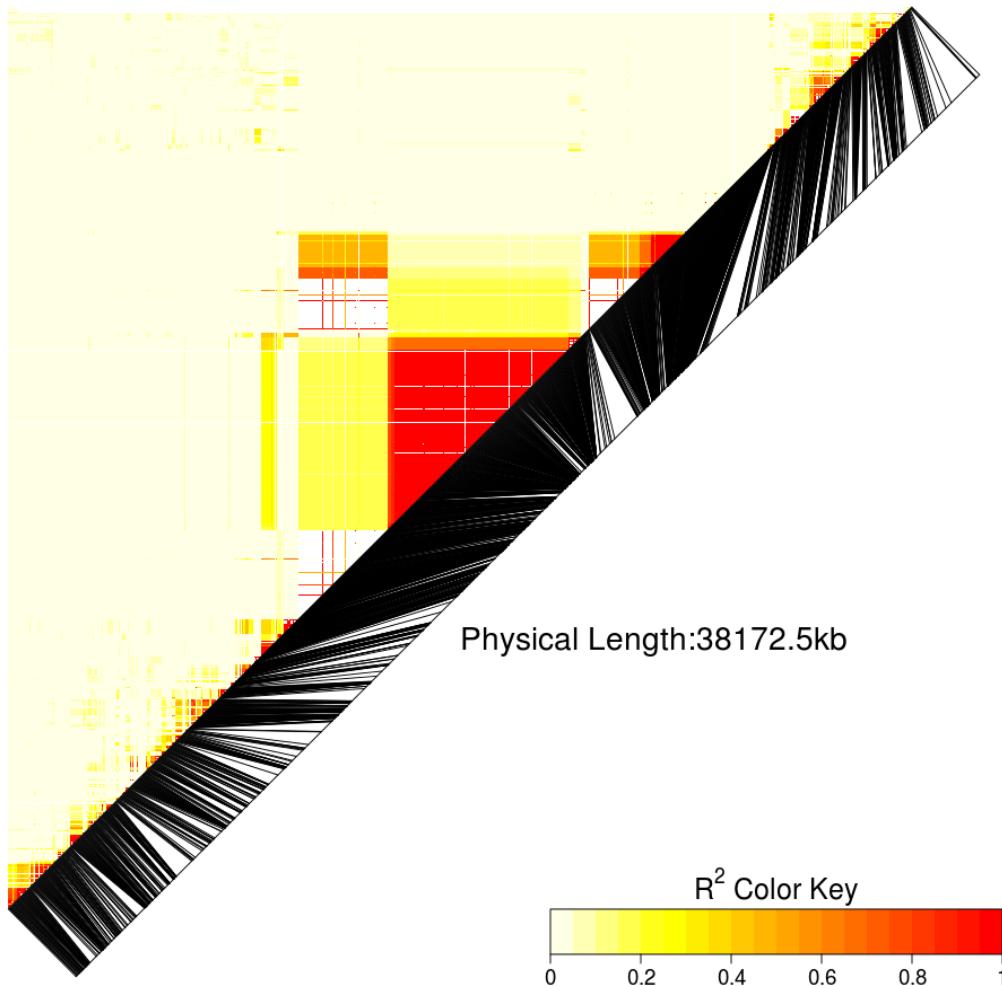
## Decay of linkage disequilibrium in different *B. napus* breeding pools



Qian et al. *BMC Genomics* (2014), Jan et al. *PLOSone* (2015), Schiessl et al. *Front Plant Sci* (2014)

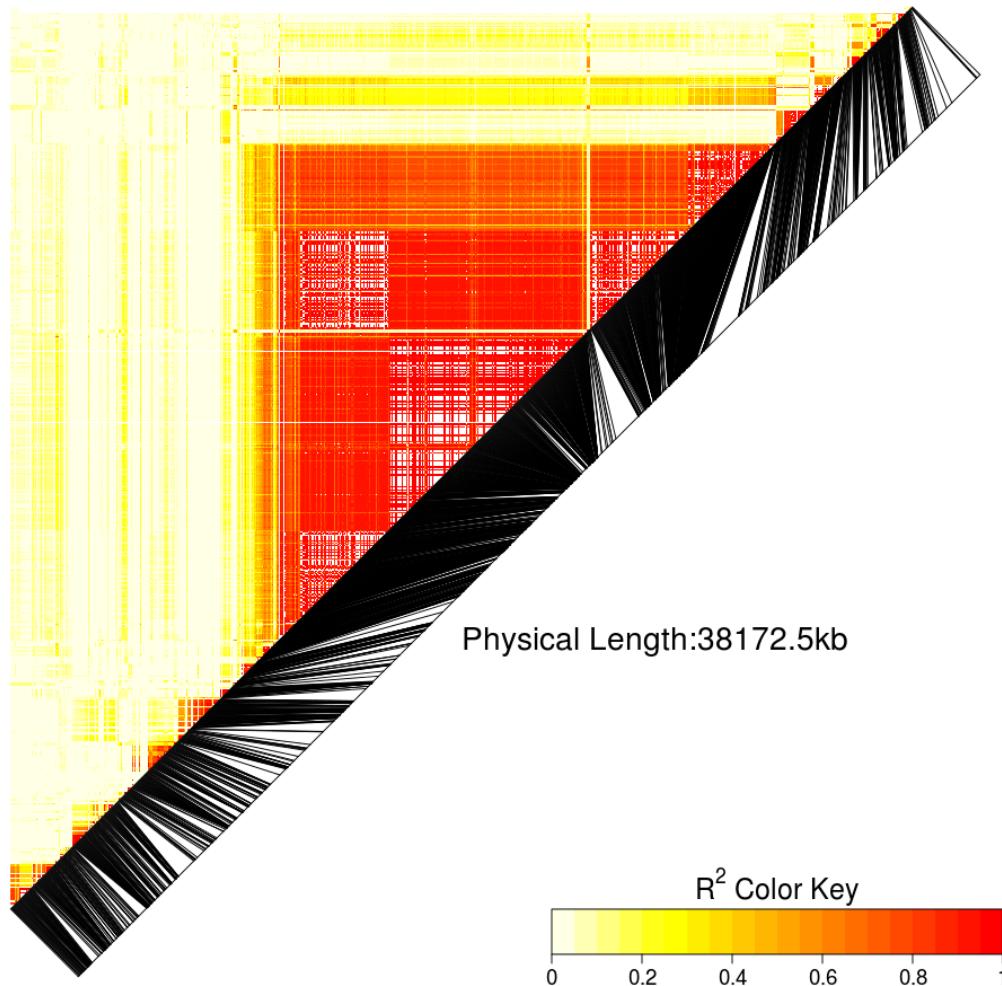
# Extreme selection signatures on some chromosomes

Chromosome C02, winter-type (I)



# Extreme selection signatures on some chromosomes

Chromosome C02, winter-type (2)



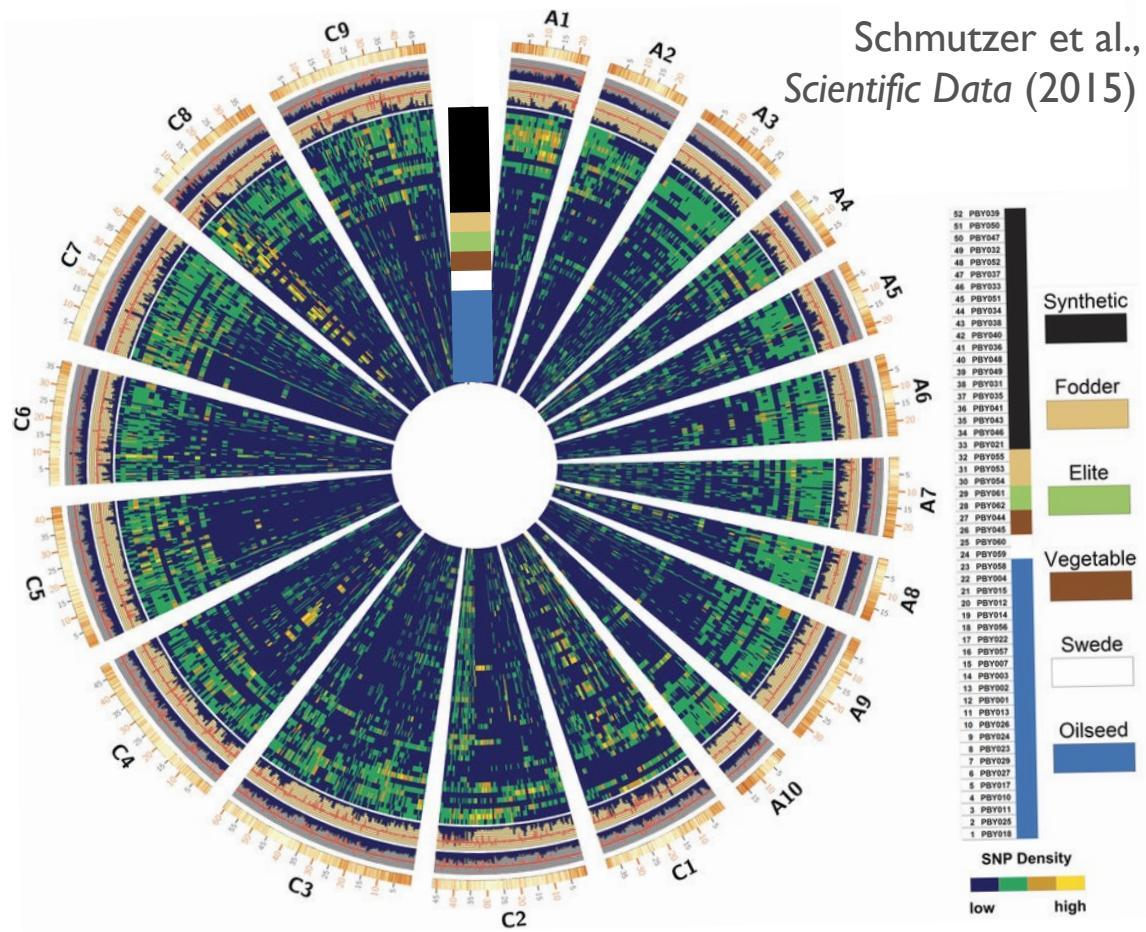
# Genomic sequencing: SNP diversity actually not so bad?

## German PreBreed-Yield BnNAM Consortium

20 highly diverse synthetic &  
32 exotic, natural *B. napus*  
(BnNAM founders)

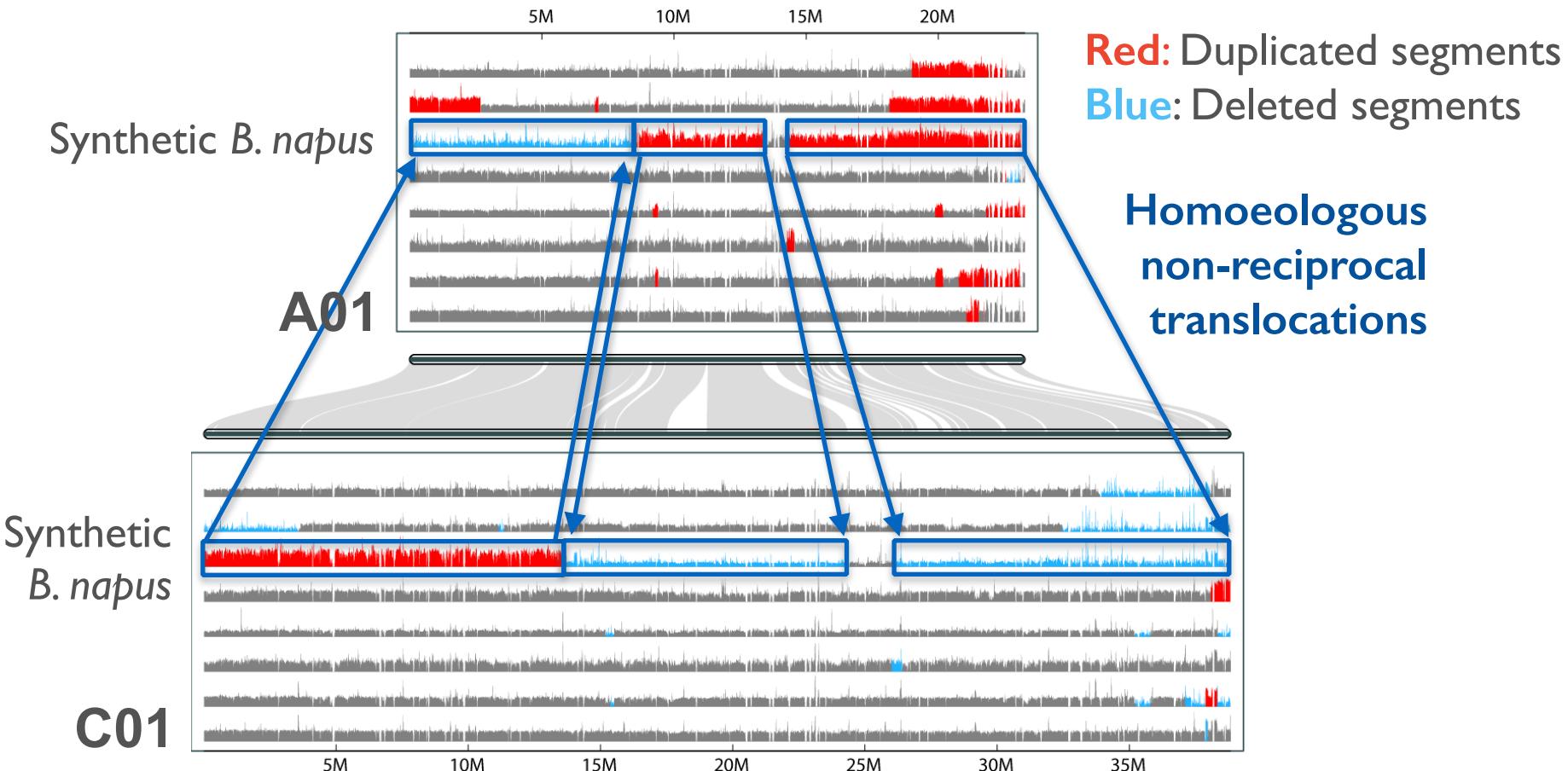
~15x coverage, 100bp PE

~4M high-quality SNPs



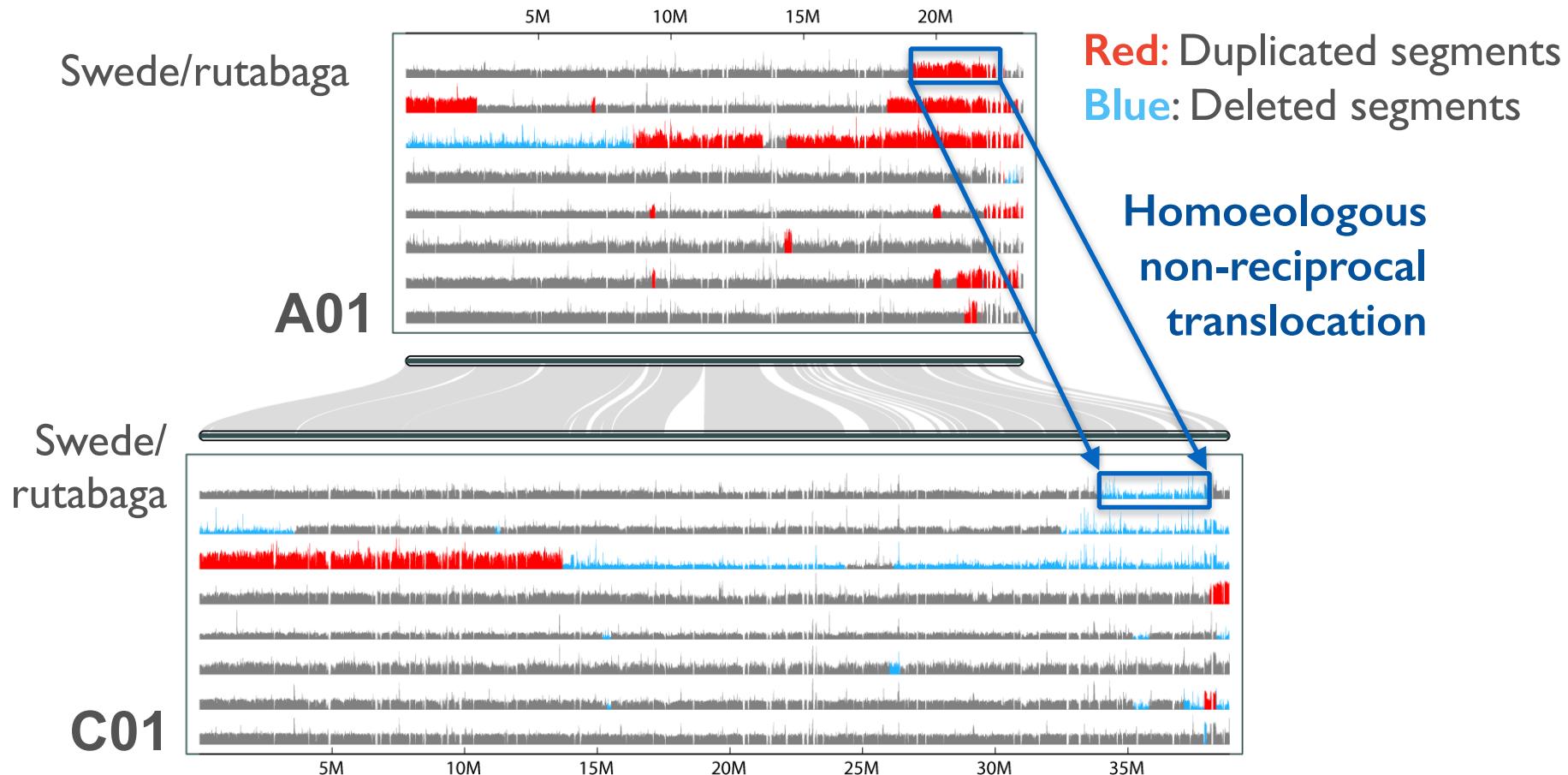
- Contradiction between array and sequence data: Is decisive sequence variation obscured in duplicated genes/genome regions?

## *B. napus*: Widespread structural genome variation



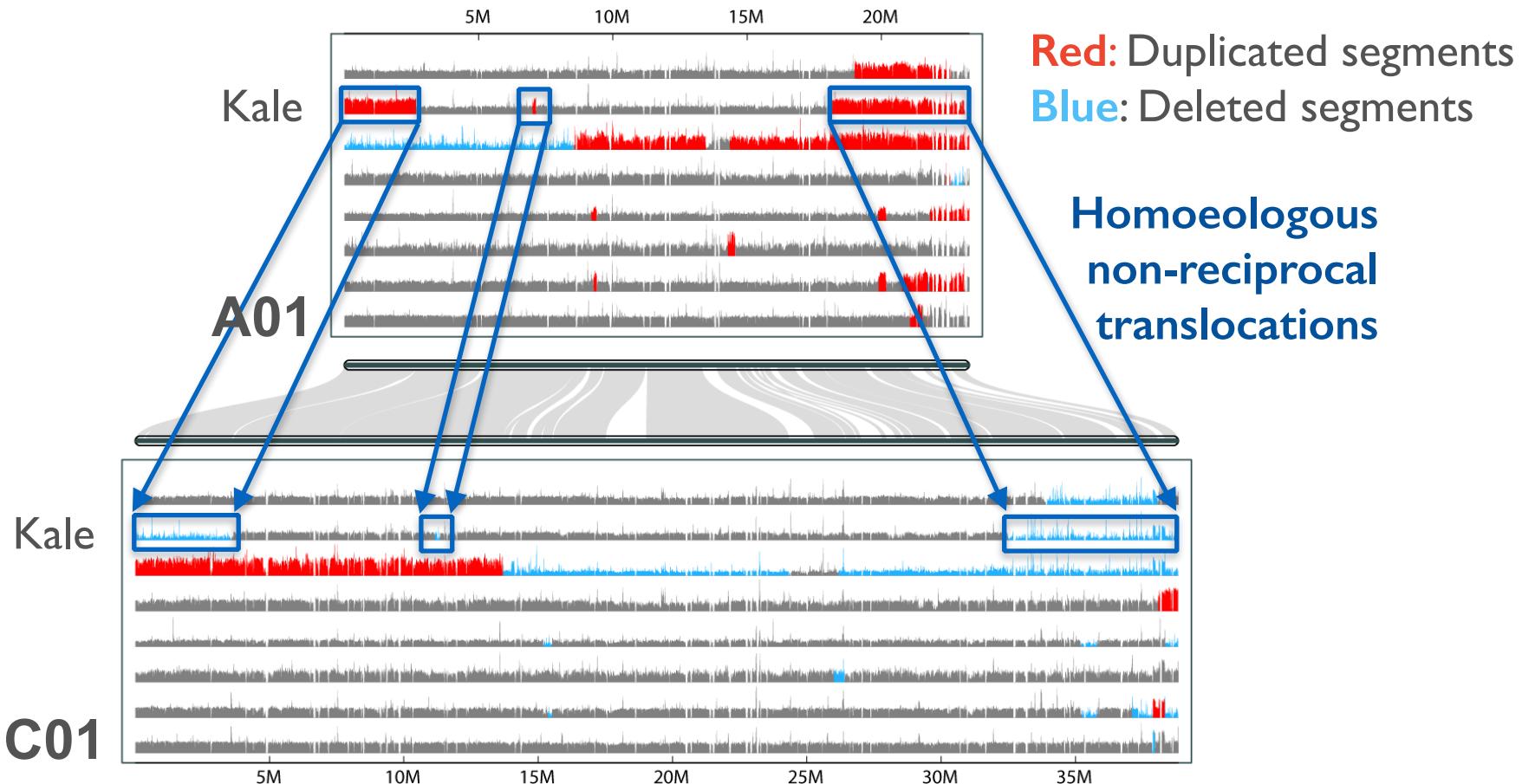
Chalhoub et al., Science (2014)

## B. napus: Widespread structural genome variation



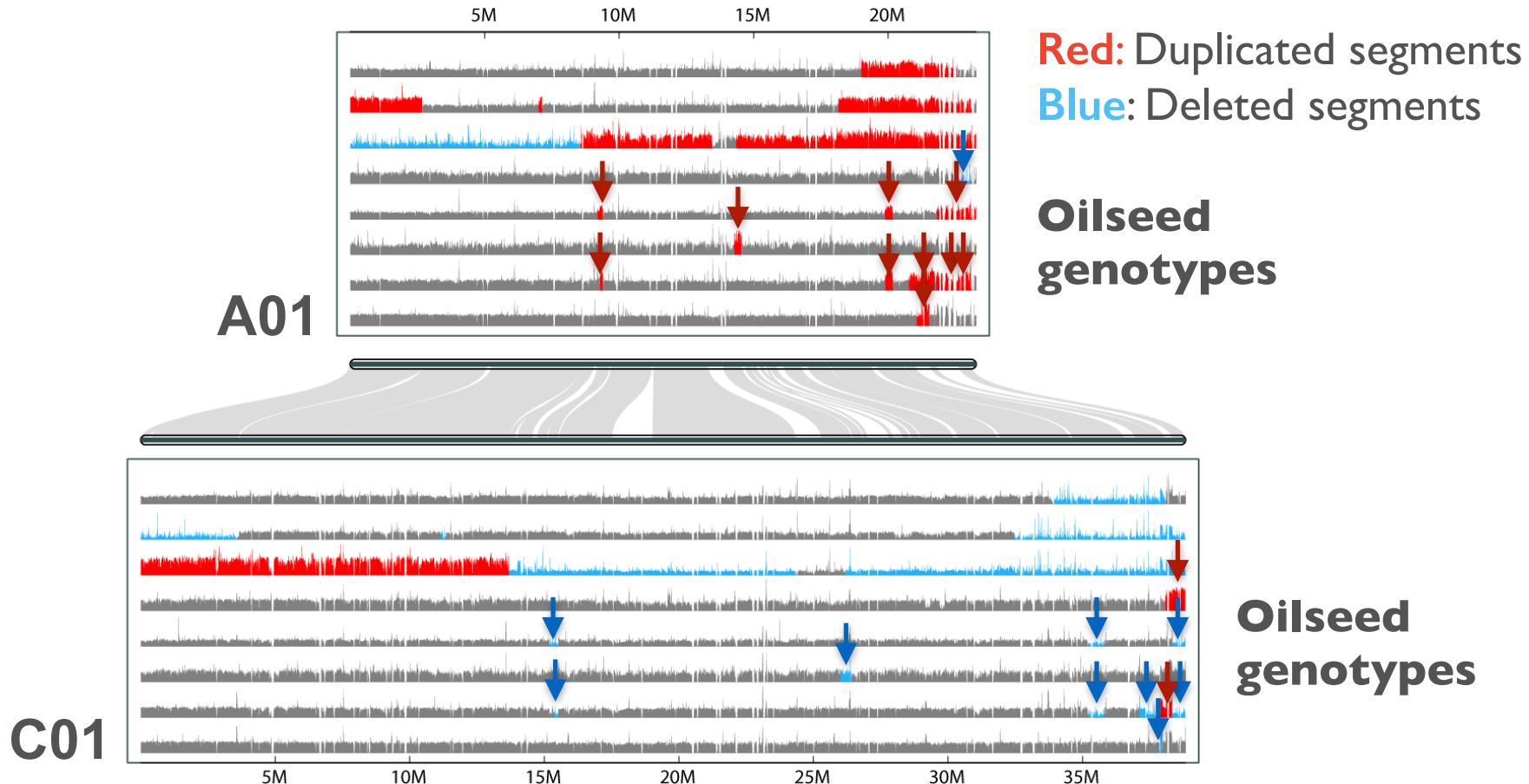
Chalhoub et al., Science (2014)

## B. napus: Widespread structural genome variation



Chalhoub et al., Science (2014)

## B. napus: Widespread structural genome variation

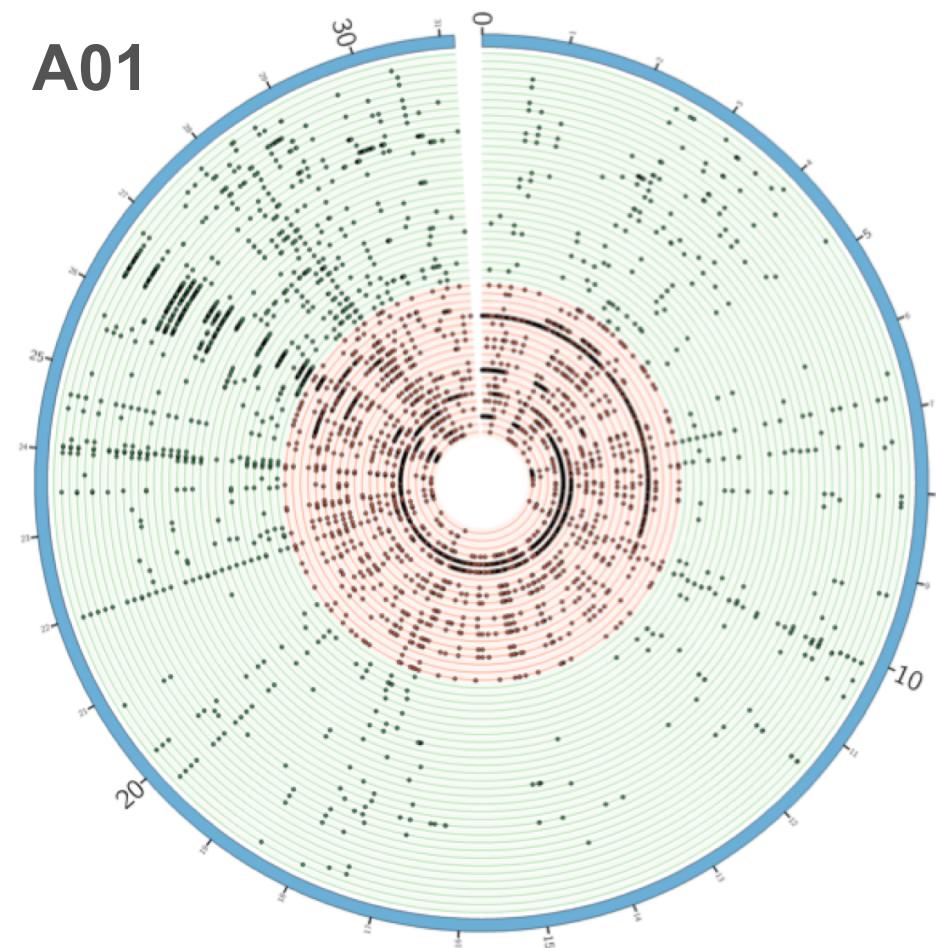


Chalhoub et al., Science (2014)

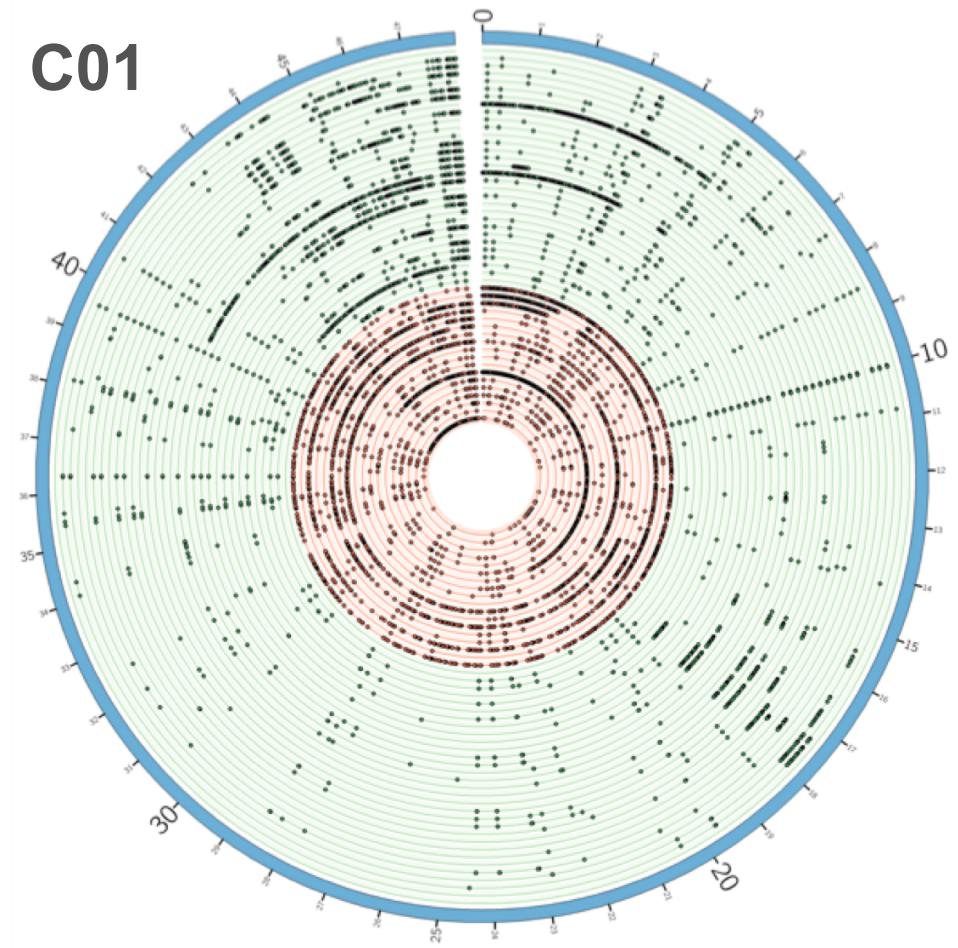
# Consequence: Extensive structural genome variation

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

A01



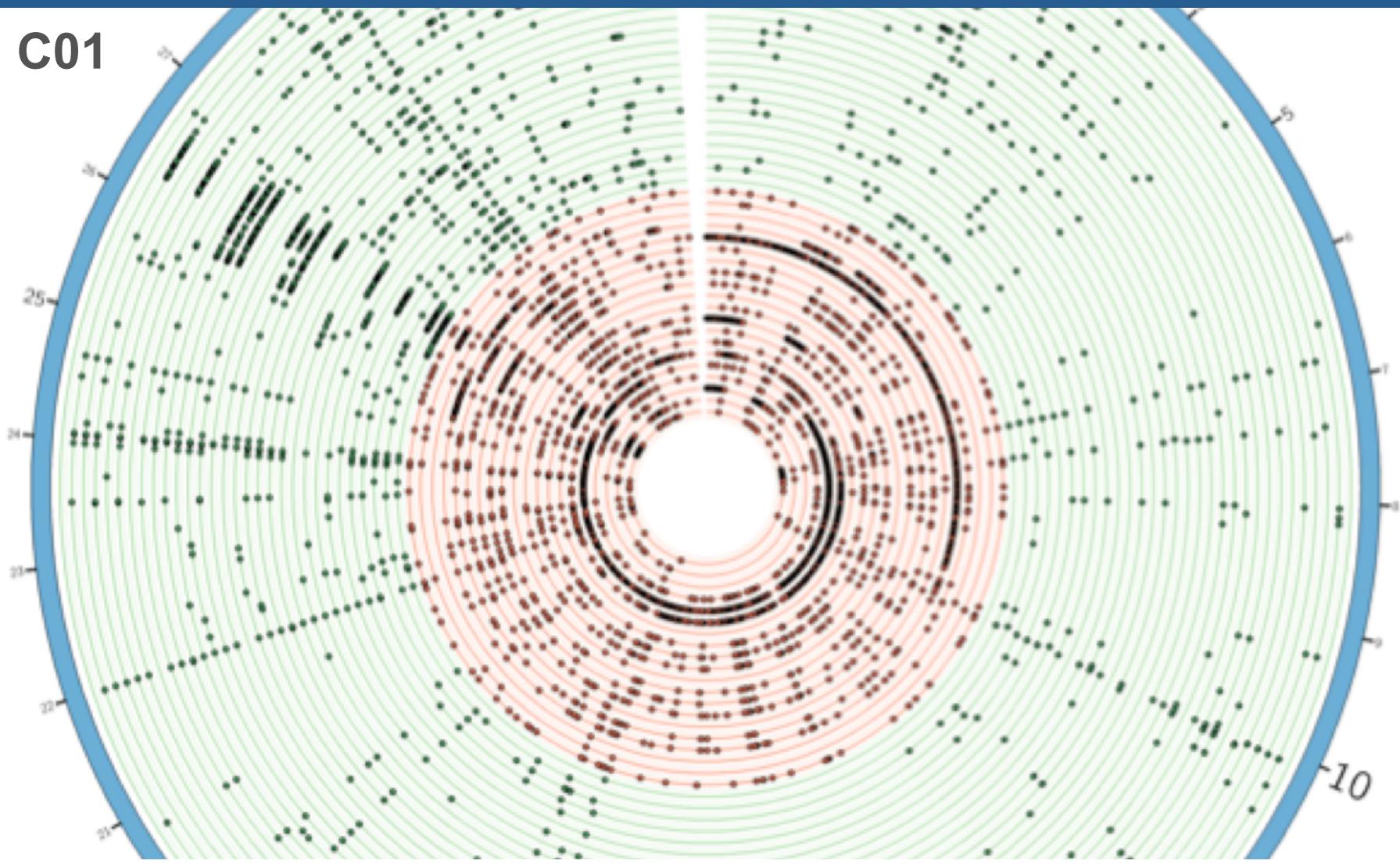
C01



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

# Significant gene presence/absence in all *B. napus*

C01

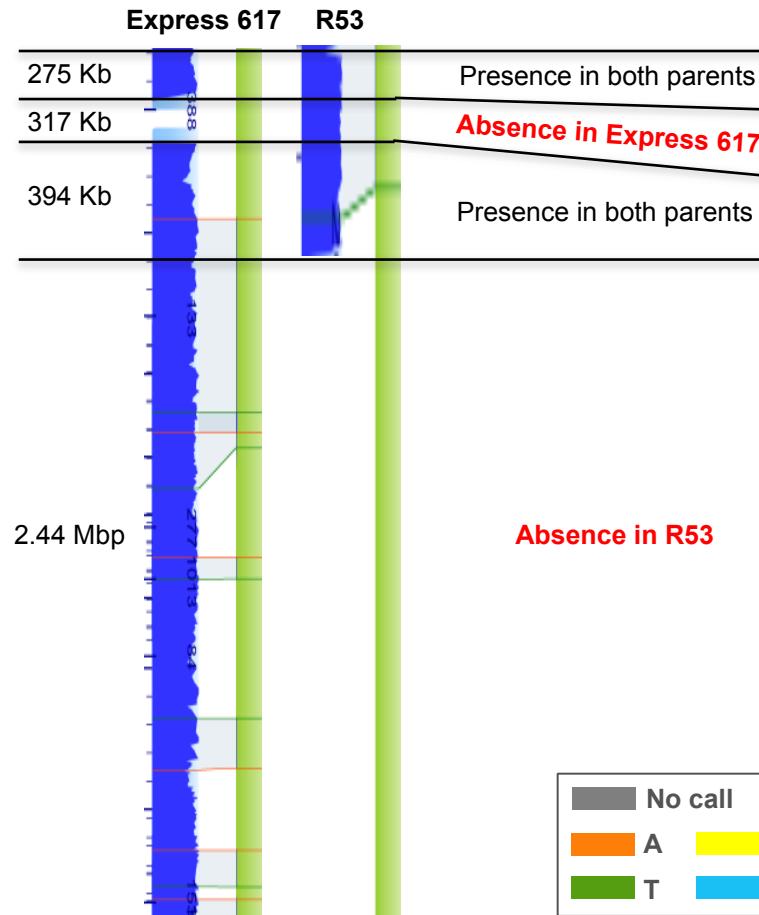


- **Consequence:** Considerable gene presence-absence and copy-number variation

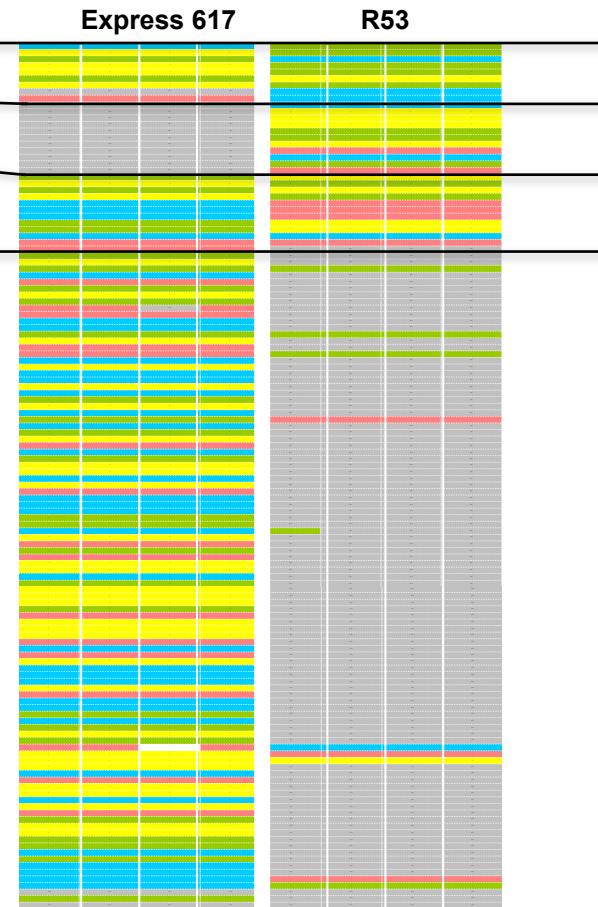
# Validation of large-scale deletions / rearrangements

**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

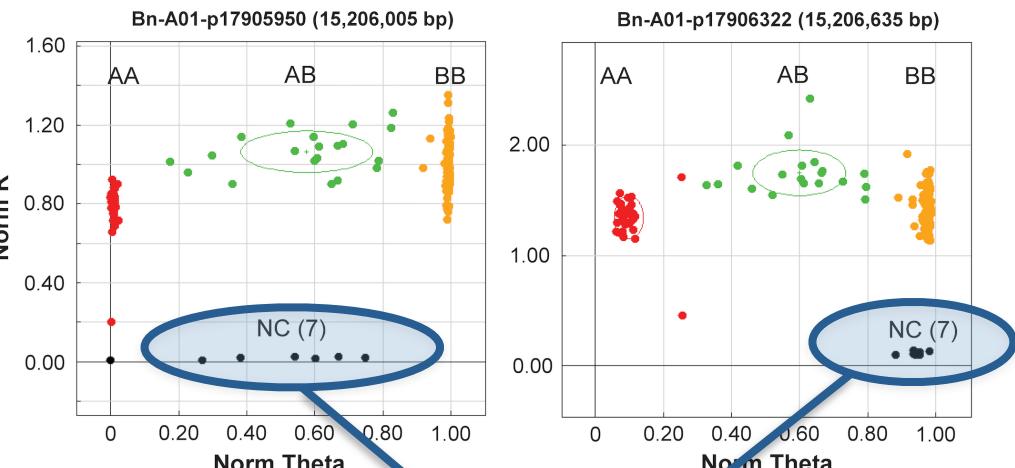


## 2) Reference-ordered SNP loci

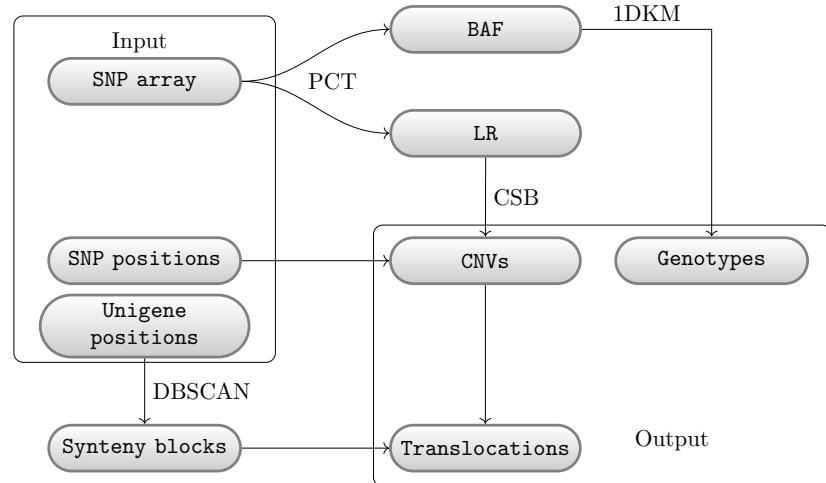


# Identifying structural variants from SNP array data

## gsrc: An R package for genome structure rearrangement calling



Physically adjacent SNPs which consistently “fail” in the same genotypes represent deletions in those genotypes and can be confidently called as **deletions**



1DKM 1-Dimensional K-Means Clustering

BAF B-Allele Frequency

CSB Circular Binary Segmentation

DBSCAN Density-Based Spatial Clustering of Applications with Noise

LR Log R ratios

PCT Polar Coordinates Transformation

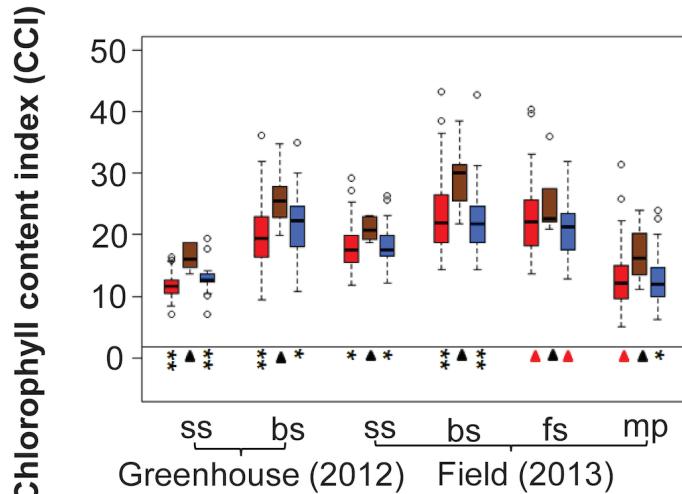
~ 5-20% of assays on **Brassica 60k SNP array**

Grandke et al., *Bioinformatics* (2016)

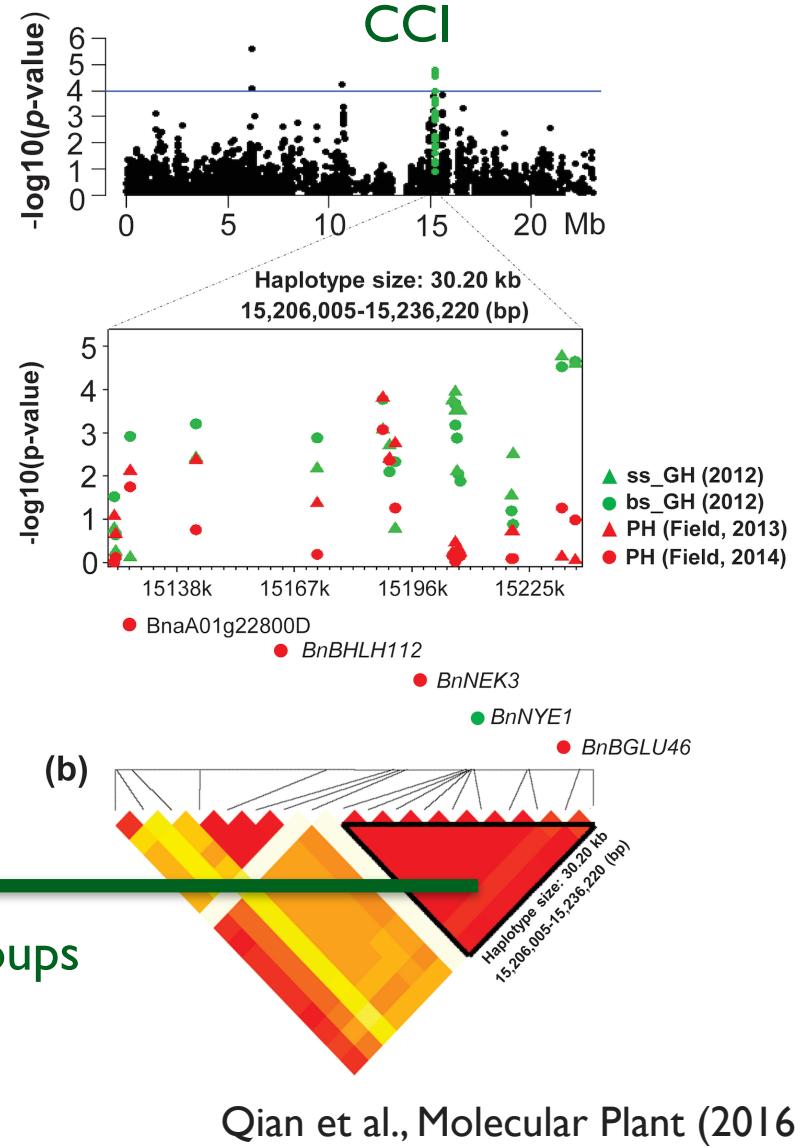
# Deletions associate with selection patterns

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

	Haplotype	No. of accessions									
		G	C	G	G	G	G	C	C	A	T
CCI-Hap1		G	C	G	G	G	G	C	C	A	T
CCI-Hap2		-	-	-	-	-	-	C	C	T	
CCI-Hap3		T	T	A	A	A	A	T	T	C	C

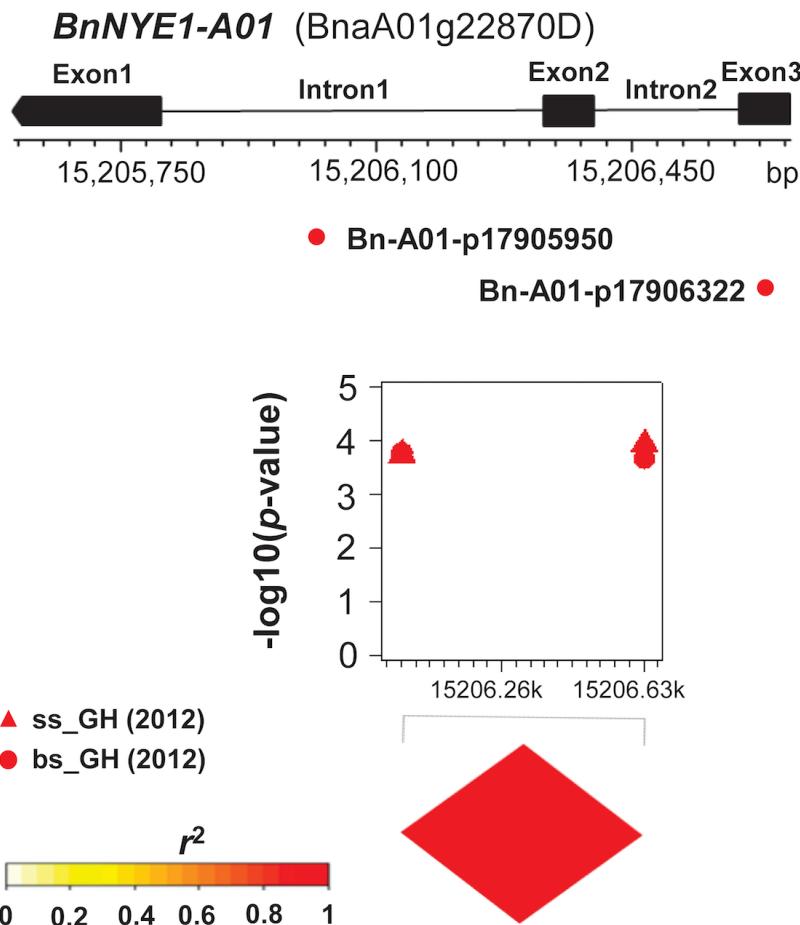


3 haplogroups

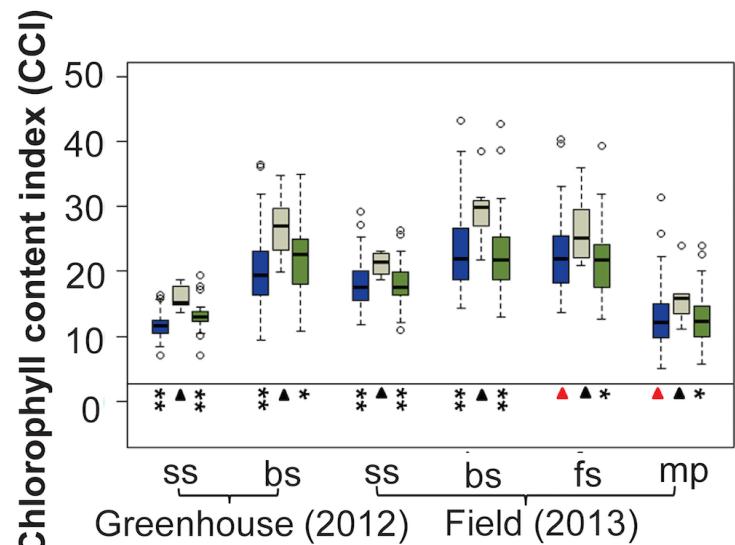


# **Deletions associate with quantitative trait variation**

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



Haplogroups	No. of accessions	
BnNYE1-A01-Hap1	G	C
BnNYE1-A01-Hap2	-	-
BnNYE1-A01-Hap3	T	T

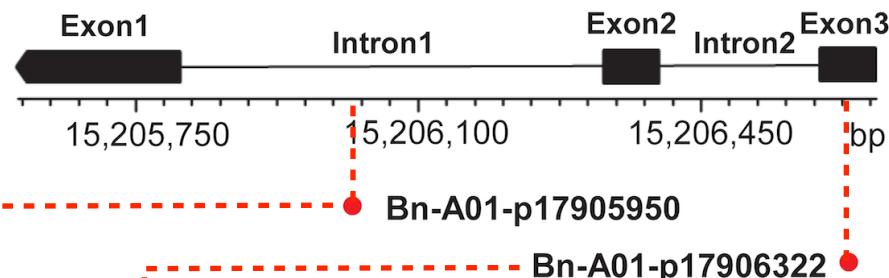
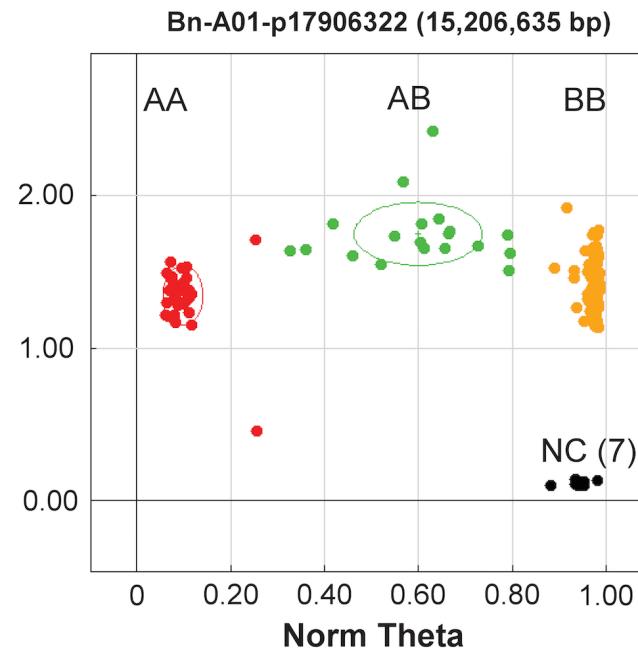
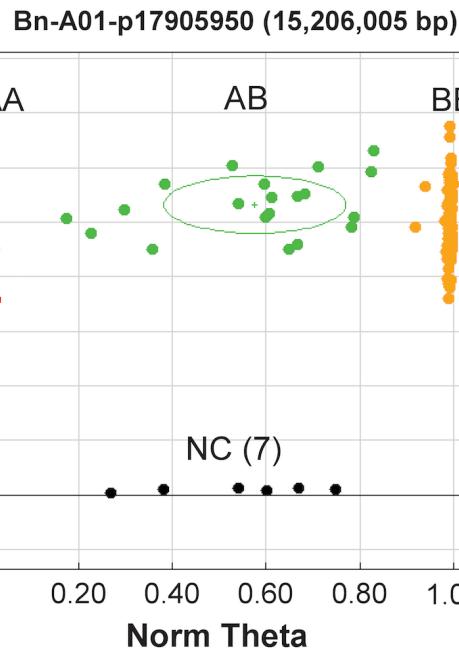


Qian et al., Molecular Plant (2016)

# Identification of deletions in SNP array data

BnNYE I-A0 I

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



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

## **“SNAP markers”**: Single-Nucleotide Absence Polymorphisms

Qian et al., Molecular Plant (2016)

# Presence-absence associates with all sorts of traits

## Chlorophyll content

- *NYE1* gene deletion associated (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 everything else too?

- We are presently revisiting historical datasets for yield, heterosis, germination, vigour, resistance, oil content, protein content, anti nutritive compounds, etc. etc etc to see if missing data (SNAPs) can uncover “invisible” QTL and perhaps help in improving selection procedures for beneficial phenotypes

# Deletions may be a key to polyploid crop adaptation

## Significantly enriched GO terms among genes affected by PAV

Font size proportional to  $-\log(p)$

meiotic DNA double-strand break formation  
regulation of cellular metabolic process  
transport of virus in host, tissue to tissue  
regulation of defense response by callose deposition  
cellular macromolecule metabolic process

defense response to bacterium  
DNA topological change translation  
sesquiterpene biosynthetic process  
immune response-regulating signaling pathway

defense response signaling pathway

defense response to virus

**plant-type hypersensitive response**

nucleic acid metabolic process  
regulation of innate immune response  
sesquiterpenoid biosynthetic process

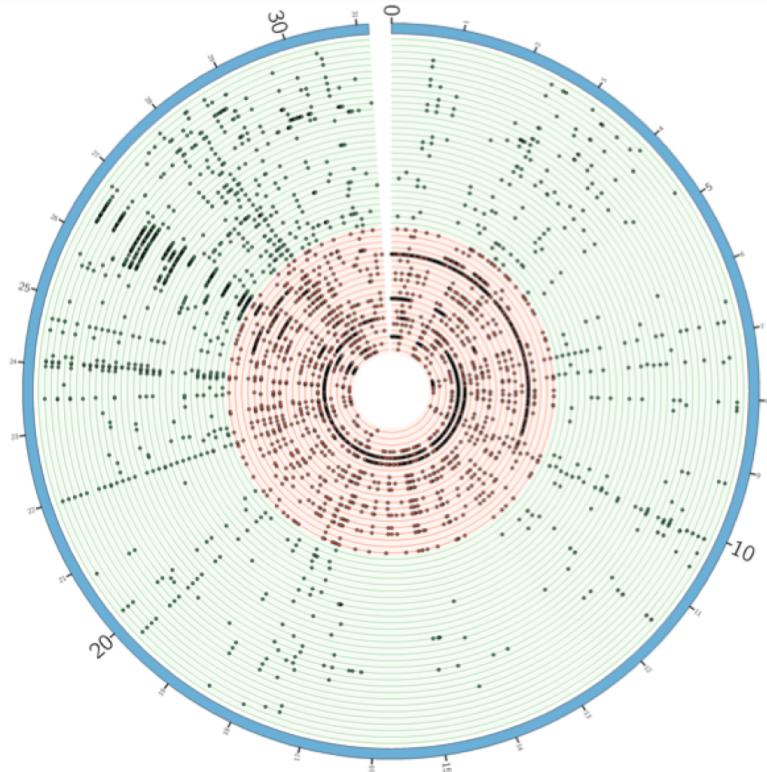
RNA 5'-end processing monocarboxylic acid transport  
cAMP biosynthetic process

single-organism process protein deglycosylation  
positive regulation of defense response

acetyl-CoA biosynthetic process

positive regulation of transport response to virus

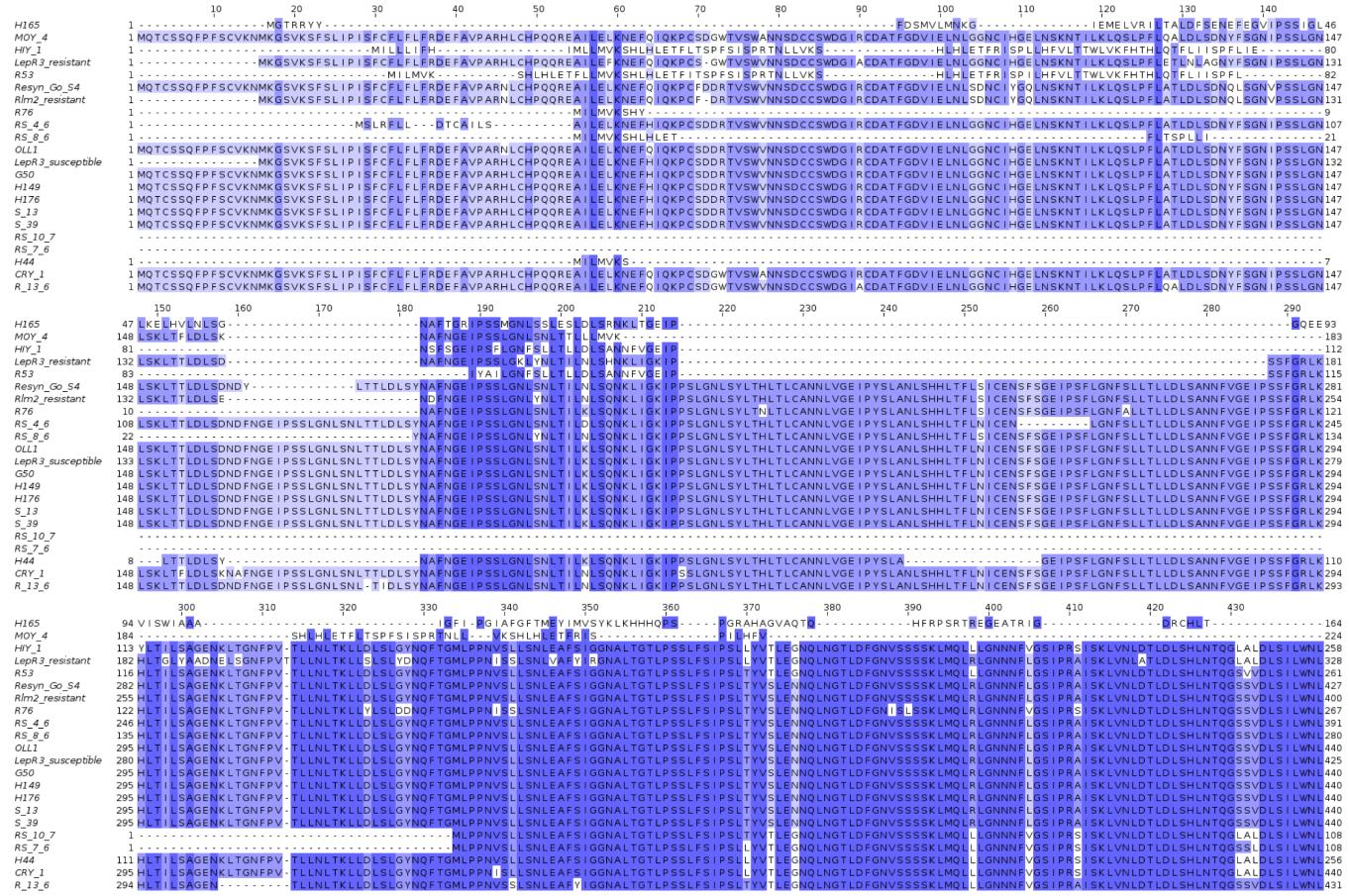
cGMP biosynthetic process



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

# What is the influence of PAV on disease resistance loci?

**Example:** Deletions within the blackleg R-gene *RLM2* are common in synthetic *Bn*



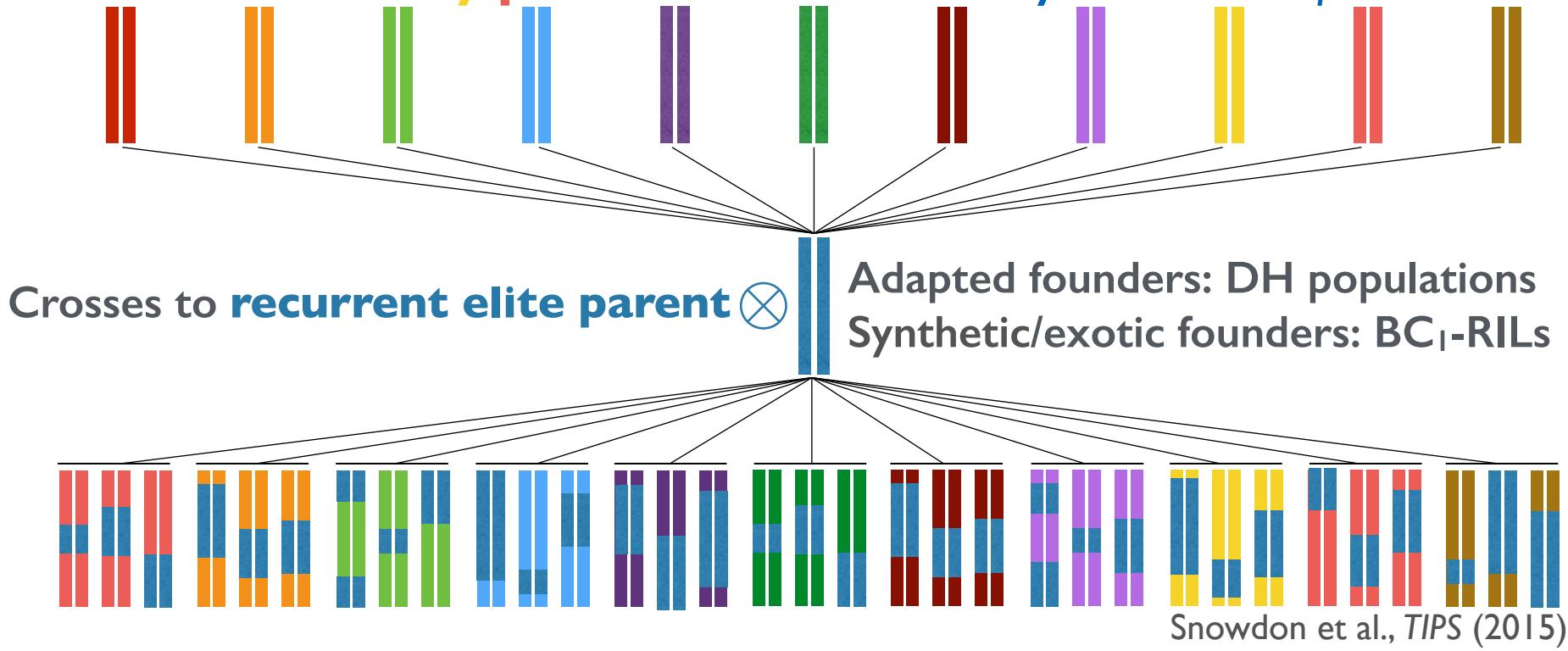
Hurgobin et al. Plant Biotech J (2017)

**Question:** What influence do deletions have on quantitative disease resistance?

# Associating missing data (SNAPs) to resistance QTL

## Nested Association Mapping populations

Genetic diversity panel – 30 natural & 20 synthetic *B. napus* founders

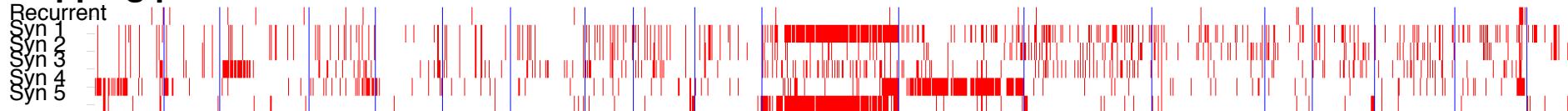


### Resistance screening in NAM subfamilies with quantitative resistance:

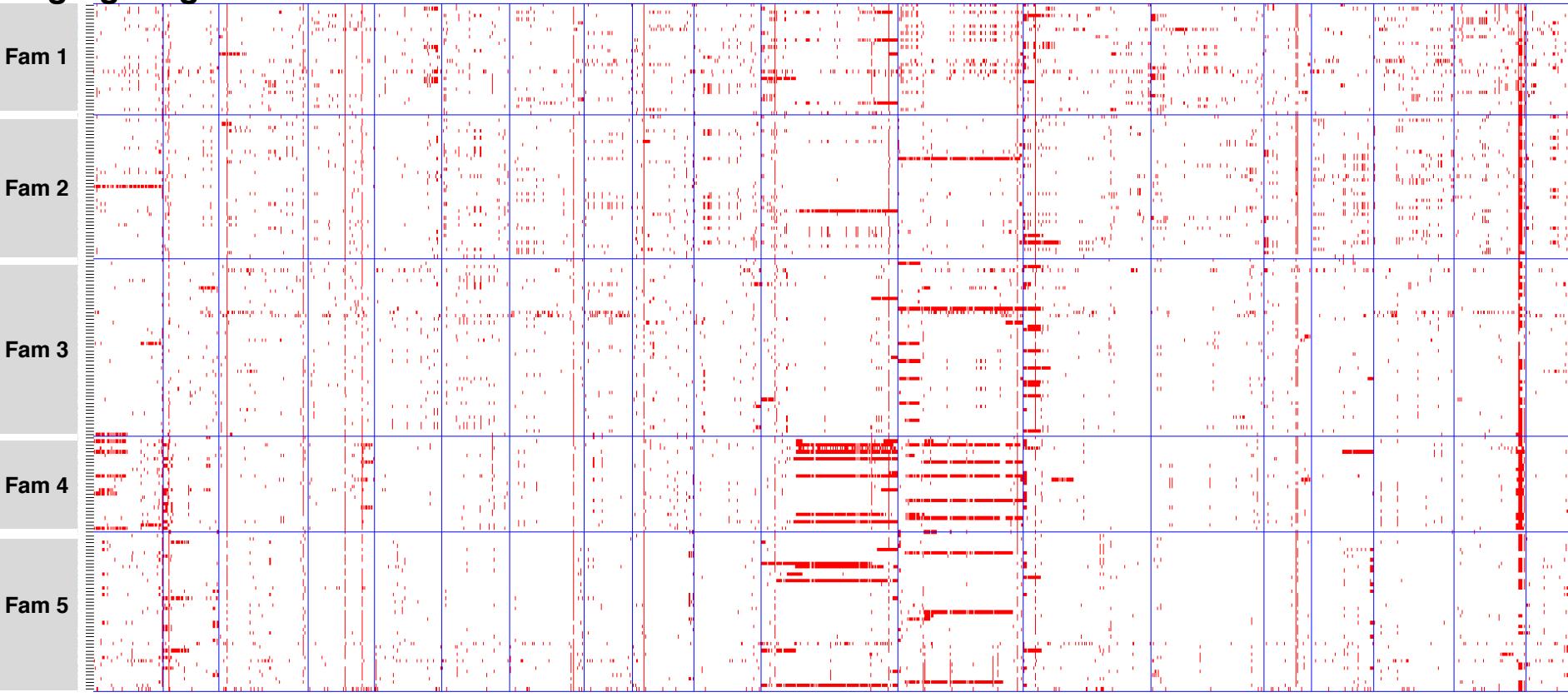
- QTL analysis in 5 subfamilies derived from synthetics carrying multiple adult plant resistances to Verticillium, Sclerotinia, blackleg and clubroot disease

# Segregation data in NAM families verifies SNAP calls

## Mapping parents



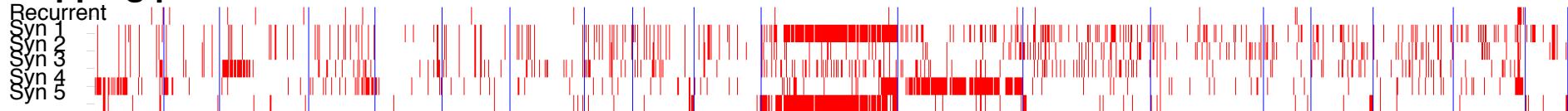
## Segregating NAM families



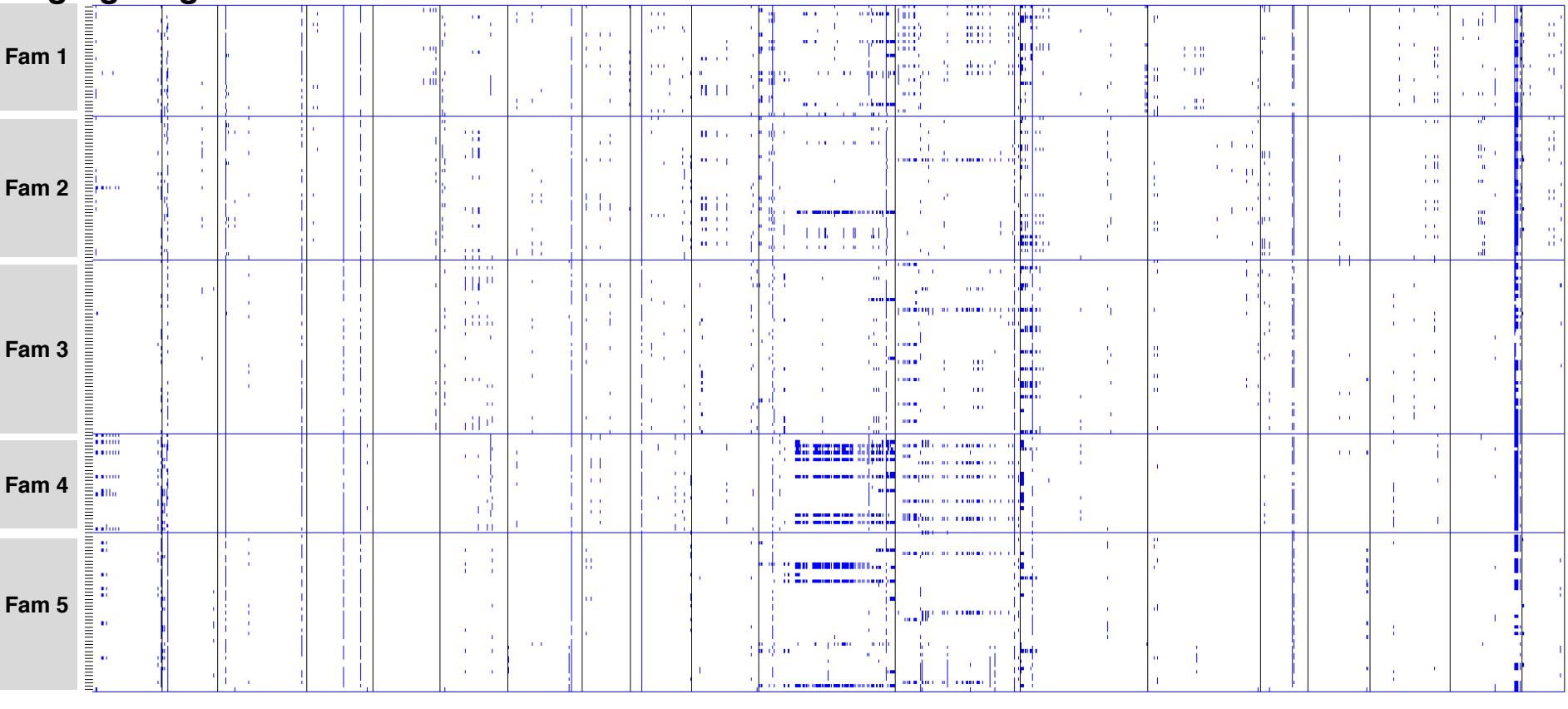
Deletion calls in 5 NAM subfamilies from resistant synthetic parents

# Segregation data in NAM families verifies SNAP calls

## Mapping parents

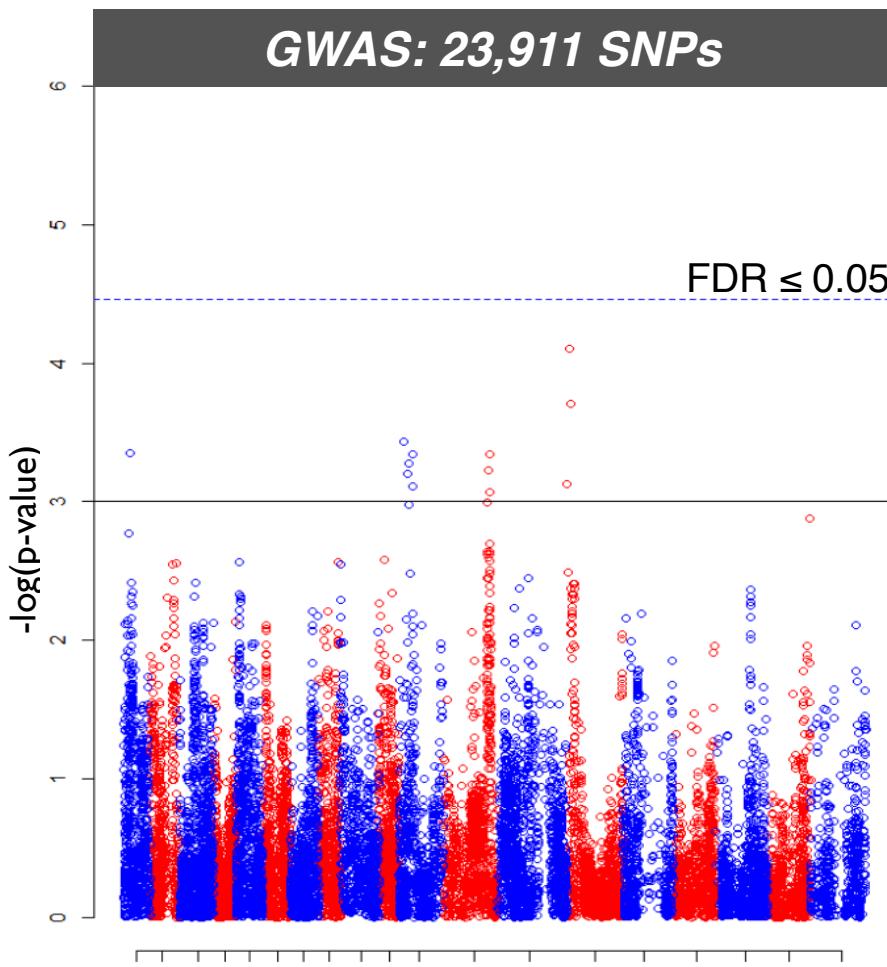


## Segregating NAM families



Deletion calls corrected for expected family-wise segregation frequency

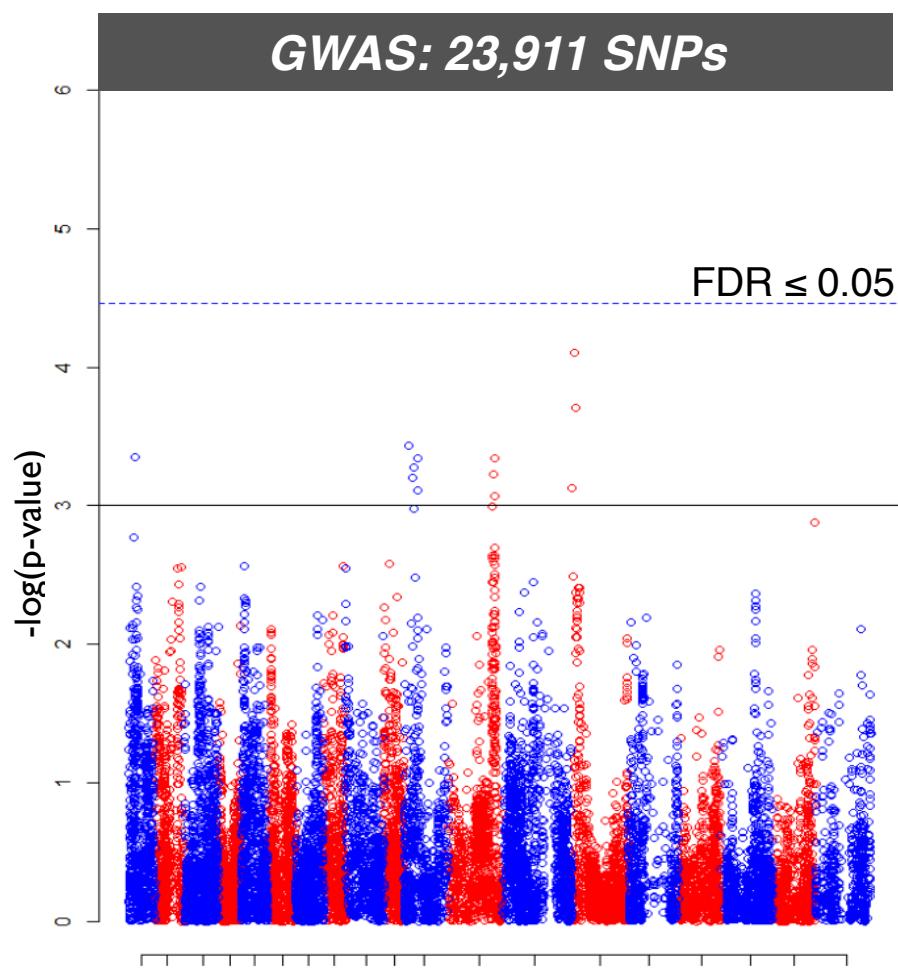
# Combining SNPs & SNAPs improves QTL detection



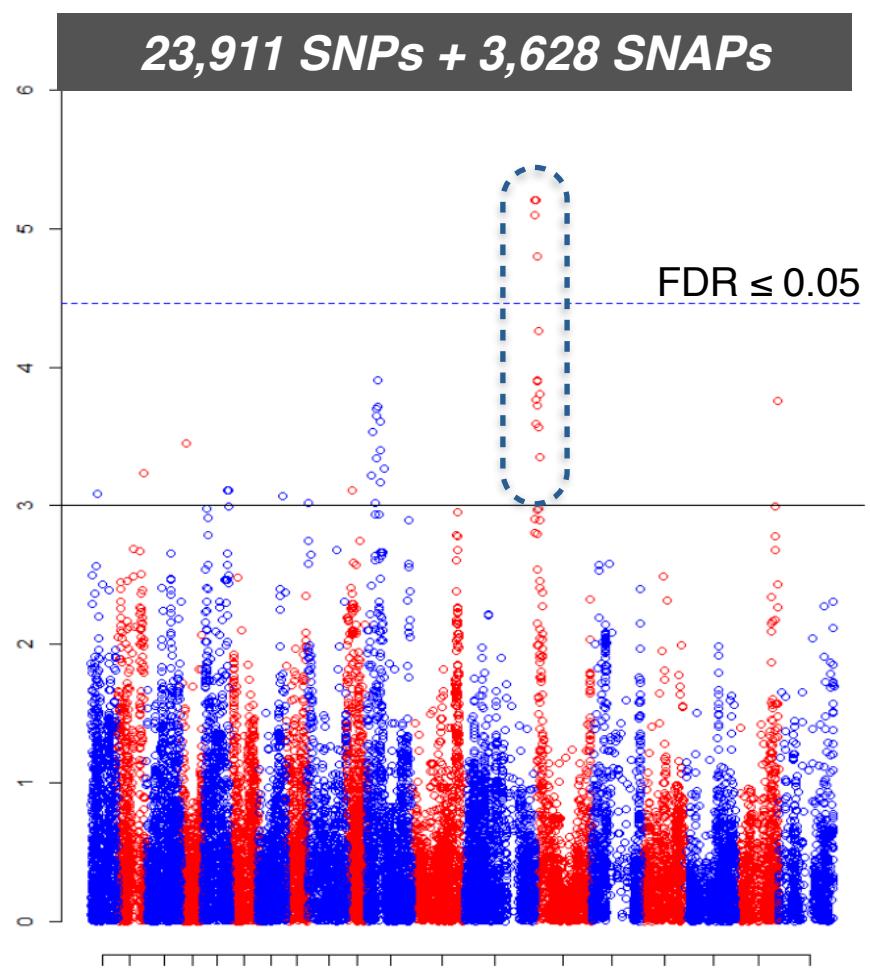
**Example:** Increased detection power and resolution of **QTL for adult plant blackleg resistance** on chromosome C04 by implementation of SNAP markers

# Combining SNPs & SNAPs improves QTL detection

**GWAS: 23,911 SNPs**



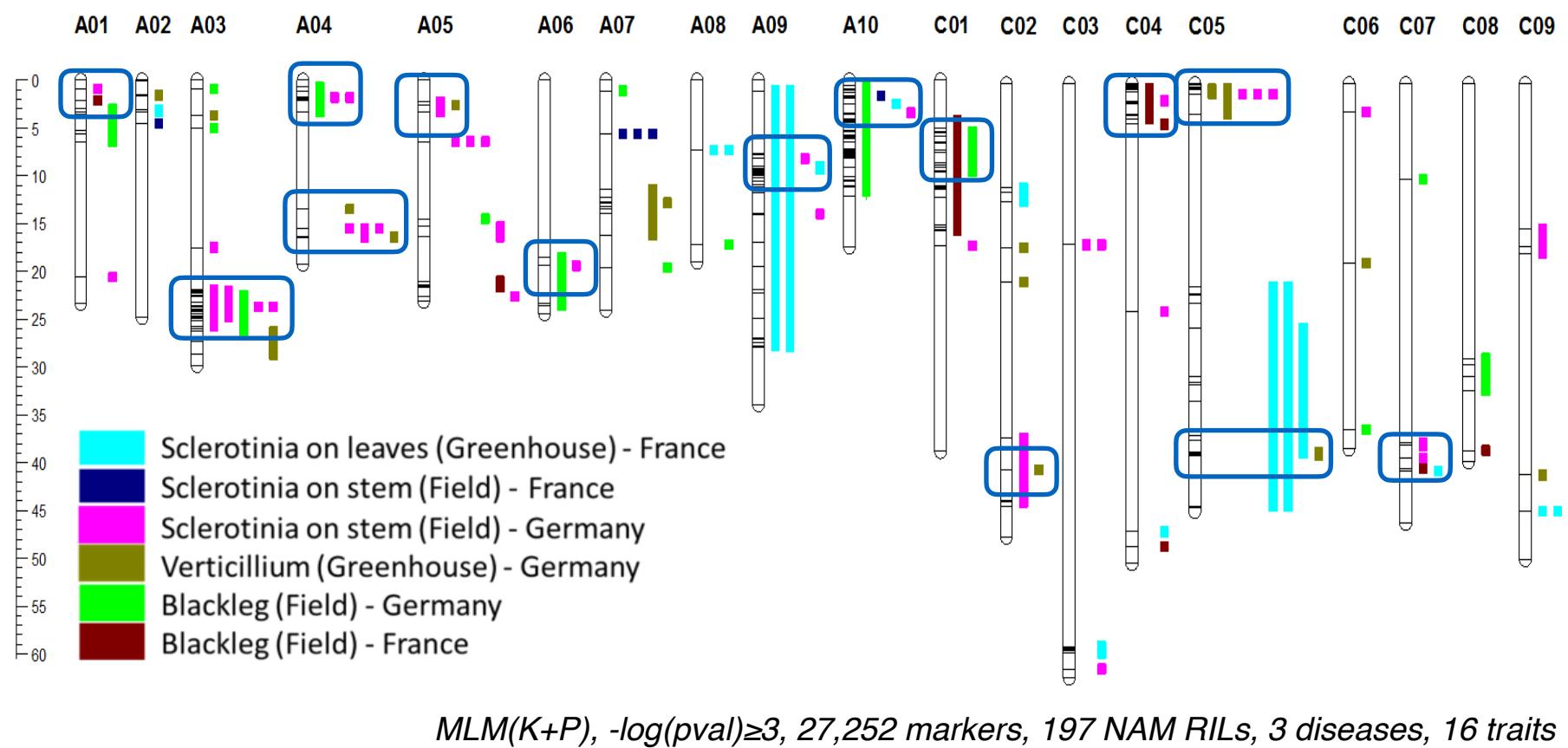
**23,911 SNPs + 3,628 SNAPs**



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

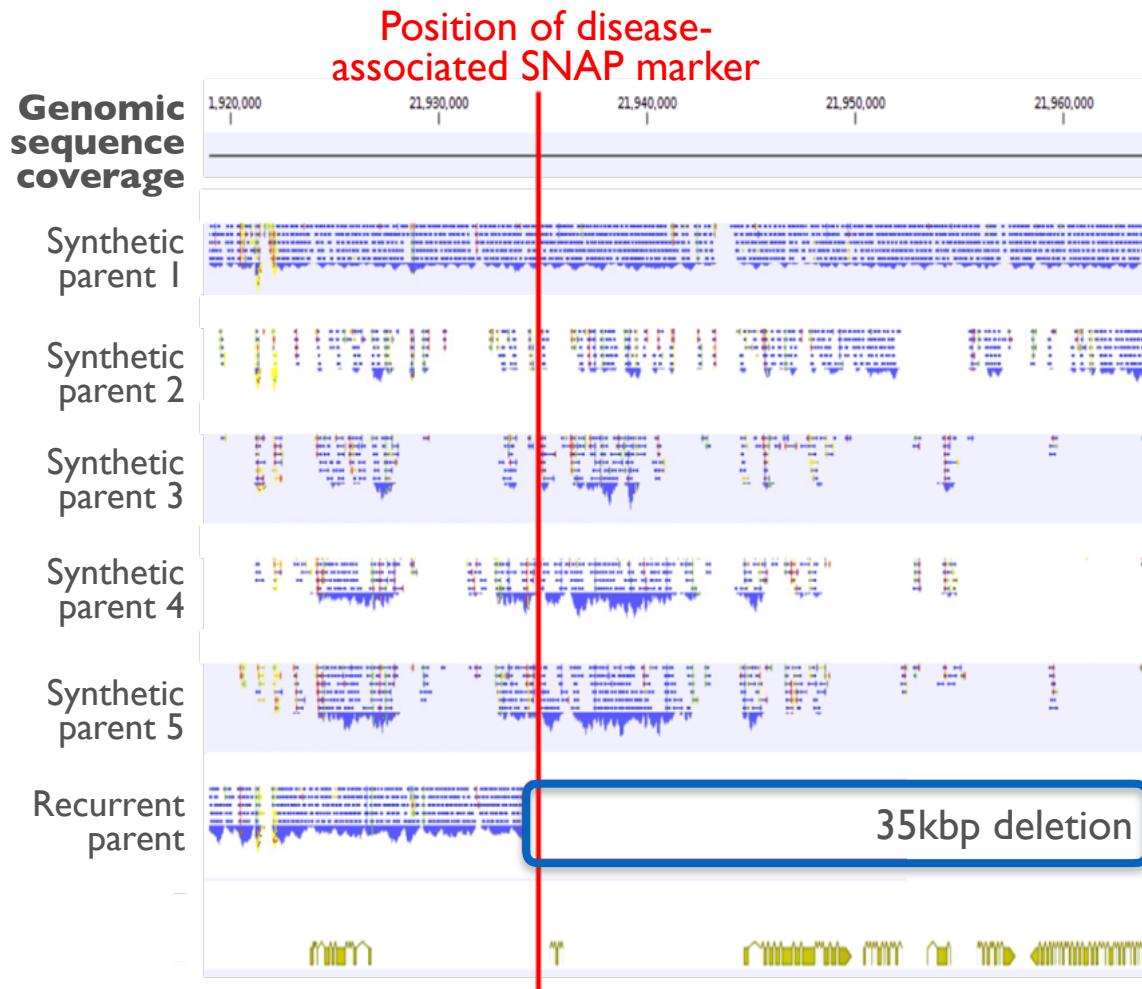
# SNAP data reveals many “invisible” resistance QTL

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

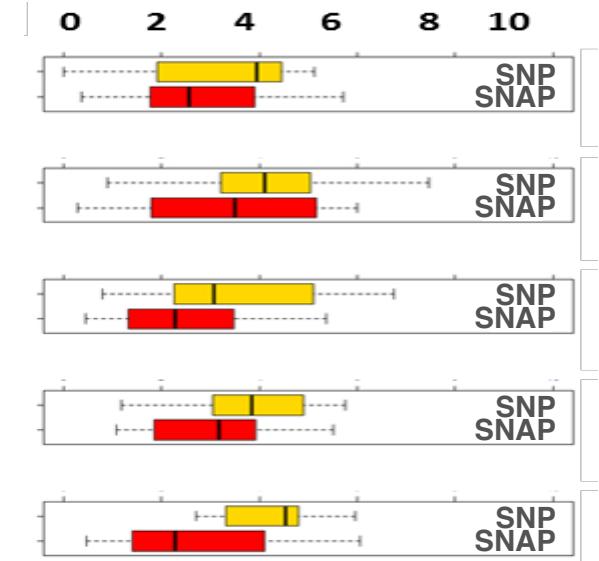


- Numerous overlapping loci influencing quantitative resistance to one or more diseases associate with SNAP markers

# Sequence data associates gene deletions to SNAP-QTL



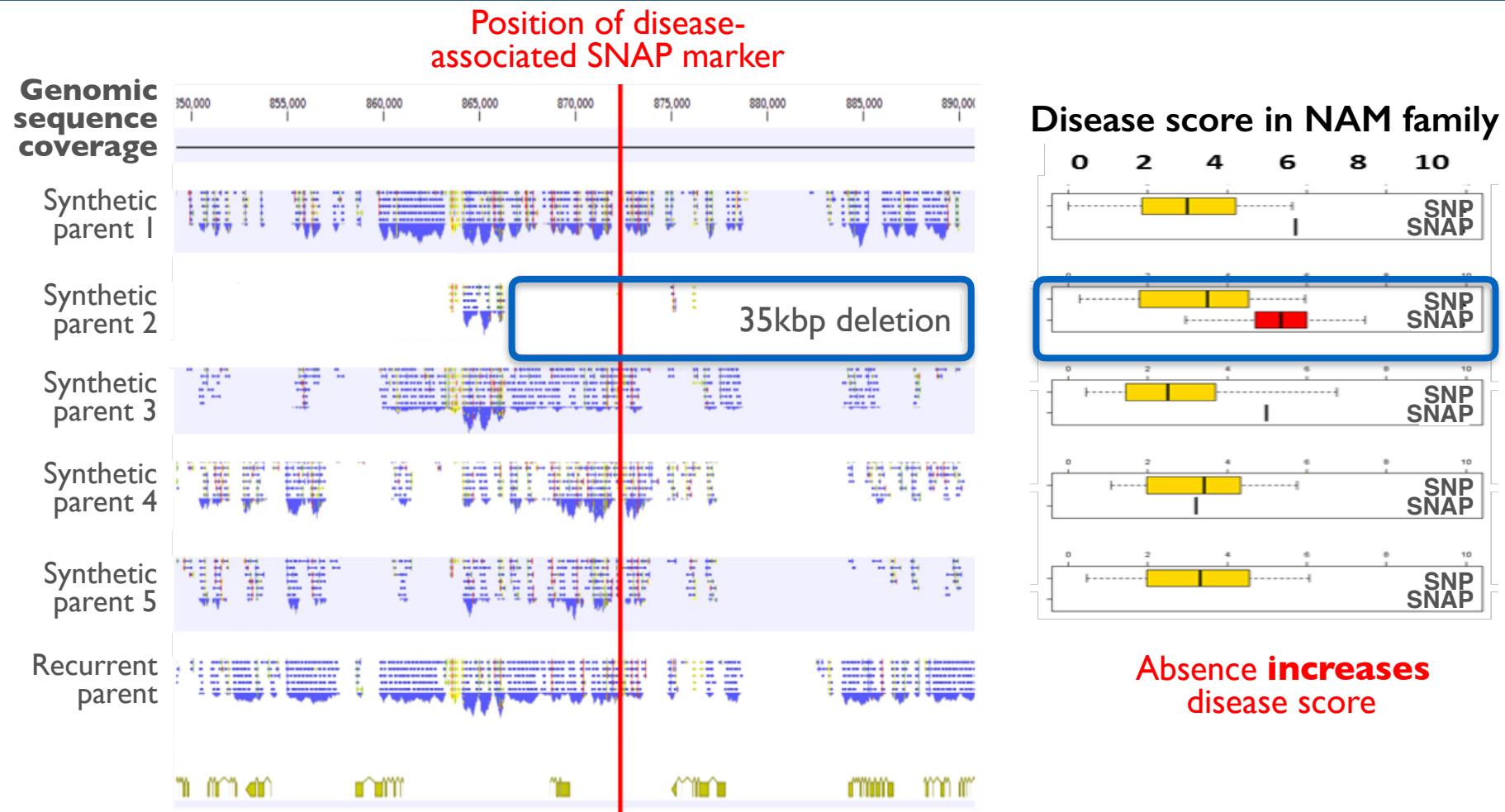
Disease score in NAM family



Absence **decreases** disease score

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

# Sequence data associates gene deletions to SNAP-QTL

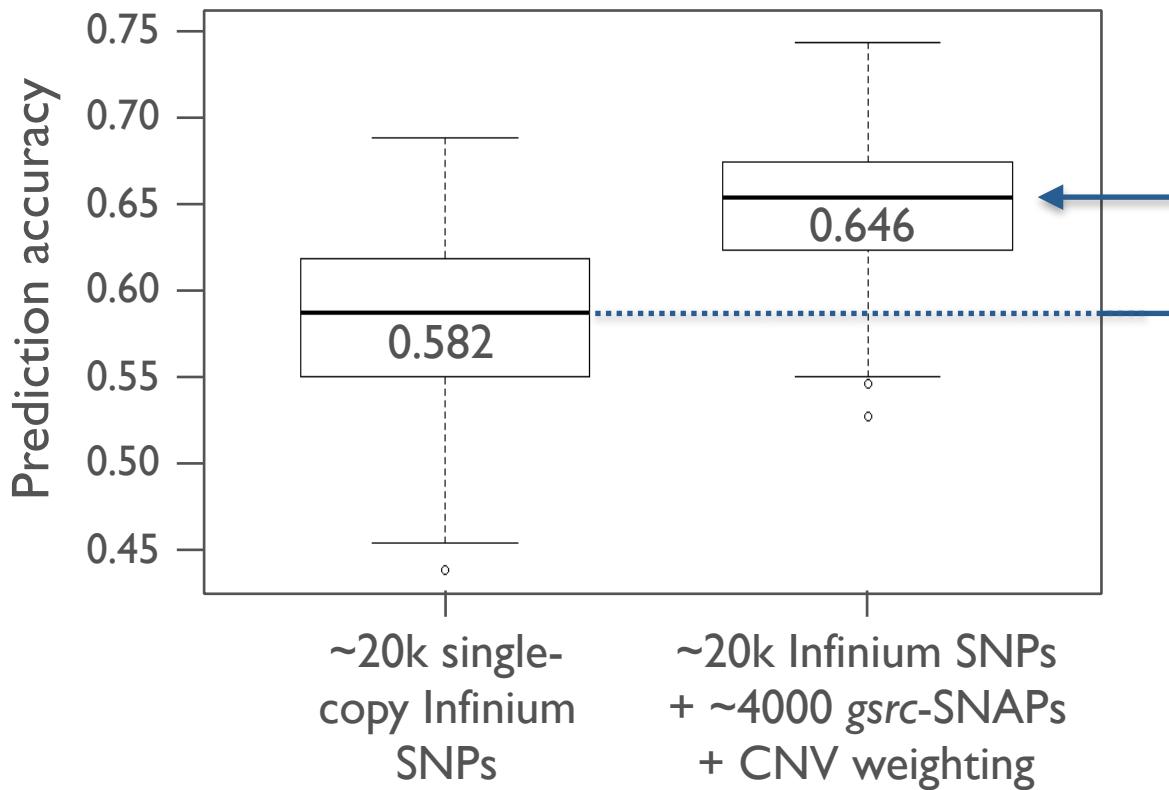


- Example: Sclerotinia stem necrosis QTL, susceptibility associated with a segmental deletion spanning four genes in the Darmor-bzh reference

# Potential to improve genomic prediction accuracy

## Example: Genomic prediction of hybrid yield performance

- rrBLUP for hybrid yield performance in 450 WOSR test hybrids
- Training population n=300, yield data from 12 environments
- Prediction using *in silico* F1 genotype data



6.4% jump in prediction accuracy for hybrid yield after adding SV calls to the prediction model (!)

# Take-home messages

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

- “Accelerated post-polyplodisation evolution” as a key to polyploid success
- Contributed to adaptive variation, distinguishes major gene pools
- Plays an important role in quantitative trait variation, particularly for stress adaptation, and is therefore highly relevant for breeding

## “Missing data” can have hidden value!

- The more we look, the more evidence we find for the impact of structural genome variation on quantitative traits
- Even if the individual effects of single loci are small, using them in genome-wide predictions can potentially improve genetic gain in a selection programme

# Department of Plant Breeding, JLU Giessen

