Exploiting omics of *B. napus* diverse accessions to address trait genetics and polyploid genome evolution toward breeding method innovation

Shengyi Liu

Email: liusy@oilcrops.cn, Lab website: https://www.ocri-genomics.net

Genomics and Disease Resistance Program of OCRI-CAAS
Rapeseed R&D Centre of China Agriculture Research System
Wuhan 430062, website: www.oilcrops.cn

1. Background

B. napus **is a polyploid crop whose genome contains several whole genome doubling/tripling events (WGD) which are highly relevant to breeding selection.**

WGD occurrence is frequent and polyploids is popular. The reasons are considered as that

- **Polyploidy has advantages of vigor, biomass, adaptation and many others (Masterson,1994; Comai, 2005; Dubkovsky and Dvorak, 2007; Jackson and Chen, 2010, and many others)**

- **Polyploid events enabled species radiation/burst which consequently have had a profound impact on the earth ecological system (Soltis and Soltis, 2003; Ainouche and Jenczewski, 2010; Van de peer 2011)**

On breeding technologies

- Line/pedigree/systemic breeding
- hybridization & hybrid breeding
- backcross breeding
- mutation breeding
- Interspecific hybridization breeding
- ploidy breeding
- cell engineering
- gene engineering
- marker-aided breeding

Since re-discovery of Mendelian inheritance law and establishment of pure line theory set the foundation of breeding 100 years ago, great progress has been made in breeding technologies.

Generally, the core: hybridization & selection.

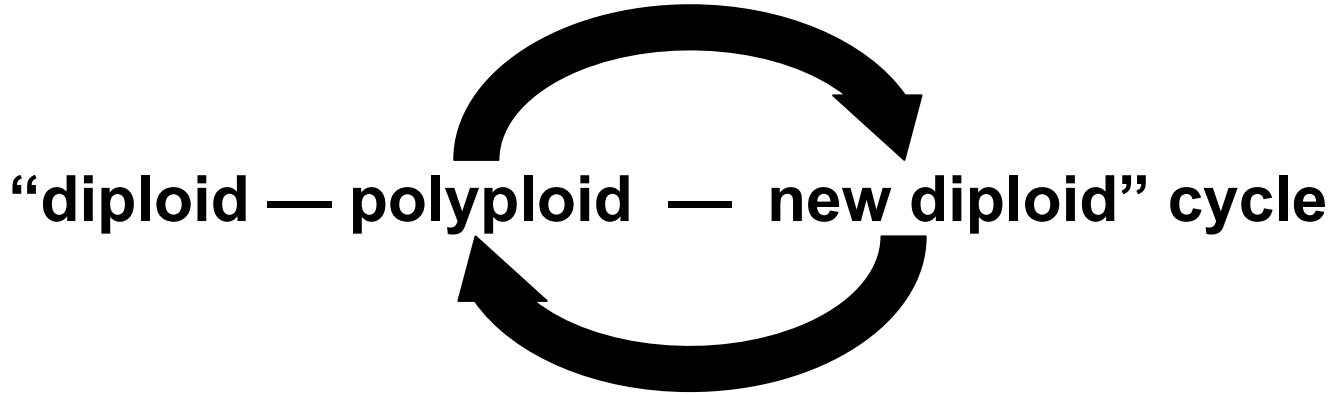
however, on the issues
1) what are elite parents and how to choose;
2) how to select elite individuals

for the 100 years, no substantial progress

Genome selection (GS) based on genotype linking to phenotype is a promise,
However, we just know the tip of an iceberg about loci/genes underlying quantitative traits

On breeding-related genome vaiantion and evolution:

- what force driving the cycle occurrence is unclear



- It has been long considered that polyploid crop genetic diversity is narrow and a challenge for breeding.

  True or not, or how to overcome this limitation?

The objectives of our genomics studies are to

- **establish omics platform toward development of breeding technologies**
- **address polyploid evolution with an aim to help development of breeding technologies**

2. Developement of populations and omics data

Materials we are using for our studies

- natural populations: three populations comprising 2500 core diverse germplasm accessions, all genomes re-sequenced, some with 60 coverage, and phenotyped in multiple environments/years for more than 20 traits.

- 5 RIL and DH recombinant populations, genotyped by 60K SNP array, F2 populations…...

- EMS mutation population, introgression lines.....

Multiple omics data generated

Ref full length transcripts + transcriptomic data of SAM, pods, roots and leaves of 320 accessions

Genome sequences of 3000 diverse accessions, 10+ *de novo* assemblies

Phenomic data of >20 traits for 8 years

Metabolomic data of 320 accessions

Pan-genome constructed from 2800 accessions

Highly dense recombination map

Database: https://BnaOmics.ocri-genomics.net , searchable & analyable

Insight into

1. genome variation and association with traits;
2. polyploid genome evolution

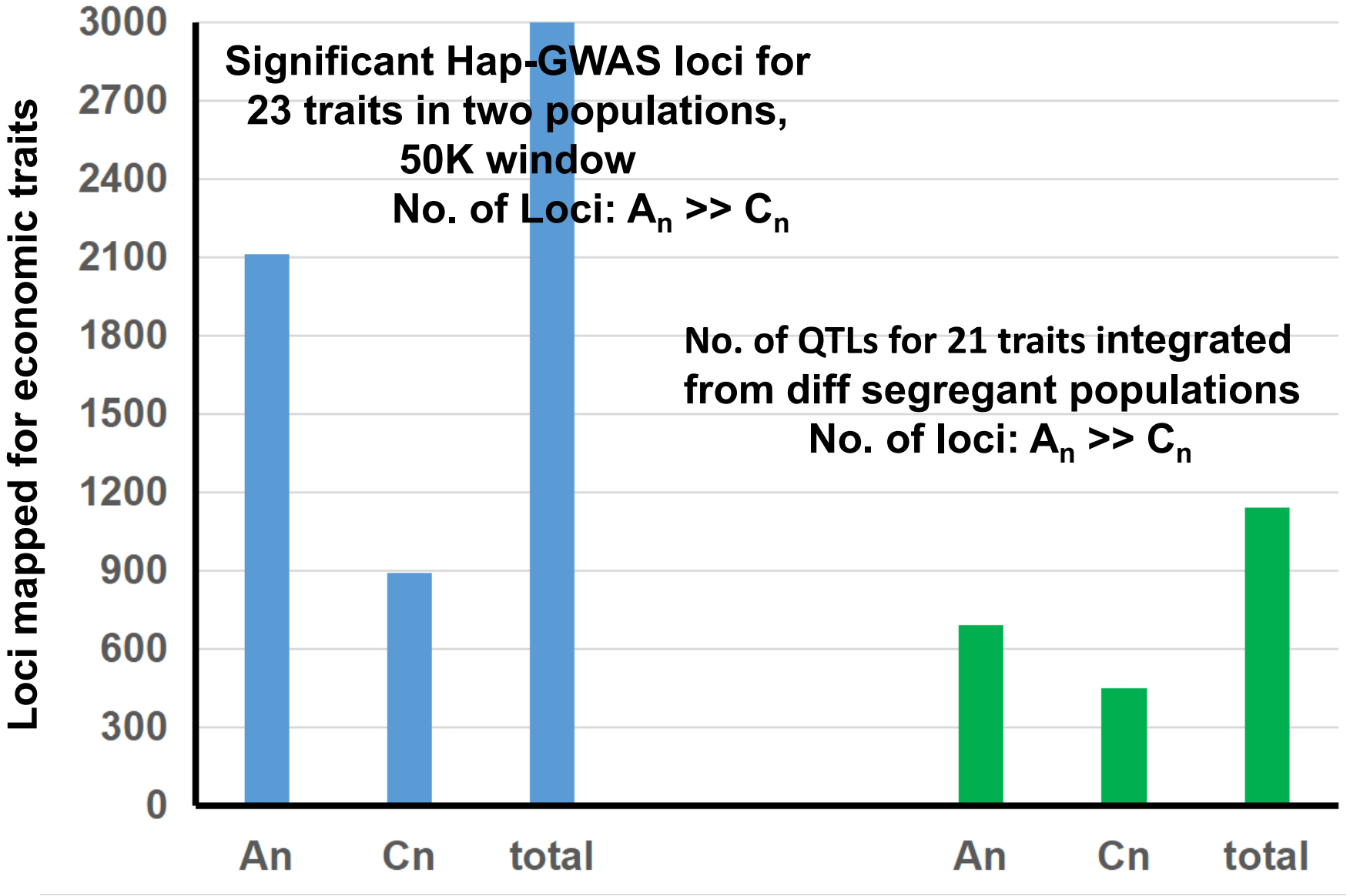
Breeding technol

Breeding by genome design (BGD):

1. choice of parents for hybridization;
2. selection after backcross and hybridization;
3. creation of new elite germplasm

3. Trait genetic dissection

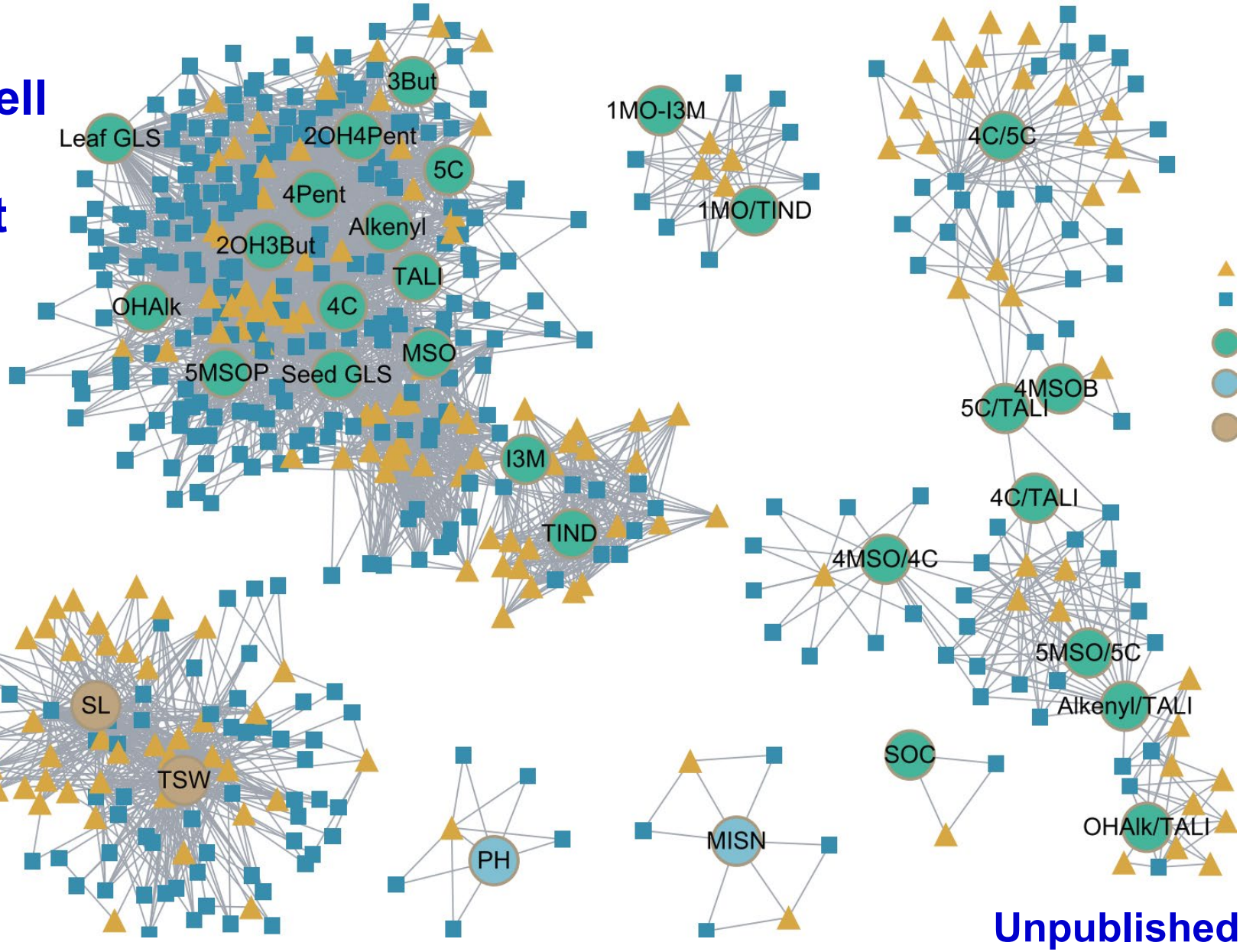
At the DNA level: **3000 trait loci mapped from GWAS**
1100 loci integrated from segregant populations



At the RNA level

Through joint analyses of GWAS — eQTL — TWAS — GWAS-eQTL colocalization

We identified a thousand of genes controlling traits as well as certain regulatory mechanisms undeelying trait variation

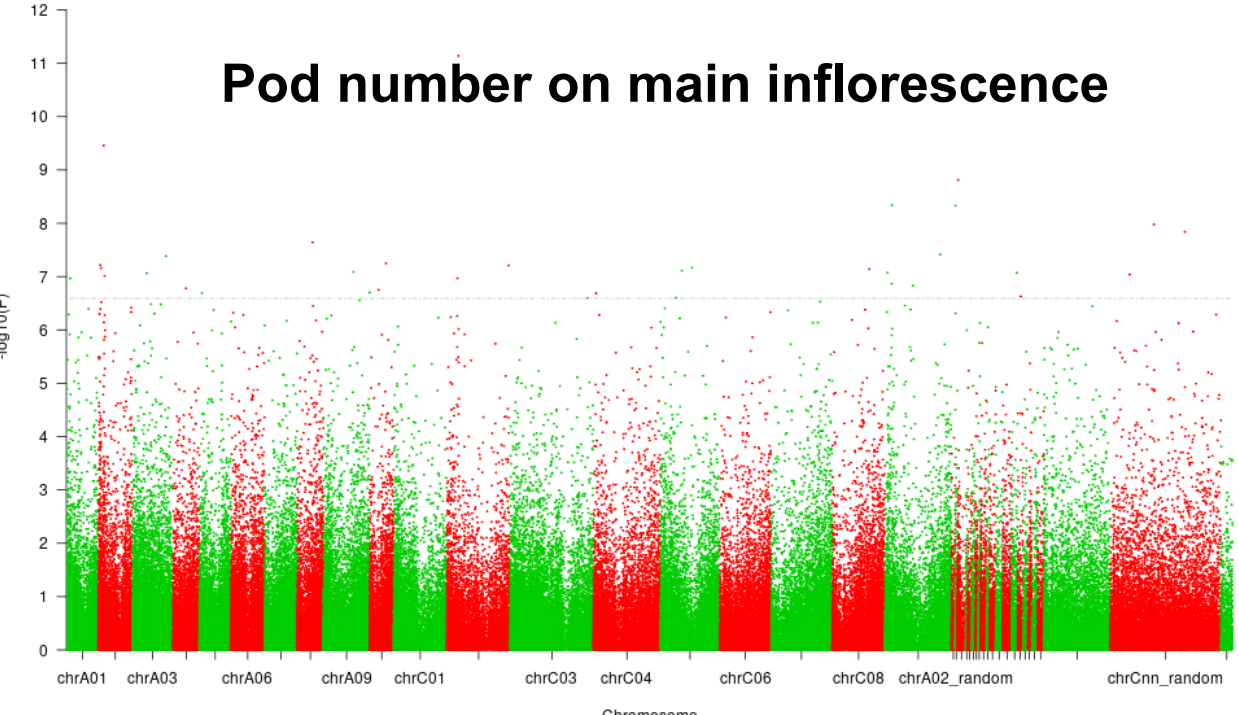


Unpublished

At the RNA level

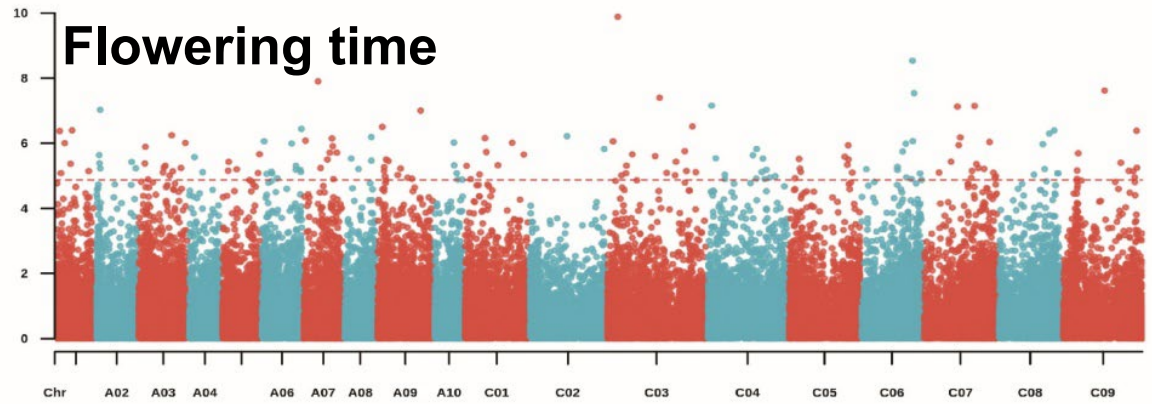
Association analyses of Alternative splicing variants identified hundreds of transcript variants of genes controlling traits as well as certain regulatory mechanisms of trait variation

Pod number on main inflorescence

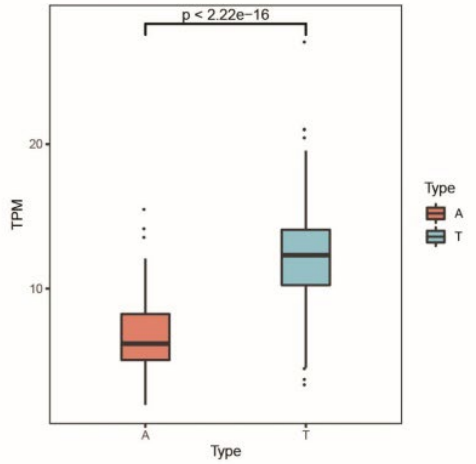


A

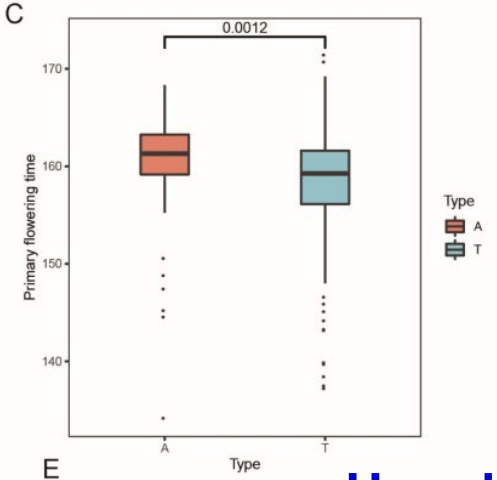
Flowering time



B



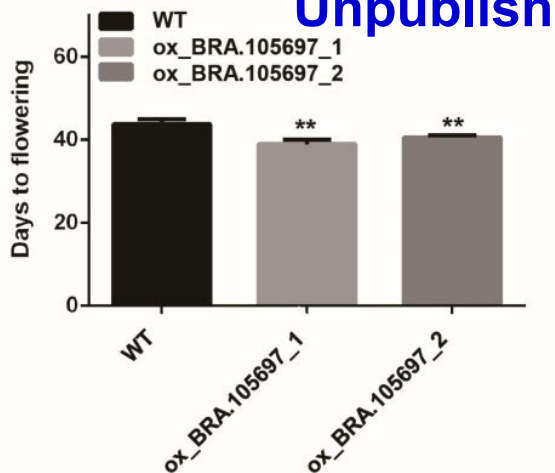
C



D

E

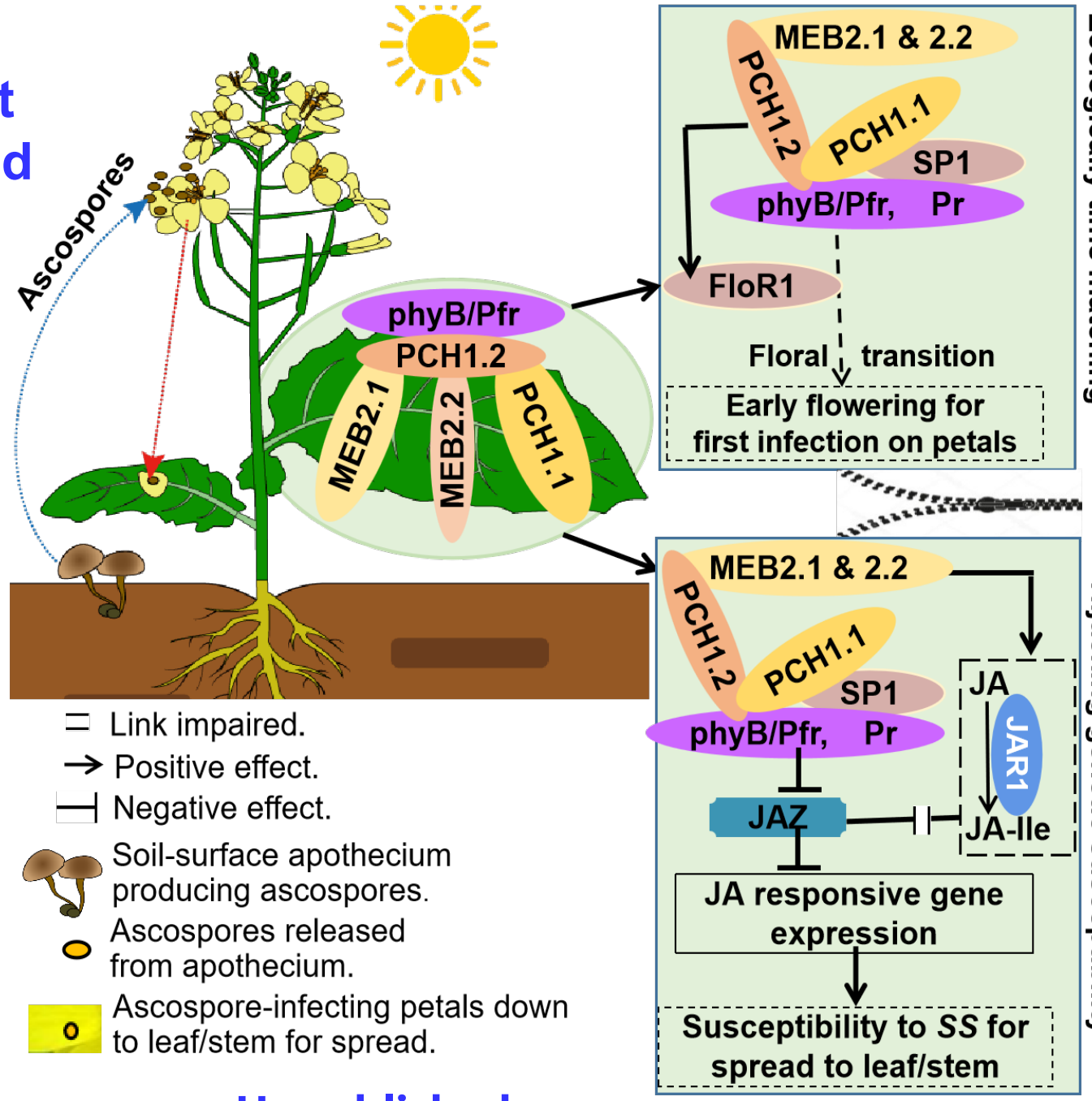
Unpublished



A series of genetic and biochemical experiments revealed a network that is wired by functionally differentiated alternative splice transcripts

A proposed model explaining two coevolution layers corresponding to two key steps of *Sclerotinia* disease cycle by promoting flowering and increasing disease susceptibility:

- **pleiotropic gene networking promotes flowering to enable time match of flowering to ascospore release for infection of petals as nutrient media**
- **the networking through phyB(Pfr/Pr), possibly through suppressing JAZ/MYC, increases susceptibility of leaves and stems**



Unpublished

4. Polyploid genome evolution

Arabidopsis—B. rapa/B. oleracea—B. napus at a single genome level

Layer	Asymmetrical evolution
1	Asymmetrical TE amplification
2	Asymmetrical Tandem duplication amplification
3	Asymmetrical divergence of duplicates' DNA sequences

From the comparative genomic analyses, we got known what are genome characteristics between *B. napus* and other species and between A and C of *B. napus*

Gill et al. 2021. Critical Reviews in Plant Sciences, 40: 157-189
Liu et al. 2021. The Plant Journal, 2020, doi: 10.1111/tpj.15037
Chen et al. 2021. Plant Biotechnology Journal 19: 615–630
Yao et al. 2020. The Plant Journal: 103:843-857
Liu et al. 2020. The Plant Journal, doi: 10.1111/tpj.15037
Guo et al. 2017. The Plant Journal, 91:34-44
Liu et al. 2014. Nature Communications 5:3930
Chalhoub et al 2014. Science 345: 950-953
Tong et al. 2013. BMC Genomics, 14:689
Zhao et al. 2013. The Plant Journal 76: 211–222

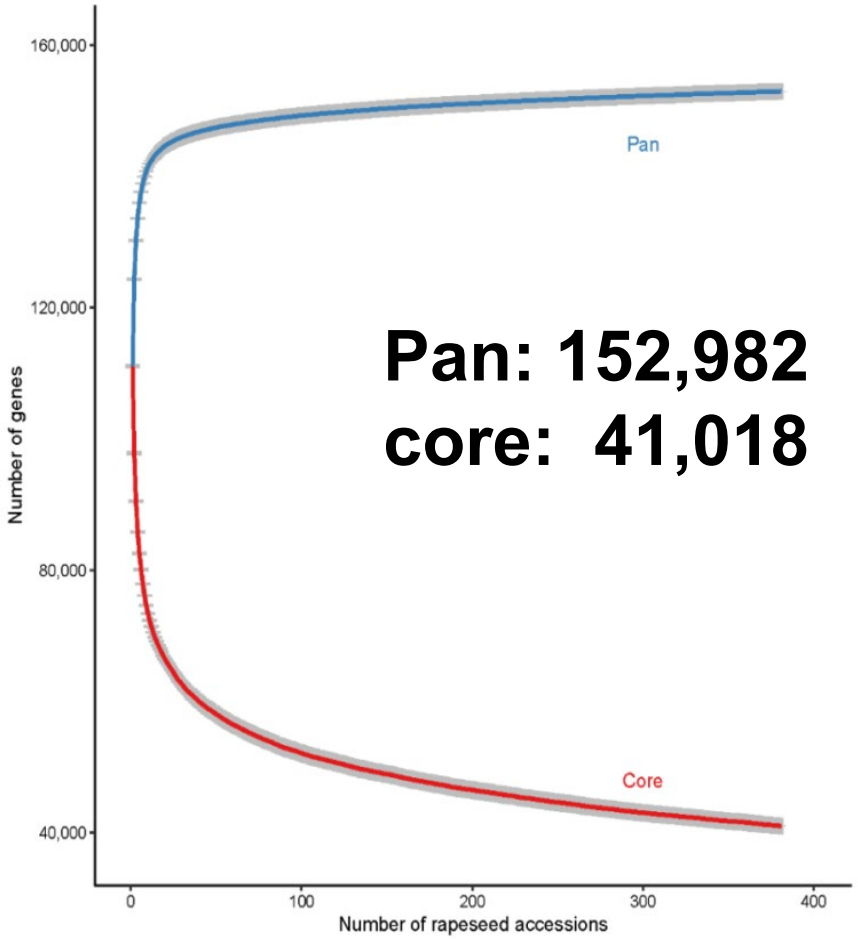
Inter-specific
Intra-specific subgenomes

The follow-up is population genomic analyses in *B. napus*

Layer	Asymmetrical evolution
4	Asymmetry in structural variations: gene loss/deletion, PAV, CNV, HE
5	Asymmetrical divergence of duplicates' expression and alternative splicing variants
6	Asymmetry in epigenetics –methylation, ncRNA
7	Recombination dominance of A over C
8	Asymmetry in QTLs/loci controlling traits

I just picked up some results related to breeding.

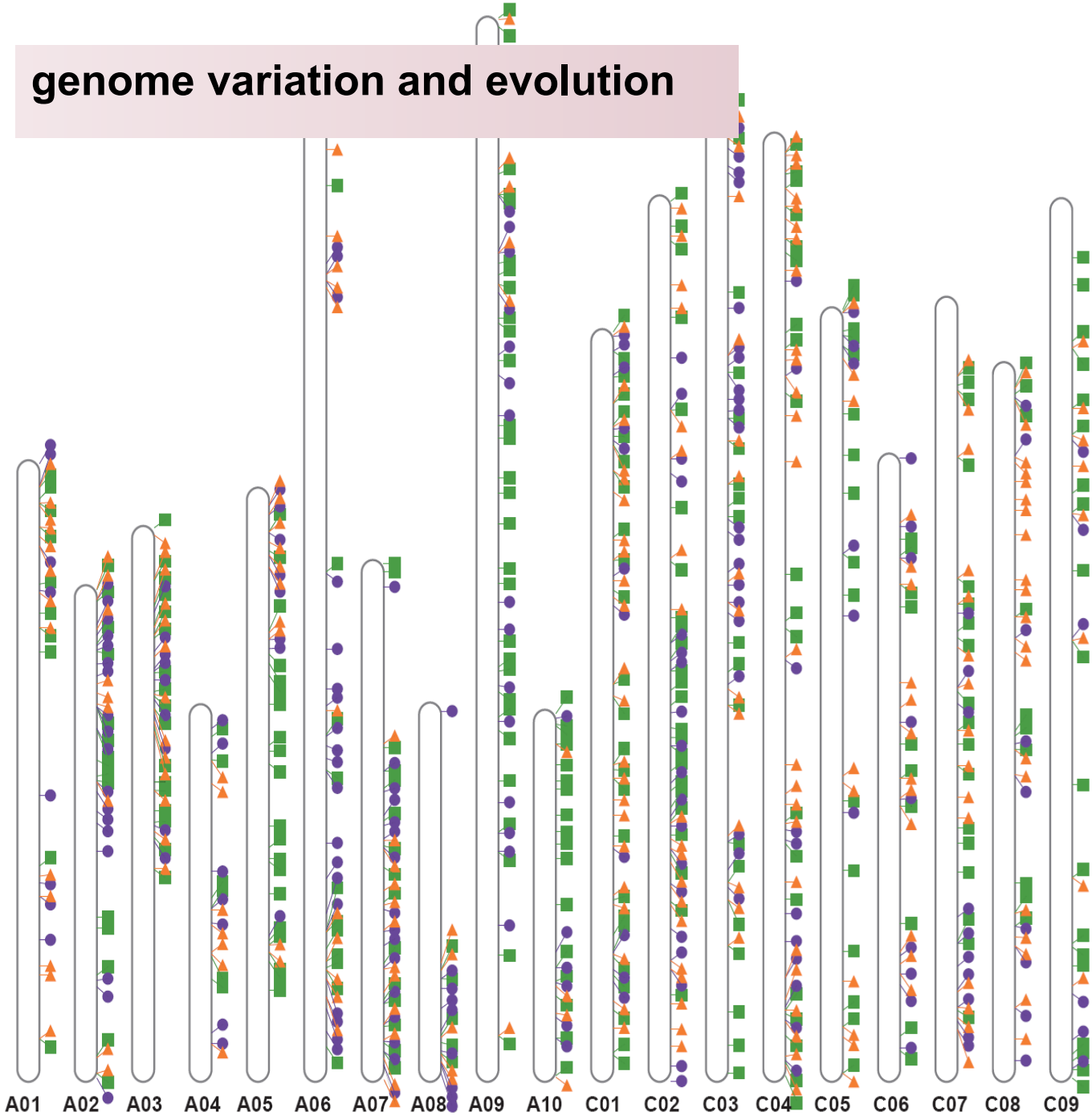
Core genes are just 1/3 of total genes, indicating great potential for breeding, But how to select? By BGD?



Pan: 152,982
core: 41,018

Unpublished

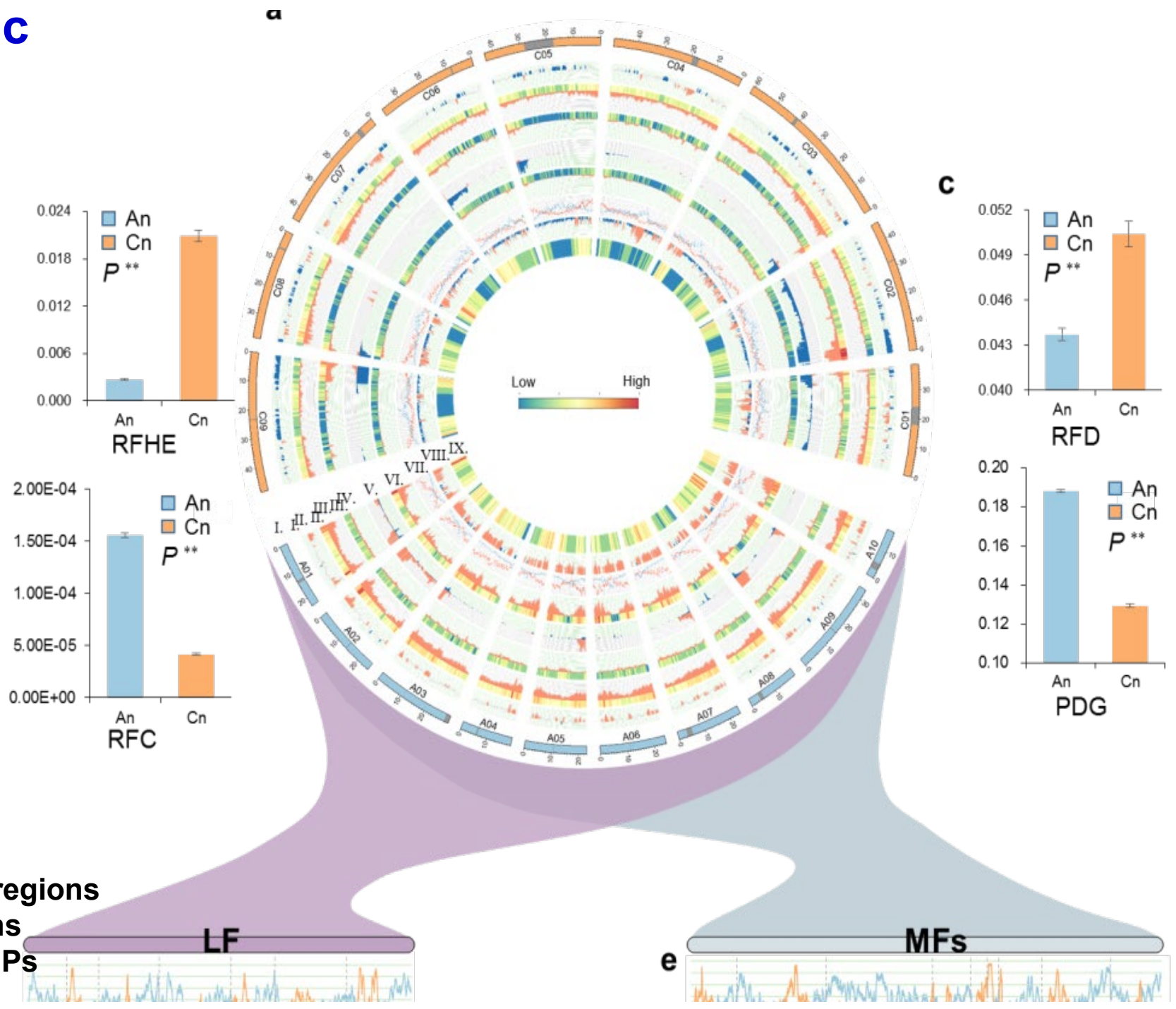
genome variation and evolution



Characteristics of genomic variation in SNP, SV (HE, PAV...), gene conversion, Rec rate

Similar trends

- **between A_n and C_n**
- **between LF and MFs**
- **between α1 and α2**

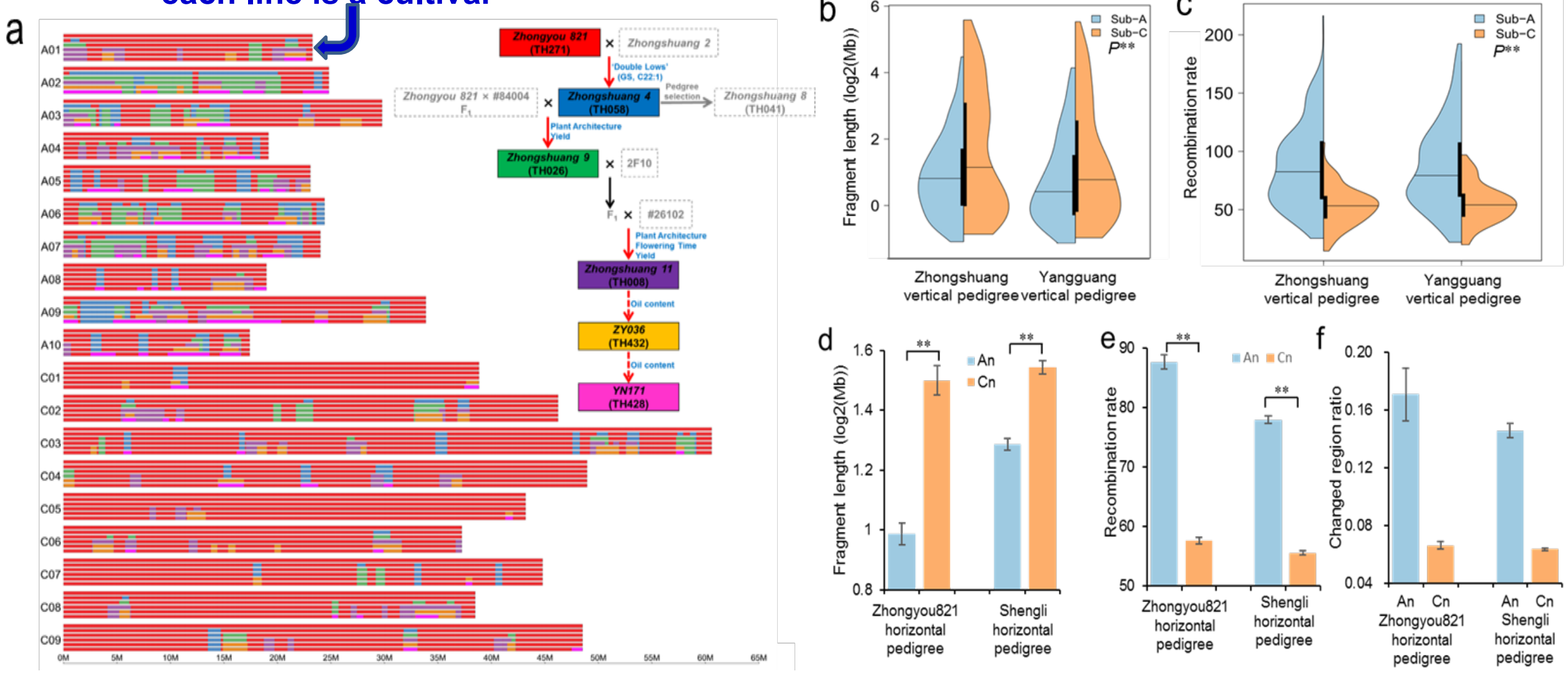


RFHE: relative frequency of HE replaced regions
RFD: relative frequency of deletion regions
RFC: relative frequency of conversion-SNPs

Breeding pedigrees demonstrate genome segment dynamics and conservation/selection providing idea on selection

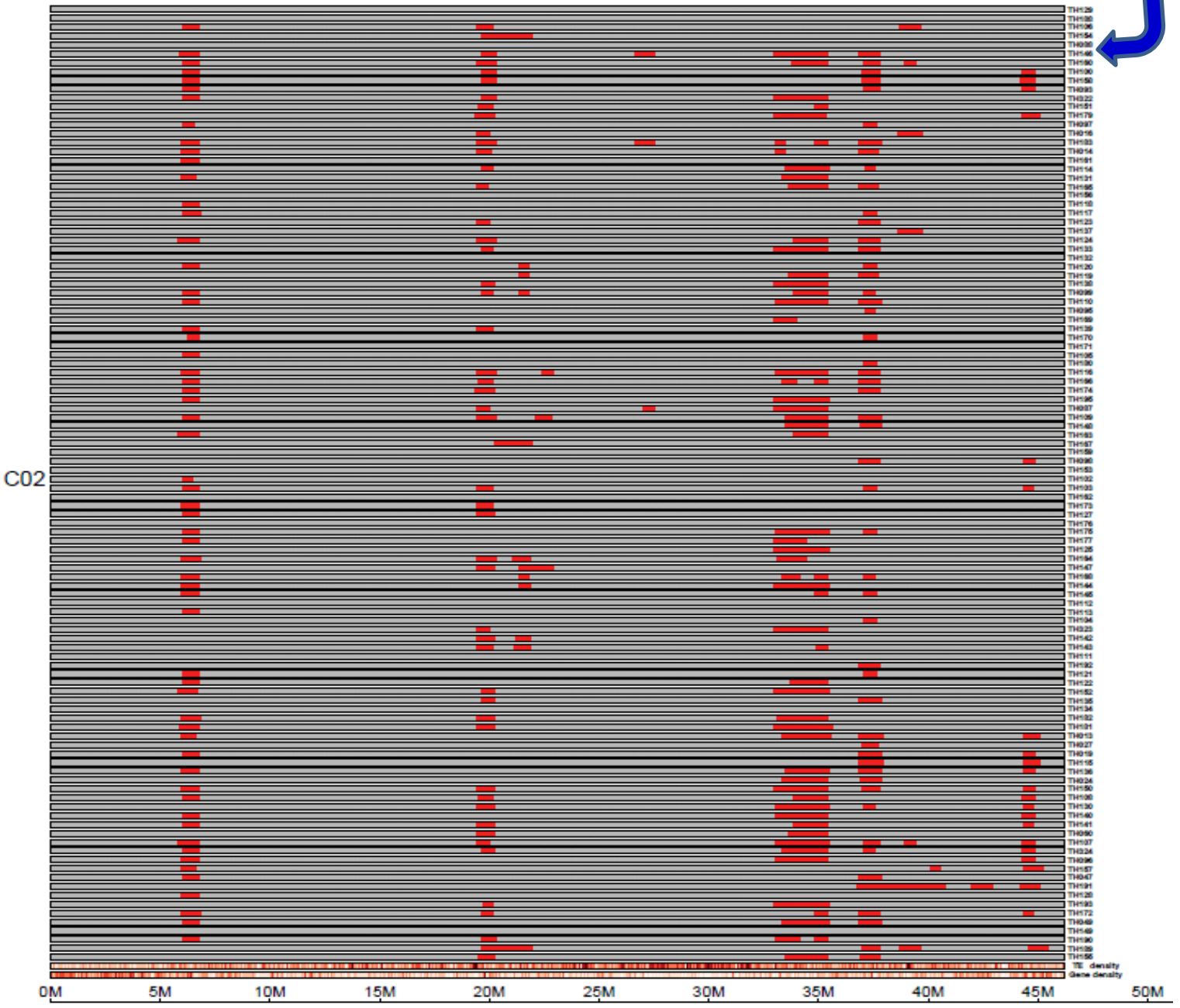
Vertical pedigrees (ABC…)

each line is a cultivar

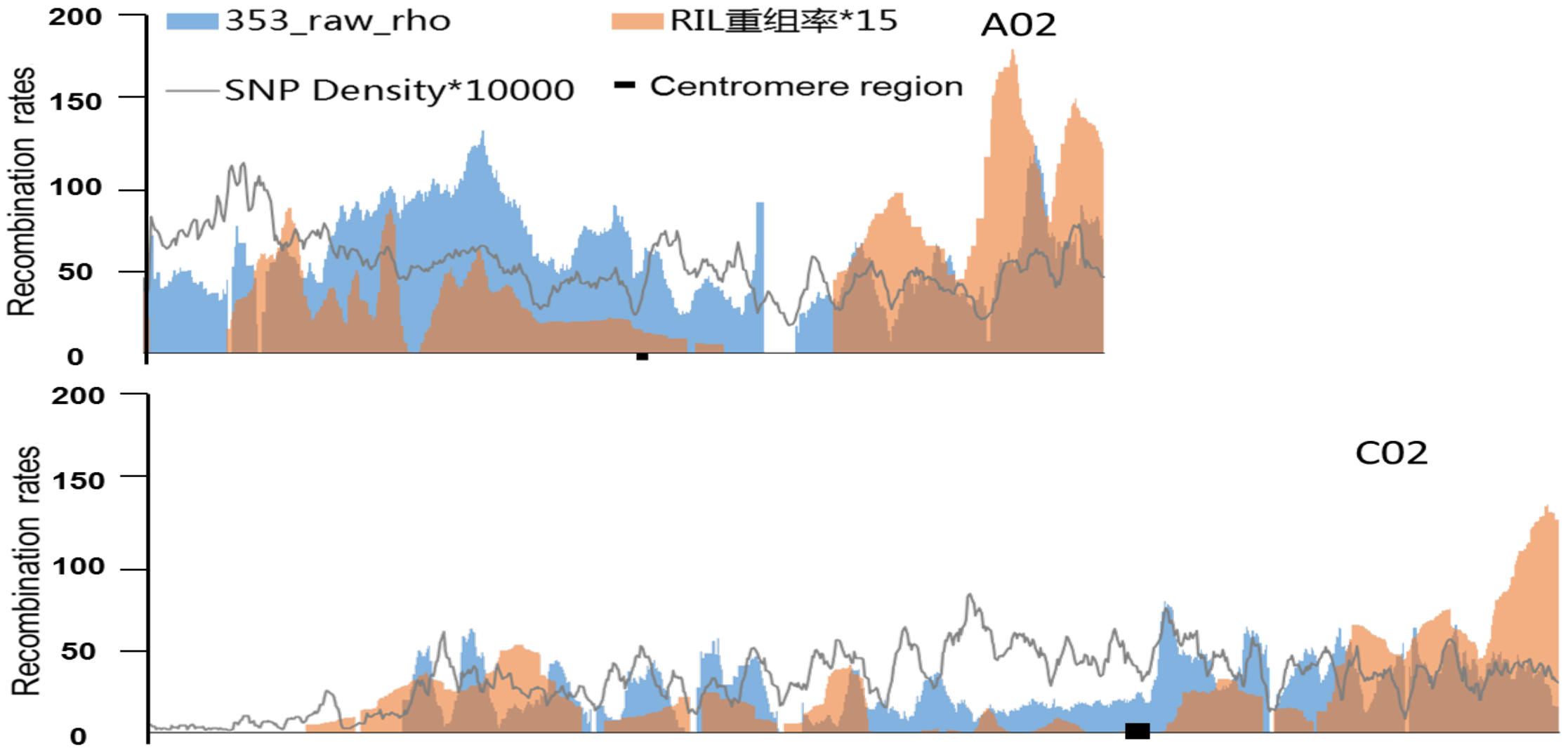


horizontal pedigrees (140 varieties shared one parent)

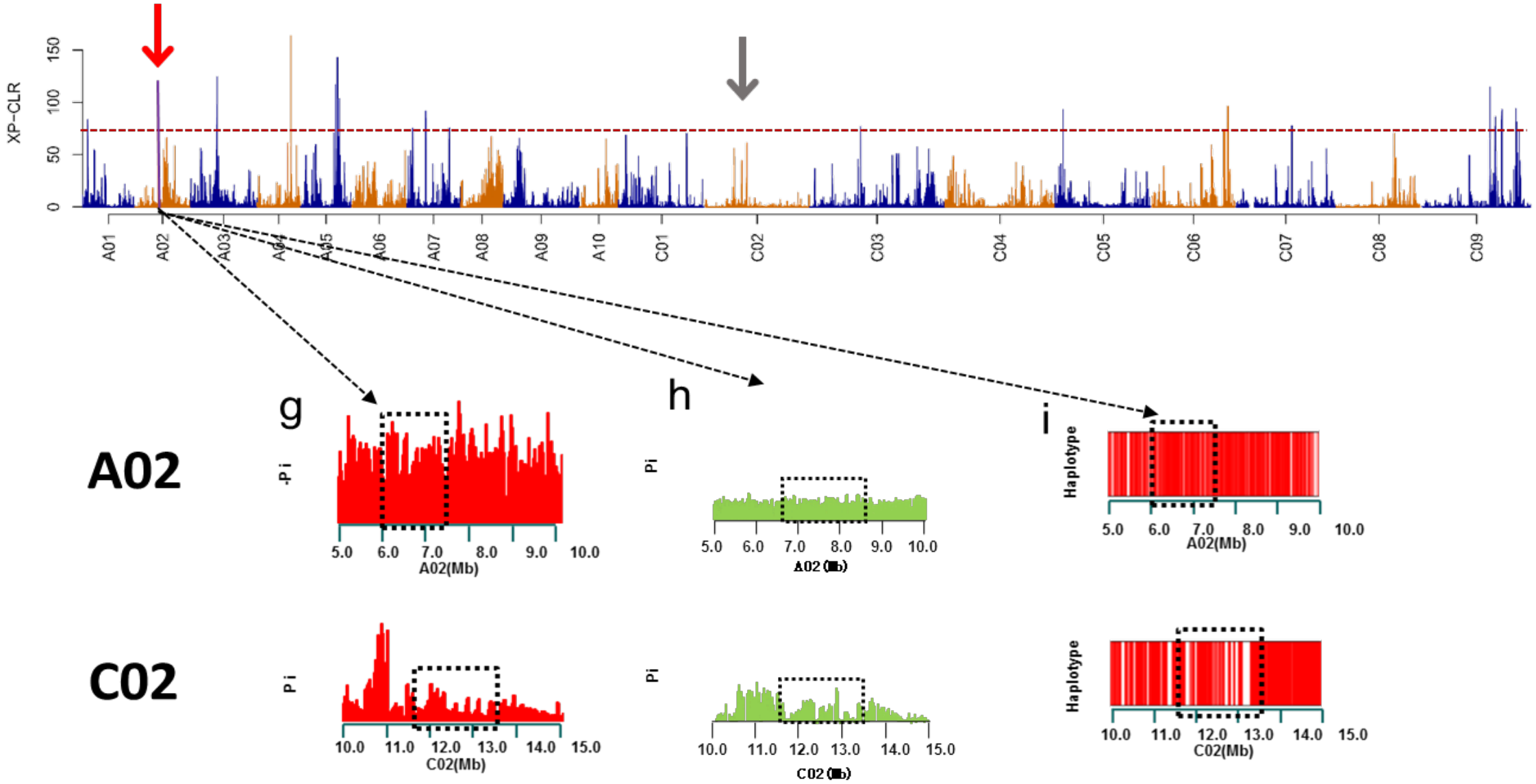
each line is an accession



Population omics data allow us to achieve a highly dense map of recombination rate (Rec rate)



In the sub-genomic syntenic regions A02 vs C02, selective sweep and QTLs just on A02, not C02, for several traits (flowering time, plant height, # branches, # pods and pod density)

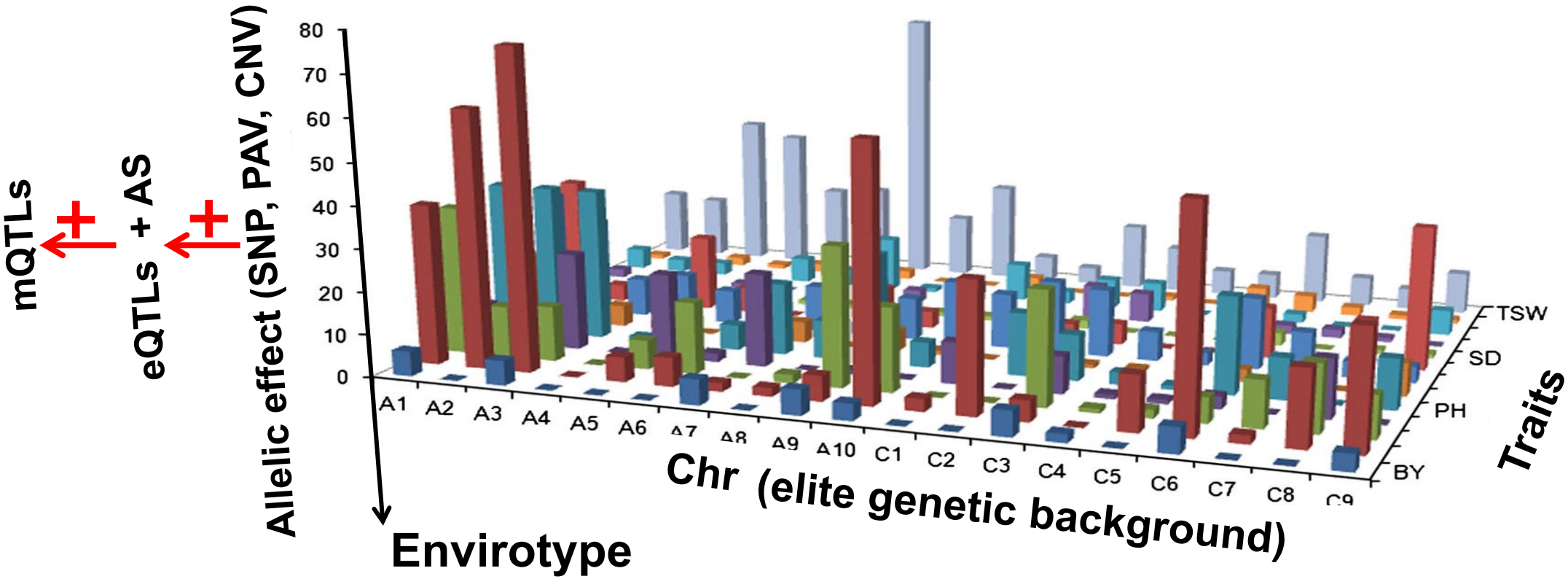


5. perspective on BGD

Breeding by genome design (BGD) (≈ synthetic biology)

For parent choice, germplasm creation, selection of hybridization offspring

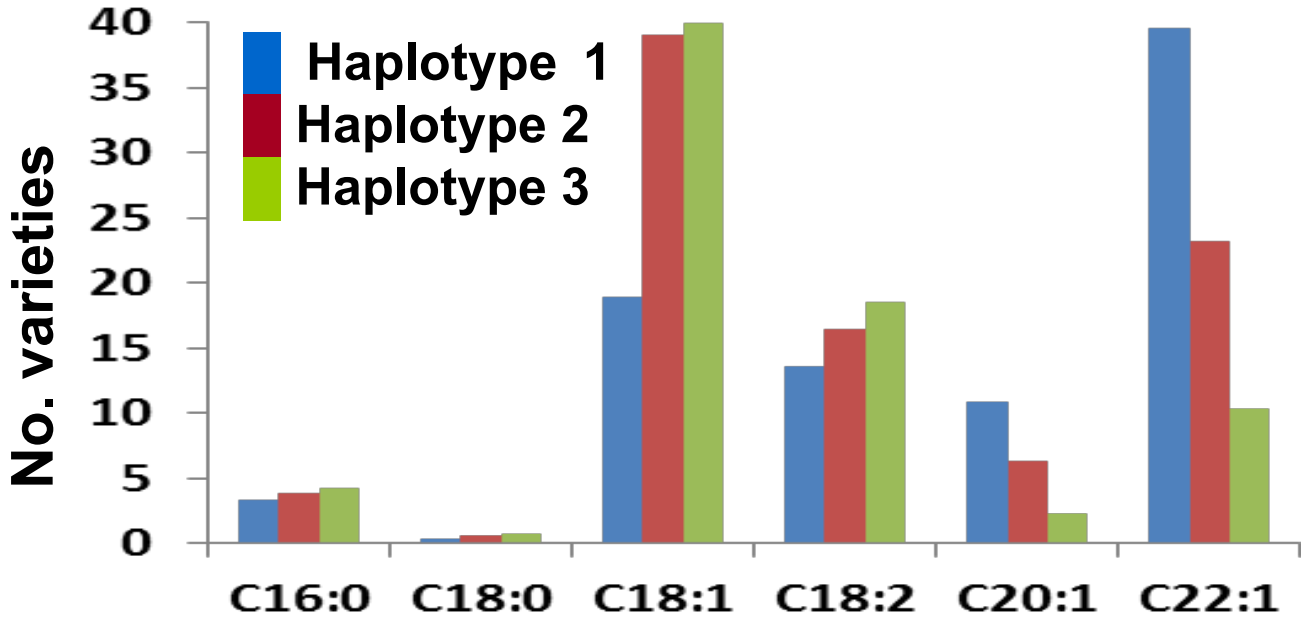
$$\text{predicted elite variety} = a + \sum_{i=1}^{n}(X_i Y_i Z_{1i} Z_{2i} Z_{3i}) + \varepsilon$$



Tartgets to be selected is an elite variety, rather than traits (not MAS)

Within BGD, some pathways can be considered as individual modules/factors of a model

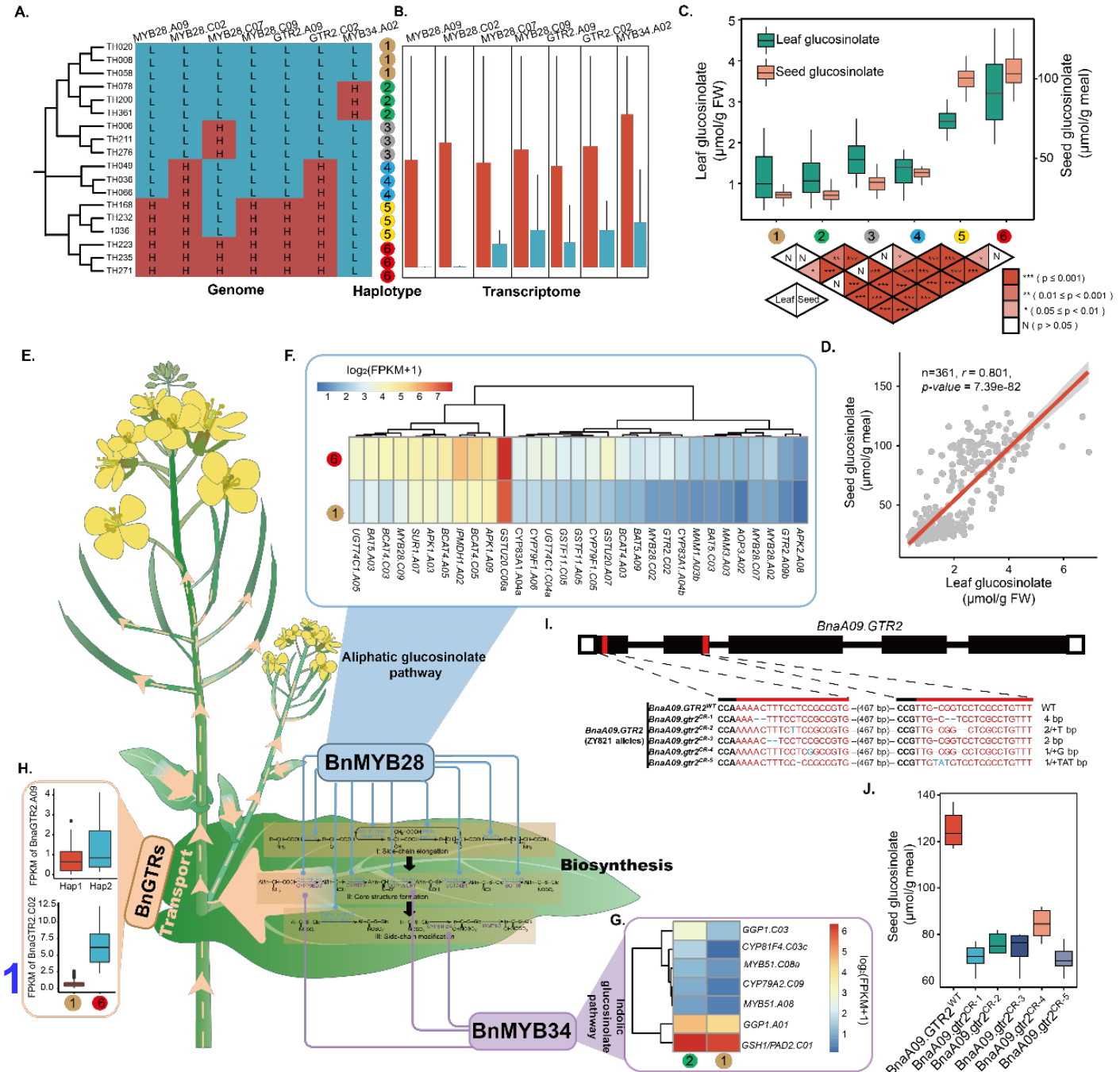
Haplotype 3 is for selection of good fatty acid profile



More complex module of contents of total glucosinolates and components

QTL and GWAS mapping:

- a complete set of haplotypes
- gene editing created high glucosinolates in leaves & low in seeds
- higher resistance to aphid and Ss.



He et al. Plant Physiology, 2022, 188:1848–51 and unpublished

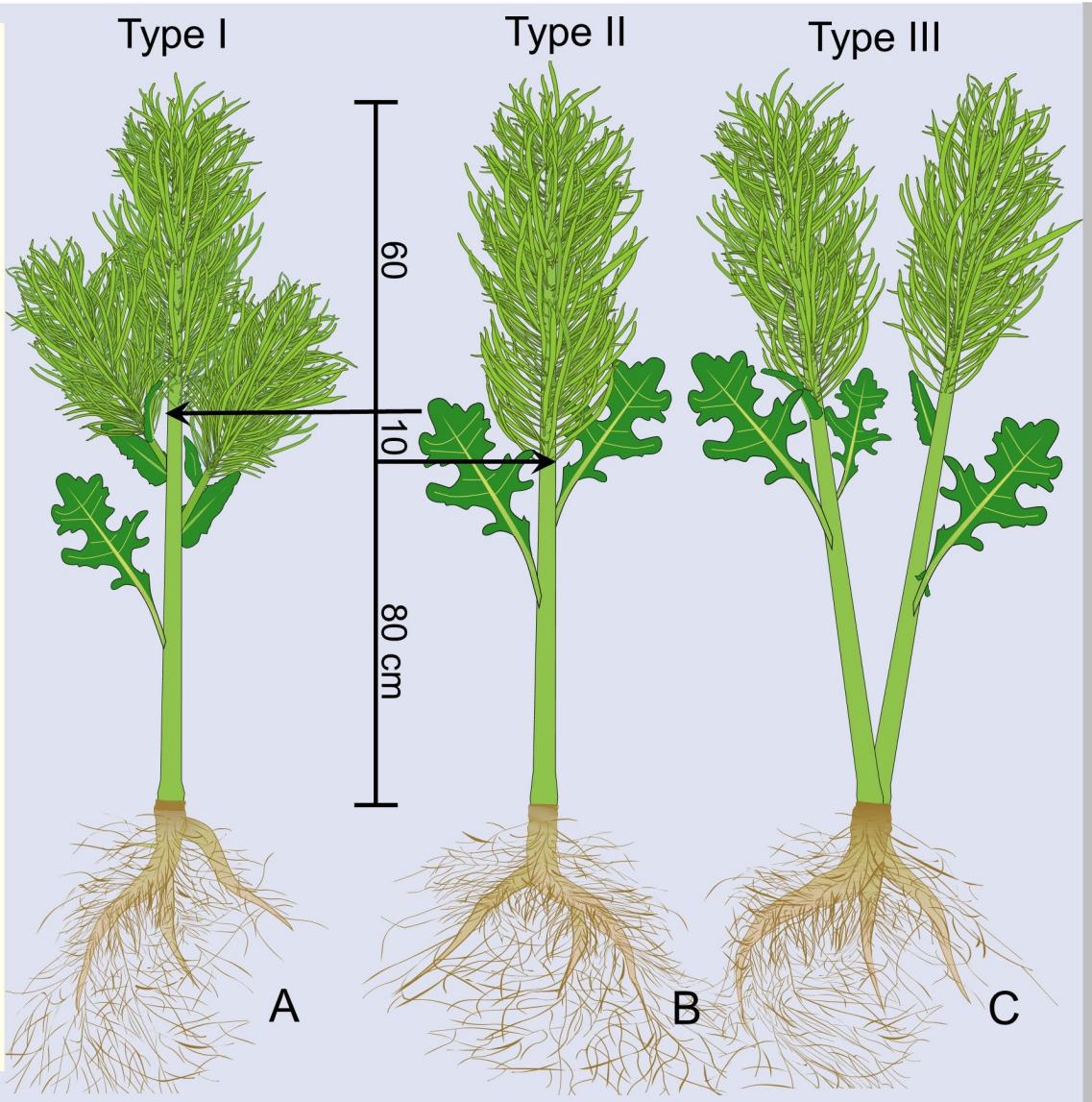
In China, canola production is facing a big challenge because of small farm and low profit. Against it, we proposed a model to double seed yield.

Liu S, Raman H, Xiang Y, Zhao C, Huang J, and Zhang Z et al. 2022. *De novo* design of future rapeseed crops: Challenges and opportunities. The Crop Journal 10: 587–596

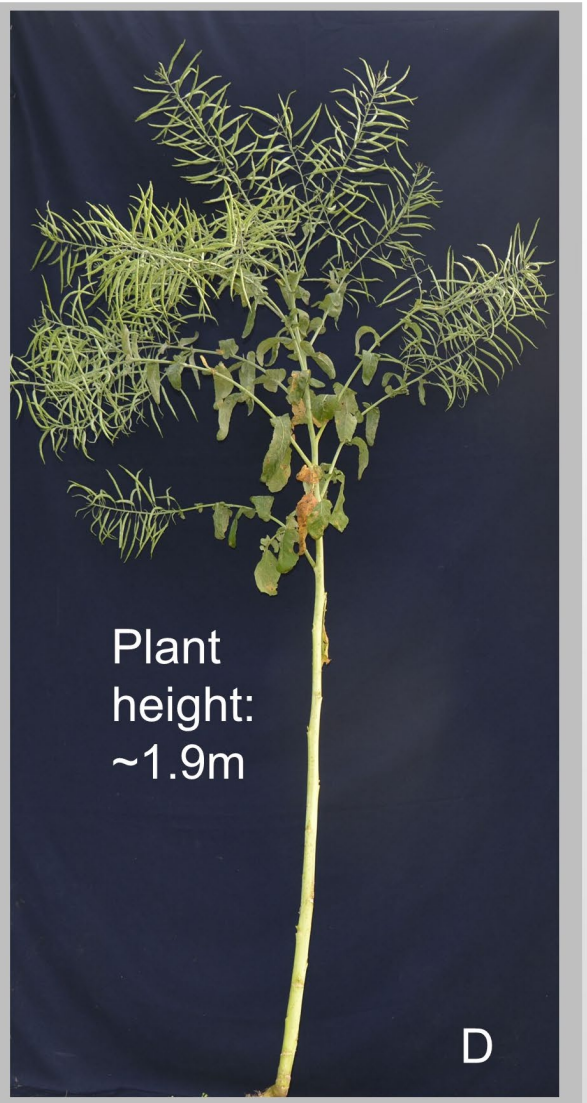
De novo design of OSR through breeding by genome design (BGD)

Indicators

Plant height：~1.5 m
Stem inflorescence:
≥60 cm
Branch angle:≤25°
Leaf angle: ≤30°
Stem: thick & firm
Plant density:
4.5-7.5x10^5 ha^{-1}
Pod density:≥2.3 cm^{-1}
Pod length:≥10 cm
Pod angle: ≤25°
#seed: ≥22 per pod
1000 seeds: ≥4.2 g
Harvest index: ≥0.4
Total siligues:
>1.4x10^8 ha^{-1}
Expected seed yield:
>12 tonnes ha^{-1}



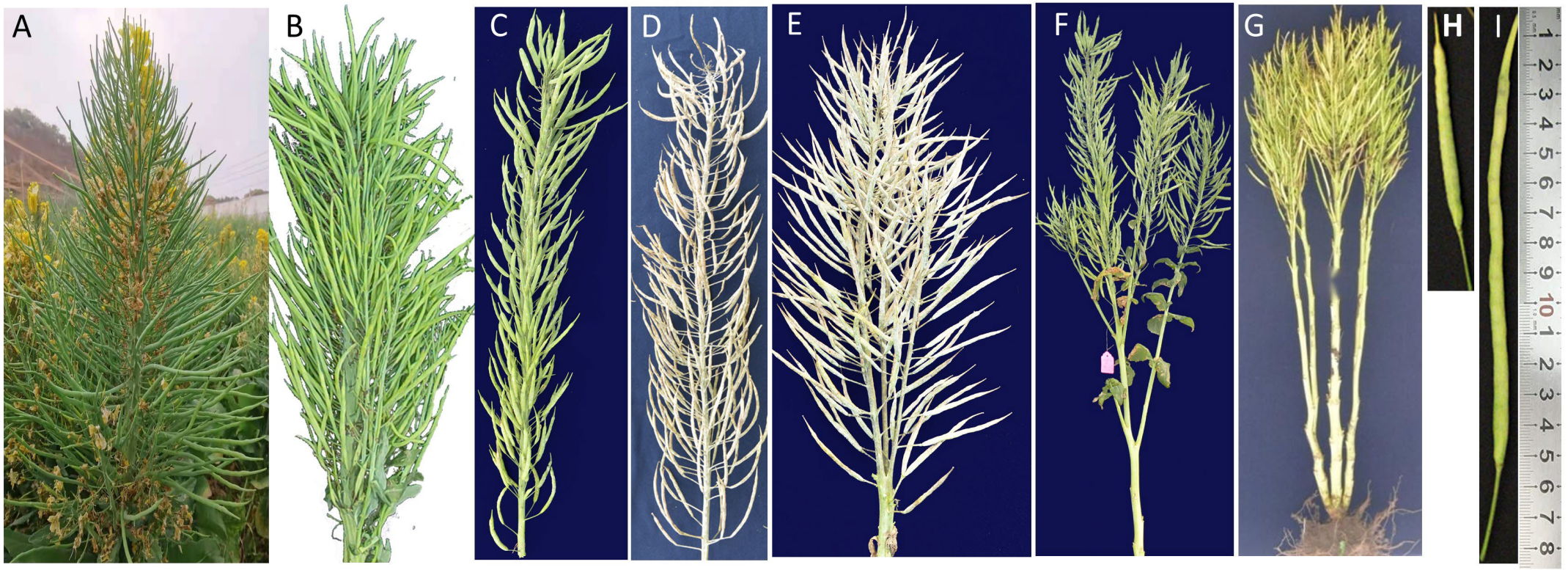
High seed-yield plant types for *de novo* design



A representative of
the current cultivars

Liu *et al. The Crop J.* 10:587–96

Representative genetic variation in rapeseed pod attributes required for the target ideotype OSR



But we need to create or find more, probably have to combine BGD and genome editing

I hope BGD can make certain contribution to this target. Within this BGD, genome editing may be a necessary tool to improve some key traits that are controlled by tightly linked genes or pleiotropic genes which have both good and bad effects.

Acknowledgement

Bioinformatics teams in my lab:

Dr. Yuanyuan Zhang & team

Dr. Chaobo Tong & team

Dr. Meili Xie & team

Wet experiment team in my lab:

Dr. Lijiang Liu & team

Dr. Feng GAo & team

Dr. Zetao Bai & team

……

External collaborators:

BGI-Shengzhen

NovoGene

BenaGene

Prof Qingyong Yang, HZAU

Dr. Harsh Raman, Wagga Wagga Institute

Dr. Yang Xiang, Guizhou Rapeseed Institute

Funding bodies:

Natural Science Foundation of China

National Key Research and Development Program of China

National Basic Research Program (“973”)

China Agriculture Research System (CARS-13)

Agricultural Science and Technology Innovation Program

Thanks!