ENHANCING CROP RESILIENCE: GENOMIC INSIGHTS AND NOVEL SOURCES OF DISEASE RESISTANCE IN WINTER WHEAT

AARHUS
UNIVERSITY
DEPARTMENT OF AGROECOLOGY

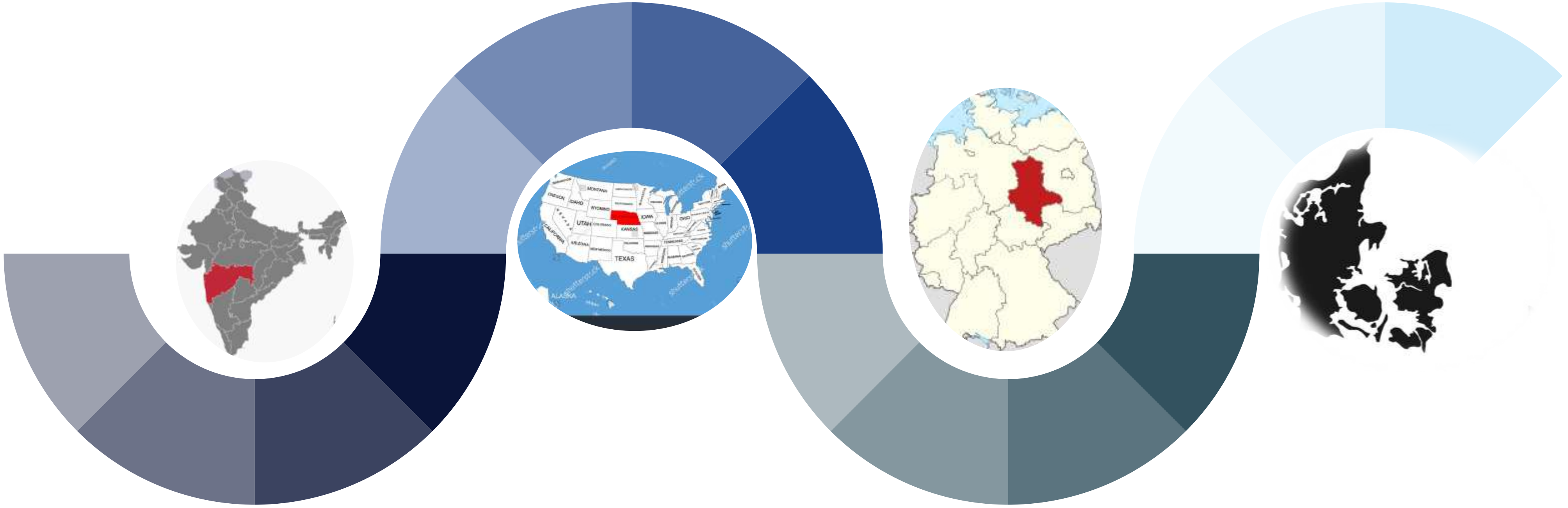
IWGSC WEBINAR
20 JUNE 2024

SANDIP MALLIKARJUN KALE
TENURE TRACK ASSISTANT PROFESSOR

India to Denmark via Germany

2013: Visiting Scientist

2021-2023 : Scientist (Carlsberg A/S)



2007-09: M.Sc (Biotechnology)
2009-14: Ph.D (Biotechnology)
2014-17: Project Scientist

2017-21:
Research Scientist

2023:
Assistant Professor
CGB, Agroecology

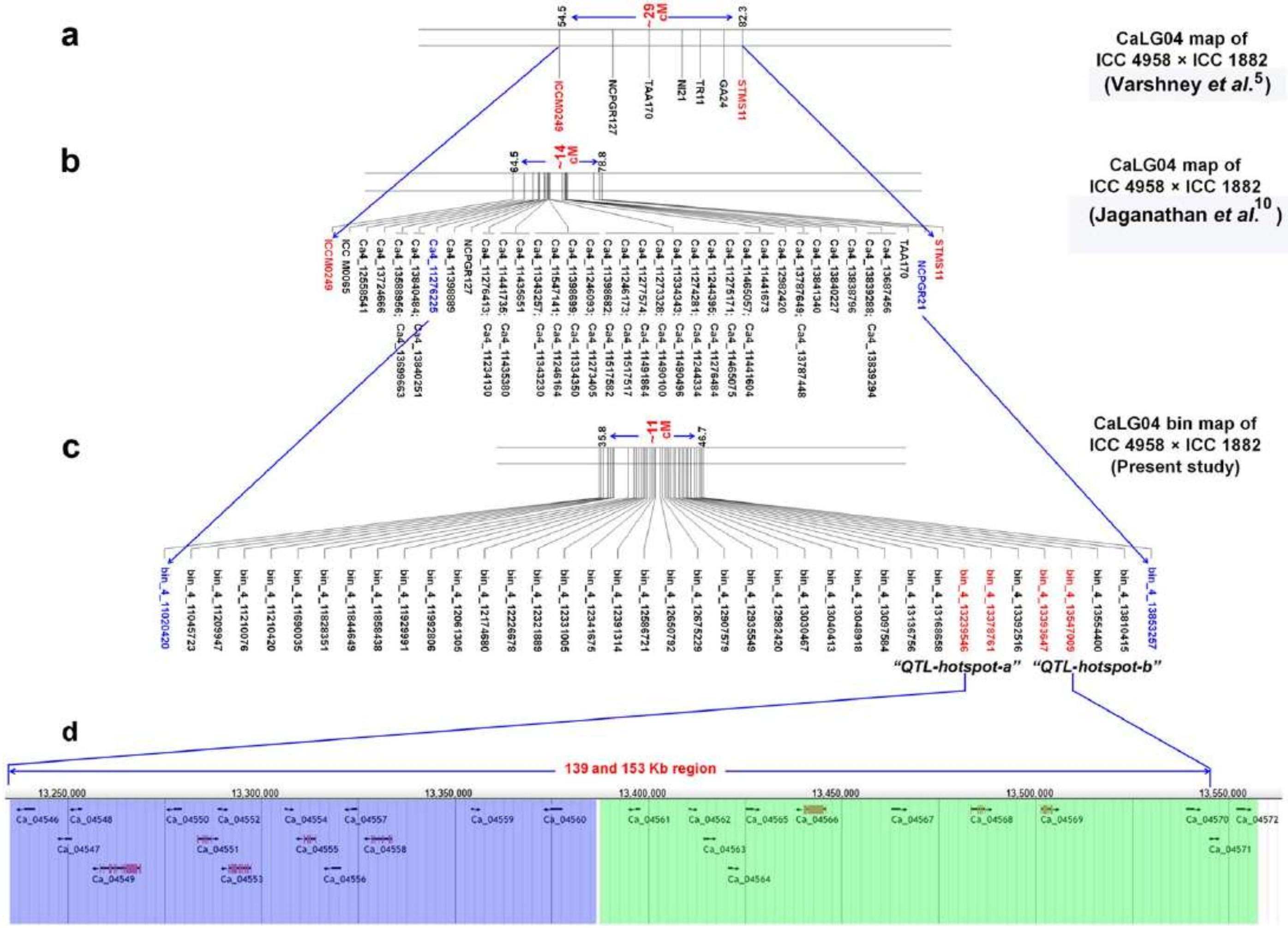
2017- 2022: IPK
Research Scientist

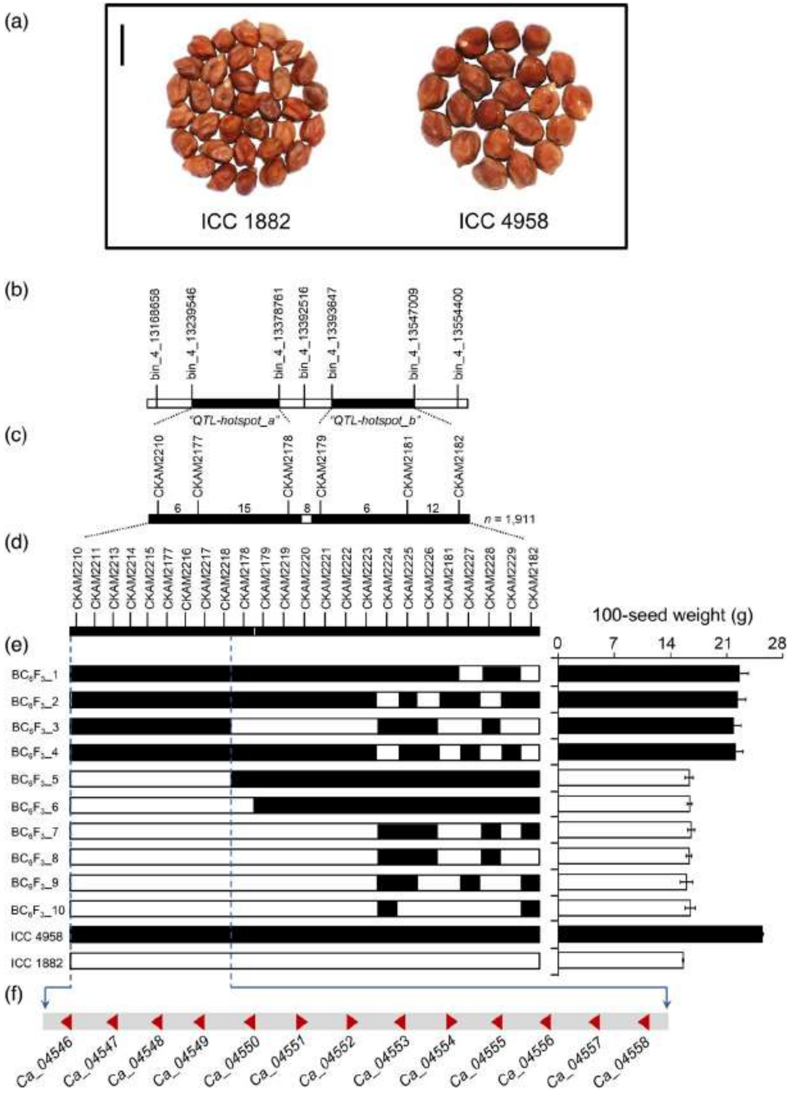
2009-14: Ph.D (Biotechnology): CSIR-NCL
Development and application of DNA markers for genetic improvement of linseed

2014-2017: ICRISAT
Special Project Scientist

MultiOmic approaches to fine map "*QTL-hotspot*" in chickpea

Kale et al., 2015





Genetic variation in *CaTIFY4b* contributes to drought adaptation in chickpea

Barmukh et al., 2022

nature genetics

Article https://doi.org/10.1038/s41588-022-01189-7

Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement

Received: 8 December 2021

Accepted: 18 August 2022

Albert W. Schulthess[1,11], Sandip M. Kale[1,9,11], Fang Liu[1,10,11], Yusheng Zhao[1], Norman Philipp[1], Maximilian Rembe[1], Yong Jiang[1], Ulrike Beukert[2], Albrecht Serfling[2], Axel Himmelbach[1], Jörg Fuchs[1], Markus Oppermann[1], Stephan Weise[1], Philipp H. G. Boeven[3], Johannes Schacht[3], C. Friedrich H. Longin[4], Sonja Kollers[5], Nina Pfeiffer[6], Viktor Korzun[5], Matthias Lange[1], Uwe Scholz[1], Nils Stein[1,7], Martin Mascher[1,8] ✉ and Jochen C. Reif[1] ✉

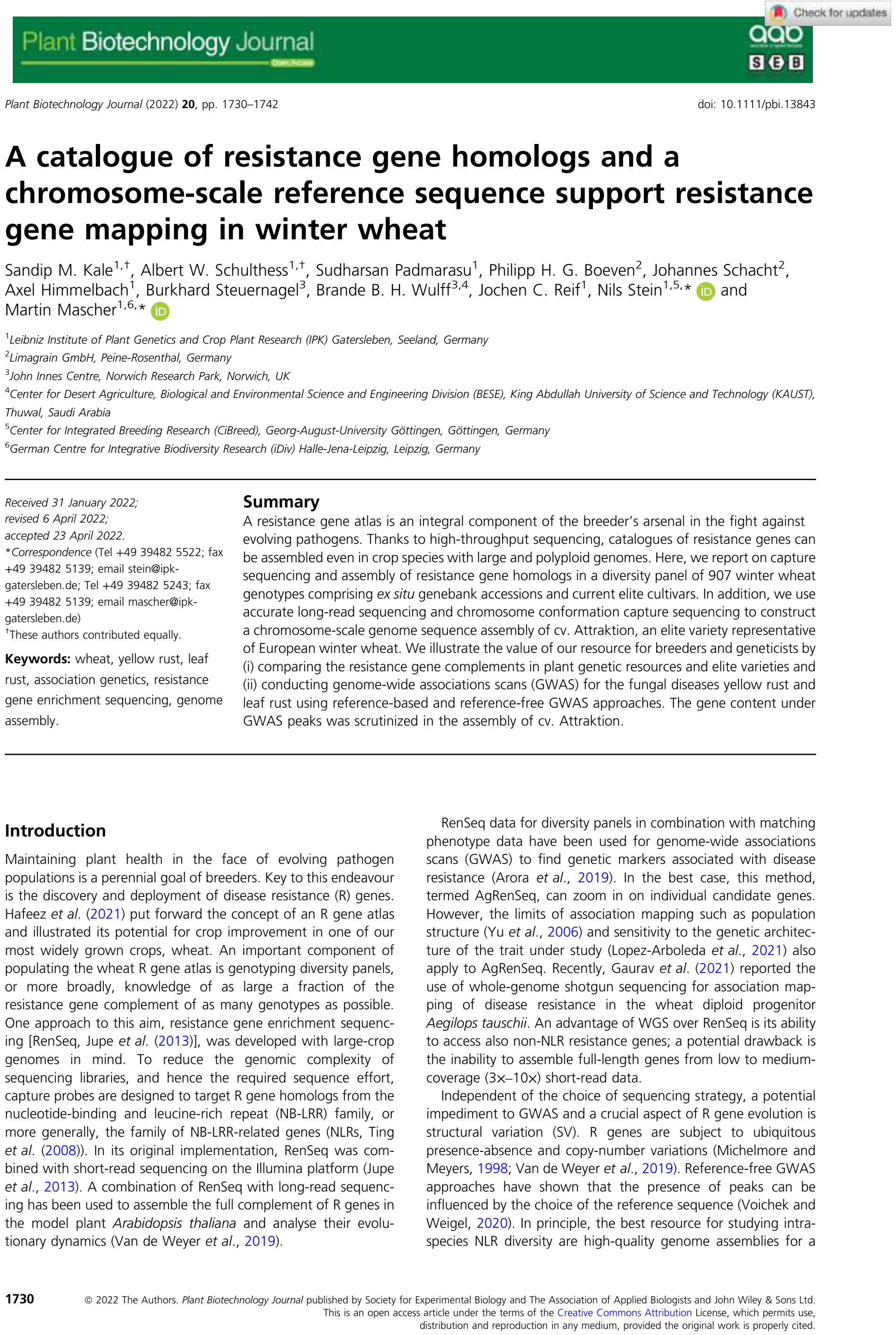
The great efforts spent in the maintenance of past diversity in genebanks are rationalized by the potential role of plant genetic resources (PGR) in future crop improvement—a concept whose practical implementation has fallen short of expectations. Here, we implement a genomics-informed prebreeding strategy for wheat improvement that does not discriminate against nonadapted germplasm. We collect and analyze dense genetic profiles for a large winter wheat collection and evaluate grain yield and resistance to yellow rust (YR) in bespoke core sets. Breeders already profit from wild introgressions but PGR still offer useful, yet unused, diversity. Potential donors of resistance sources not yet deployed in breeding were detected, while the prebreeding contribution of PGR to yield was estimated through 'Elite × PGR' F_1 crosses. Genomic prediction within and across genebanks identified the best parents to be used in crosses with elite cultivars whose advanced progenies can outyield current wheat varieties in multiple field trials.

Genebanks are committed worldwide to maintain PGR, some of which have not been grown on farmer's field for a century or never been used as crops at all[1]. PGR underperform dramatically in current agricultural environments[2]. Most of them, for example, succumb to pathogens currently at large, preventing an unbiased assessment of their breeding value[3]. Prebreeders often end up in choosing 'exotic' genotypes too closely related to the elite gene pool or with the inadvertent loss of novel haplotypes by selection in the field[4,5]. We and others[6,7] have bemoaned the disconnect between genebank management and breeding resulting from a lack of effective and generally applicable strategies to identify valuable germplasm as donors in prebreeding programs.

---

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany. [2]Julius Kühn Institute (Federal Research Centre for Cultivated Plants), Quedlinburg, Germany. [3]Limagrain GmbH, Peine-Rosenthal, Germany. [4]State Plant Breeding Institute, University of Hohenheim, Stuttgart, Germany. [5]KWS SAAT SE & Co. KGaA, Einbeck, Germany. [6]KWS LOCHOW GmbH, Bergen, Germany. [7]Center for Integrated Breeding Research (CiBreed), Georg-August-University, Göttingen, Germany. [8]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [9]Present address: Carlsberg Research Laboratory, Copenhagen, Denmark. [10]Present address: Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China. [11]These authors contributed equally: Albert W. Schulthess, Sandip M. Kale, Fang Liu. ✉e-mail: mascher@ipk-gatersleben.de; reif@ipk-gatersleben.de

---



Plant Biotechnology Journal

Plant Biotechnology Journal (2022) **20**, pp. 1730–1742 doi: 10.1111/pbi.13843

A catalogue of resistance gene homologs and a chromosome-scale reference sequence support resistance gene mapping in winter wheat

Sandip M. Kale[1,†], Albert W. Schulthess[1,†], Sudharsan Padmarasu[1], Philipp H. G. Boeven[2], Johannes Schacht[2], Axel Himmelbach[1], Burkhard Steuernagel[3], Brande B. H. Wulff[3,4], Jochen C. Reif[1], Nils Stein[1,5,*] and Martin Mascher[1,6,*]

[1]*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany*
[2]*Limagrain GmbH, Peine-Rosenthal, Germany*
[3]*John Innes Centre, Norwich Research Park, Norwich, UK*
[4]*Center for Desert Agriculture, Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia*
[5]*Center for Integrated Breeding Research (CiBreed), Georg-August-University Göttingen, Göttingen, Germany*
[6]*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany*

Received 31 January 2022;
revised 6 April 2022;
accepted 23 April 2022.

**Correspondence* (Tel +49 39482 5522; fax +49 39482 5139; email stein@ipk-gatersleben.de; Tel +49 39482 5243; fax +49 39482 5139; email mascher@ipk-gatersleben.de)
[†]These authors contributed equally.

Keywords: wheat, yellow rust, leaf rust, association genetics, resistance gene enrichment sequencing, genome assembly.

Summary

A resistance gene atlas is an integral component of the breeder's arsenal in the fight against evolving pathogens. Thanks to high-throughput sequencing, catalogues of resistance genes can be assembled even in crop species with large and polyploid genomes. Here, we report on capture sequencing and assembly of resistance gene homologs in a diversity panel of 907 winter wheat genotypes comprising *ex situ* genebank accessions and current elite cultivars. In addition, we use accurate long-read sequencing and chromosome conformation capture sequencing to construct a chromosome-scale genome sequence assembly of cv. Attraktion, an elite variety representative of European winter wheat. We illustrate the value of our resource for breeders and geneticists by (i) comparing the resistance gene complements in plant genetic resources and elite varieties and (ii) conducting genome-wide associations scans (GWAS) for the fungal diseases yellow rust and leaf rust using reference-based and reference-free GWAS approaches. The gene content under GWAS peaks was scrutinized in the assembly of cv. Attraktion.

Introduction

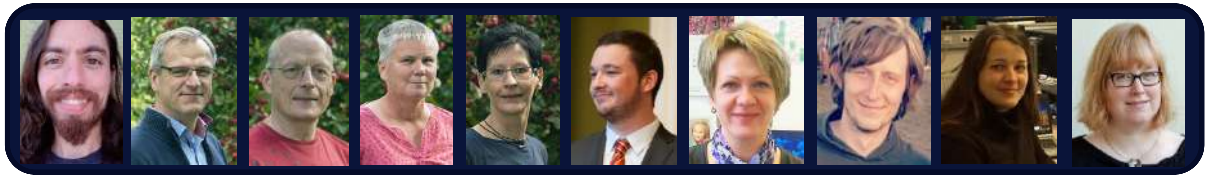
Maintaining plant health in the face of evolving pathogen populations is a perennial goal of breeders. Key to this endeavour is the discovery and deployment of disease resistance (R) genes. Hafeez *et al*. (2021) put forward the concept of an R gene atlas and illustrated its potential for crop improvement in one of our most widely grown crops, wheat. An important component of populating the wheat R gene atlas is genotyping diversity panels, or more broadly, knowledge of as large a fraction of the resistance gene complement of as many genotypes as possible. One approach to this aim, resistance gene enrichment sequencing [RenSeq, Jupe *et al*. (2013)], was developed with large-crop genomes in mind. To reduce the genomic complexity of sequencing libraries, and hence the required sequence effort, capture probes are designed to target R gene homologs from the nucleotide-binding and leucine-rich repeat (NB-LRR) family, or more generally, the family of NB-LRR-related genes (NLRs, Ting *et al*. (2008)). In its original implementation, RenSeq was combined with short-read sequencing on the Illumina platform (Jupe *et al*., 2013). A combination of RenSeq with long-read sequencing has been used to assemble the full complement of R genes in the model plant *Arabidopsis thaliana* and analyse their evolutionary dynamics (Van de Weyer *et al*., 2019).

RenSeq data for diversity panels in combination with matching phenotype data have been used for genome-wide associations scans (GWAS) to find genetic markers associated with disease resistance (Arora *et al*., 2019). In the best case, this method, termed AgRenSeq, can zoom in on individual candidate genes. However, the limits of association mapping such as population structure (Yu *et al*., 2006) and sensitivity to the genetic architecture of the trait under study (Lopez-Arboleda *et al*., 2021) also apply to AgRenSeq. Recently, Gaurav *et al*. (2021) reported the use of whole-genome shotgun sequencing for association mapping of disease resistance in the wheat diploid progenitor *Aegilops tauschii*. An advantage of WGS over RenSeq is its ability to access also non-NLR resistance genes; a potential drawback is the inability to assemble full-length genes from low to medium-coverage (3×–10×) short-read data.

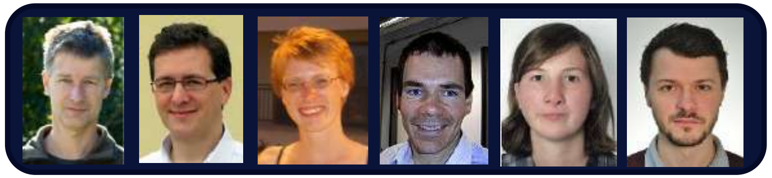
Independent of the choice of sequencing strategy, a potential impediment to GWAS and a crucial aspect of R gene evolution is structural variation (SV). R genes are subject to ubiquitous presence-absence and copy-number variations (Michelmore and Meyers, 1998; Van de Weyer *et al*., 2019). Reference-free GWAS approaches have shown that the presence of peaks can be influenced by the choice of the reference sequence (Voichek and Weigel, 2020). In principle, the best resource for studying intraspecies NLR diversity are high-quality genome assemblies for a

Acknowledgement

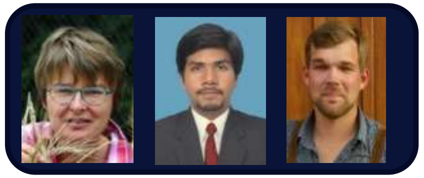
WP-1: ***Access and Genotyping***



WP-2: ***Biotic Stress***

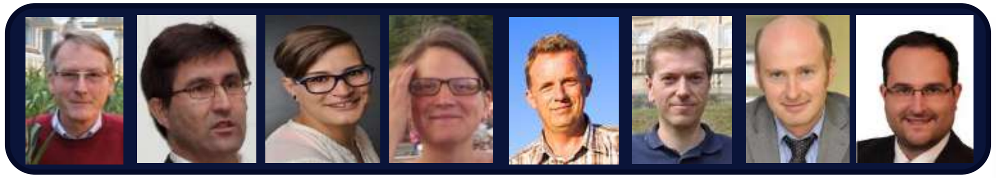


WP-3: ***Novel trait discovery***



WP-4: ***Nitrogen Use Efficiency***

WP-5: ***Pre-breeding***



WP-6: ***Biodiversity Informatics***

PtJ
Projektträger Jülich
Forschungszentrum Jülich

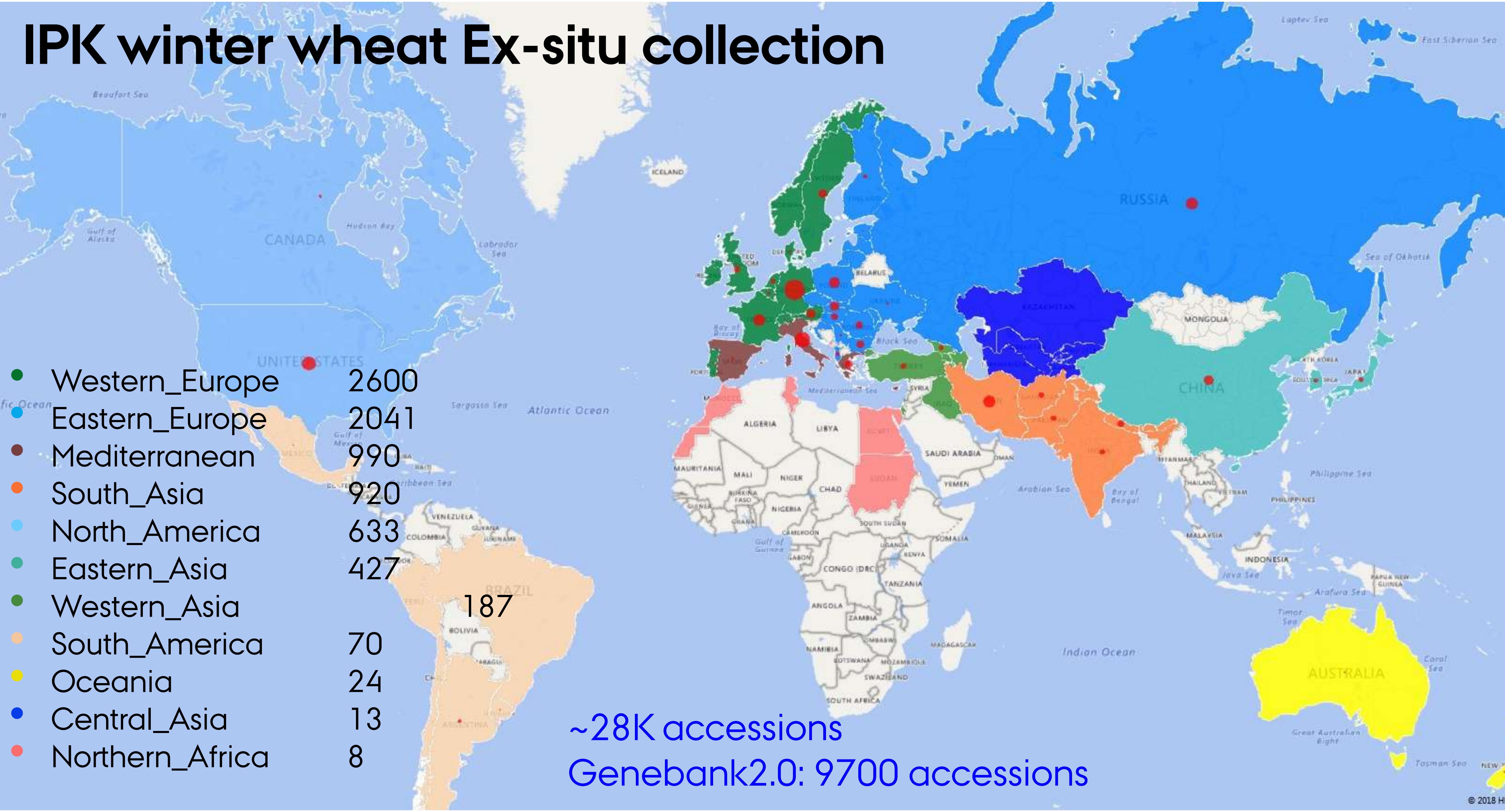
GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Gene Bank 2.0

Outline

- ☐ GBS of 9K winter wheat accessions.
- ☐ Construction of TCCC.
- ☐ WGS and introgression identification
- ☐ RenSeq and GWAS for YR and LR

IPK winter wheat Ex-situ collection



●	Western_Europe	2600
●	Eastern_Europe	2041
●	Mediterranean	990
●	South_Asia	920
●	North_America	633
●	Eastern_Asia	427
●	Western_Asia	187
●	South_America	70
●	Oceania	24
●	Central_Asia	13
●	Northern_Africa	8

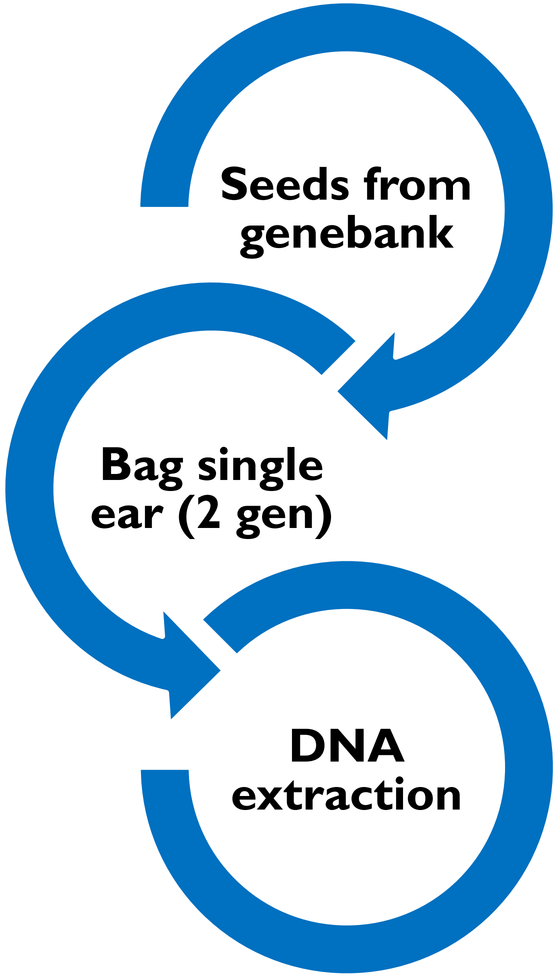
~28K accessions

Genebank2.0: 9700 accessions

Spike diversity

Disease resistance

GB2.0: Genebank activation pipeline



Seeds from genebank

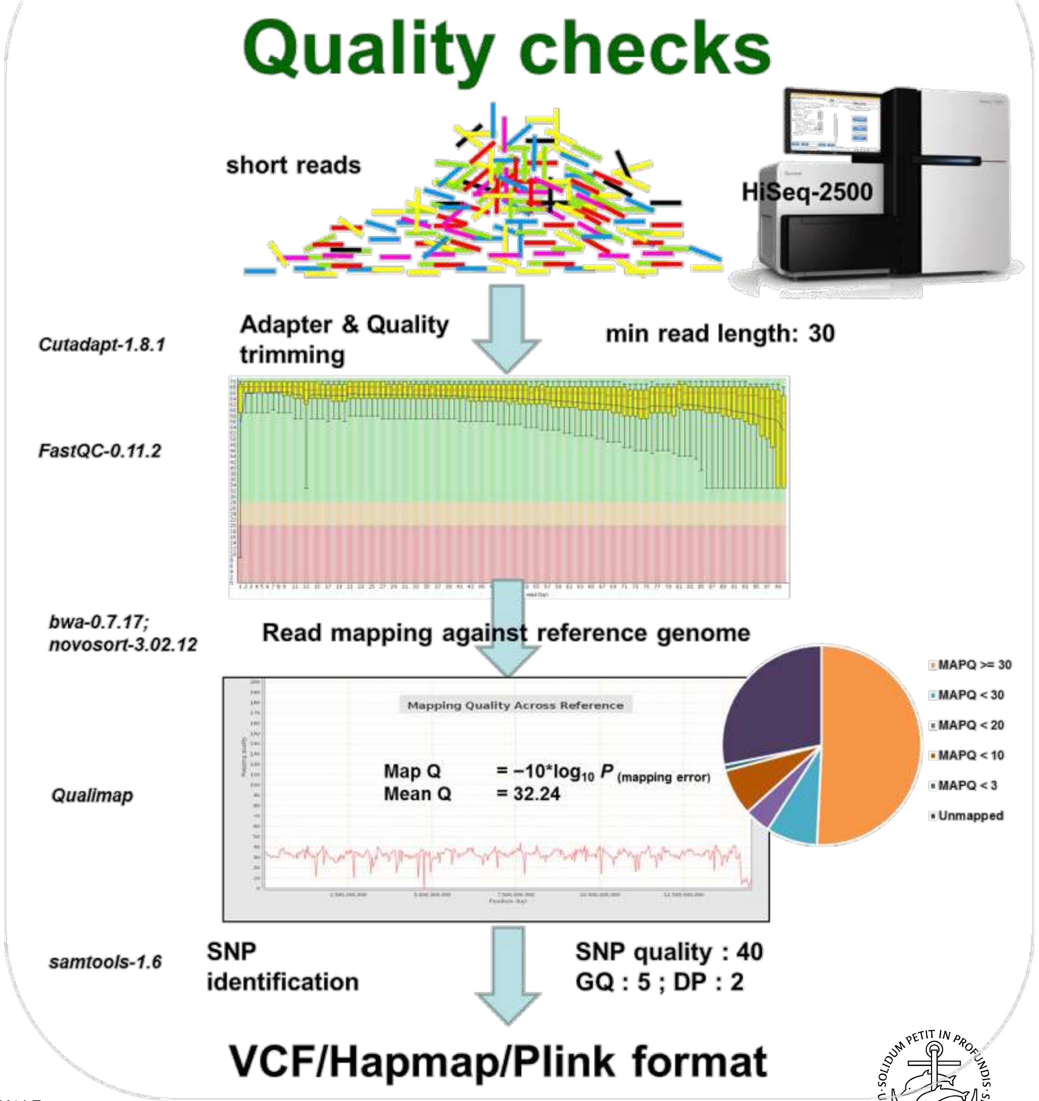
Bag single ear (2 gen)

DNA extraction

SSD purification

- REs: *PstI* and *MspI* (*Wendler et al. 2014*)
- 540-plex libraries
- ~2.50 millions reads per sample
- 8,070: 7,745 ISO-PGR + 325 Elite lines

GBS libraries



Quality checks

Cutadapt-1.8.1 Adapter & Quality trimming — min read length: 30

FastQC-0.11.2

bwa-0.7.17; novosort-3.02.12 Read mapping against reference genome

Qualimap

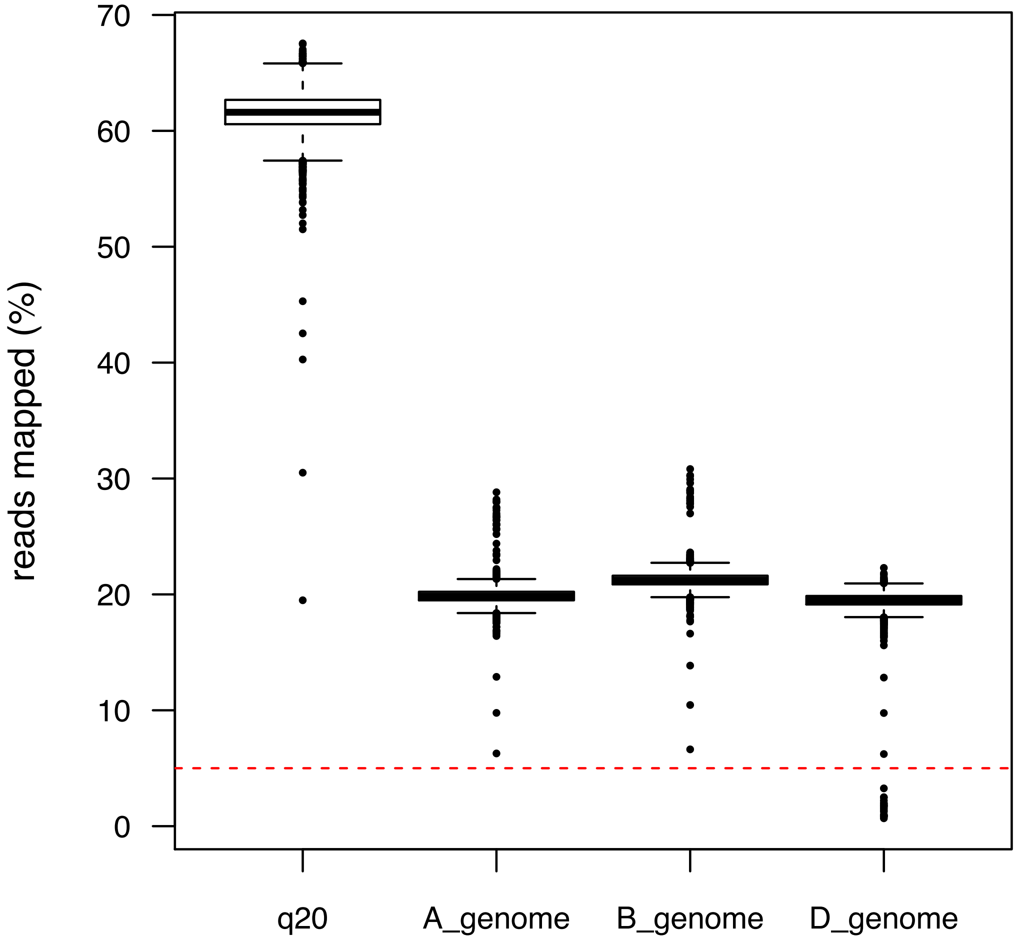
Map Q $= -10*\log_{10} P_{\text{(mapping error)}}$

Mean Q $= 32.24$

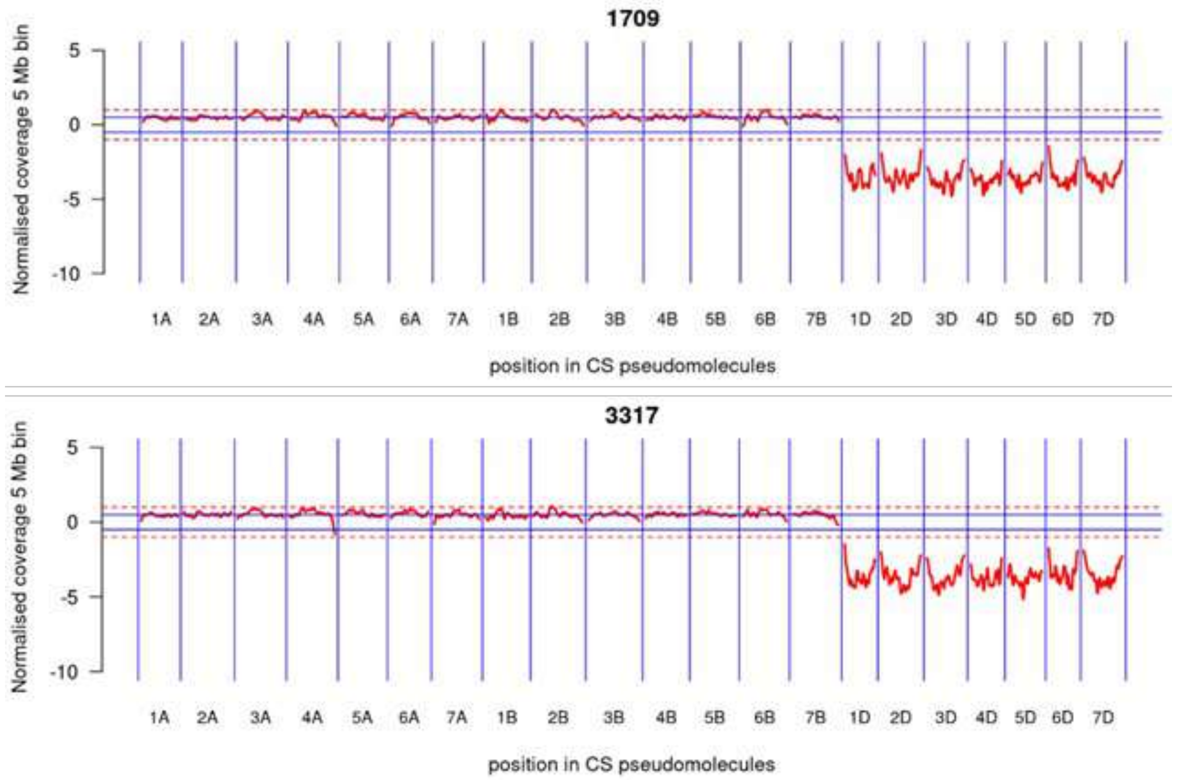
samtools-1.6 SNP identification — SNP quality : 40, GQ : 5 ; DP : 2

VCF/Hapmap/Plink format

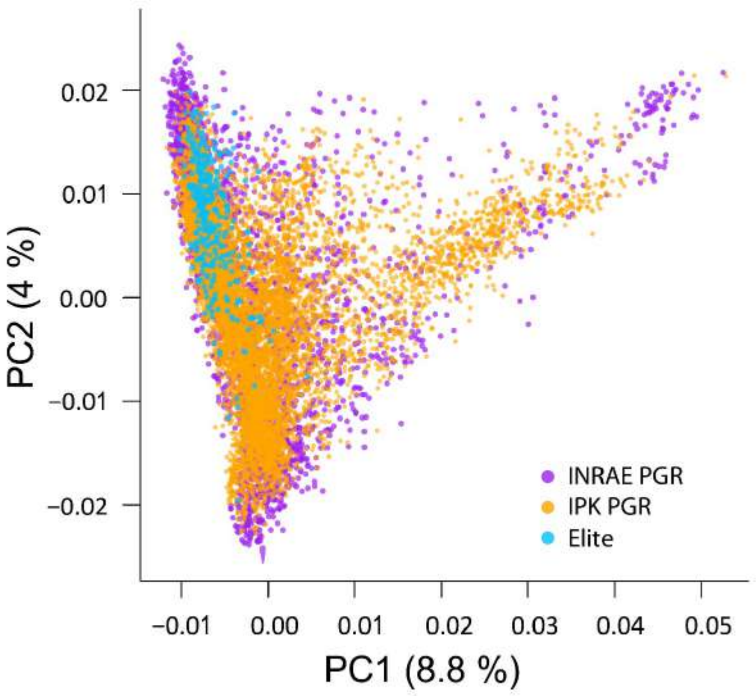
ABD or only AB? Correcting passport data

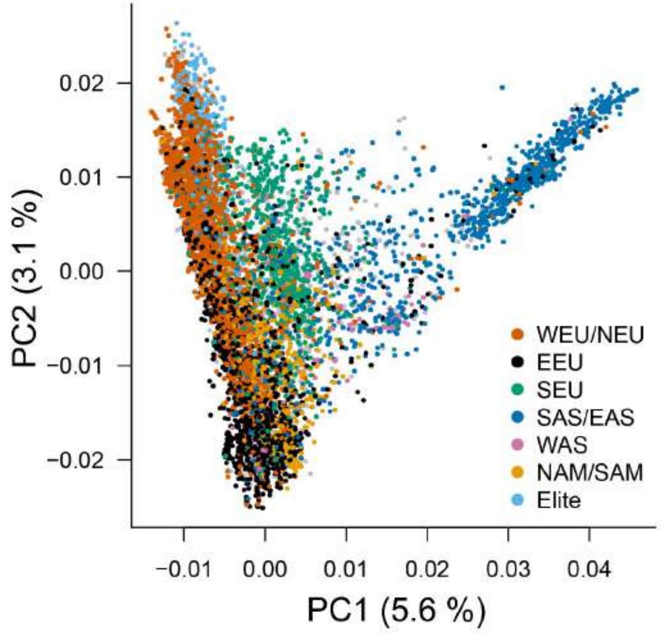


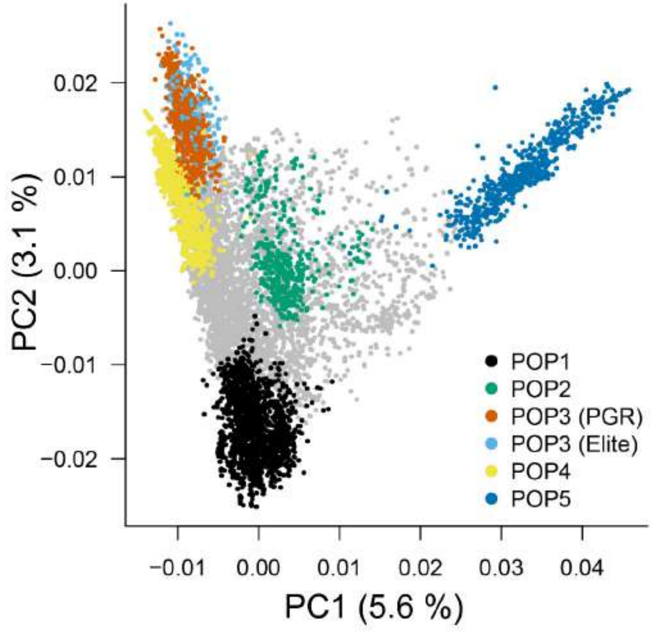
Mapping against RefSeq v1.0

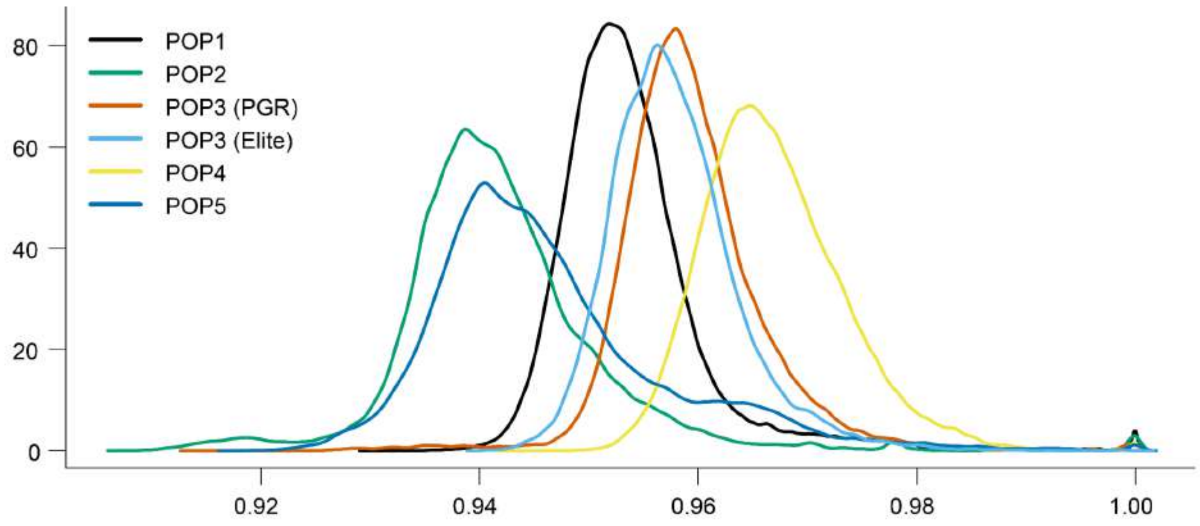


Exploring neutral diversity in the IPK genebank

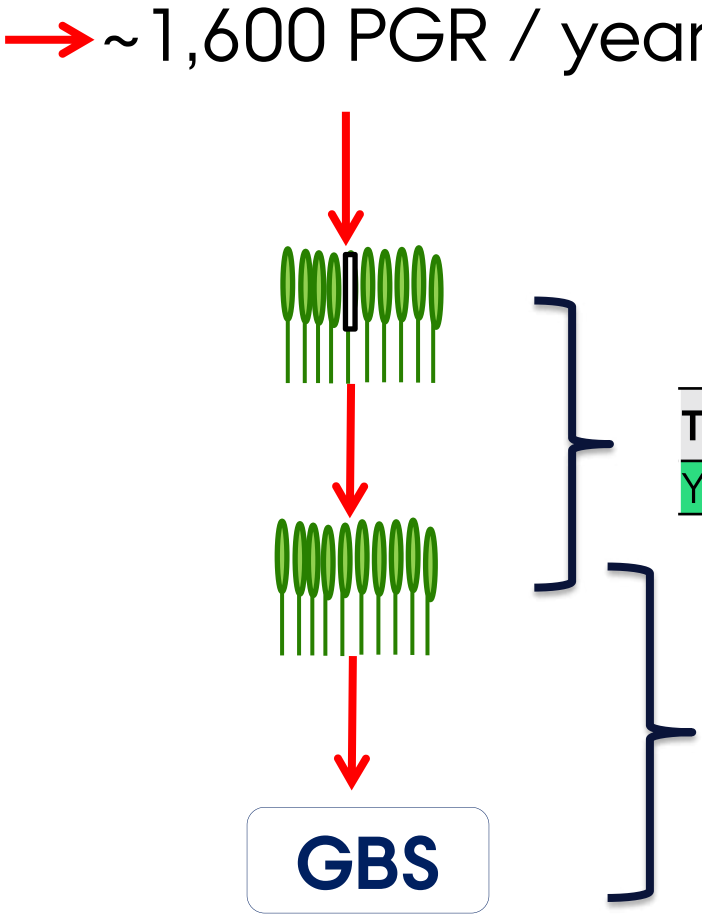








Breeding for disease resistance



Wheat diseases:

- Septoria tritici blotch
- Fusarium ear blight
- **Leaf rust**
- Tan spot
- Eyespot
- Take-all
- **Powdery mildew**
- **Yellow rust**
- Septoria nodorum
- Soil-borne mosaic virus

Trait	h^2
Yellow rust	0.82

Trait	**N ISO-PGR**
YR, LR and PM	6,300

Reference based v/s Reference free GWAS

Read Mapping

SNP identification

GWAS

- Single assembly is not enough!
- Errors in assembly
- Structural variations

Reference based GWAS

Raw reads

k-mer identification

k-mer matrix

GWAS

Significant *k*-mers

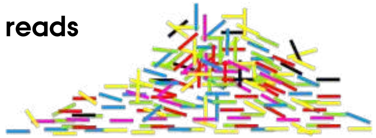
Mapping/denovo assembly

Reference free GWAS

K-mer GWAS

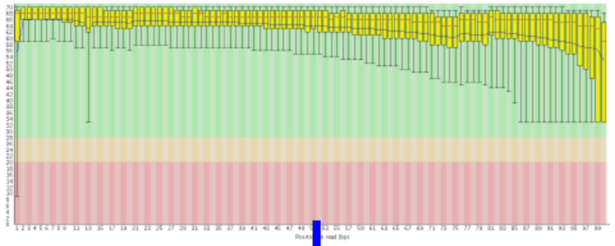
Voichek and weigel: Nat Genet. 52, 534–540(2020)

250 bp PE short reads



Adapter & Quality trimming

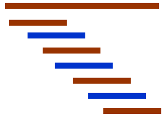
min read length: 30



Kmer counting

All k-mers without canonization

canonical k-mers (kmer with its reverse complement)



filteration

Kmer	count
K1	10
K2	20
K3	45
Kn	5

	K1	K2	K3	K4	Kn
S1	1	0	1	1	1
S2	1	1	1	0	1
S3	0	1	1	1	1
Sn	0	0	0	1	0

Kmers Pre/Abs Matrix

Phenotypic data

Permute phenotype

	K1	K2	K3	K4	Kn
S1	1	0	1	1	1
S2	1	1	1	0	1
S3	0	1	1	1	1
Sn	0	0	0	1	0

	P1
S1	-2.45
S2	-1.71
S3	-2.31
S4	-2.41
Sn	-2.39

Linear Mixed Model

Kinship matrix

Identify top N kmers based on kmer scores

Gemma V0.96

Calculate the likelihood ratio test p-values

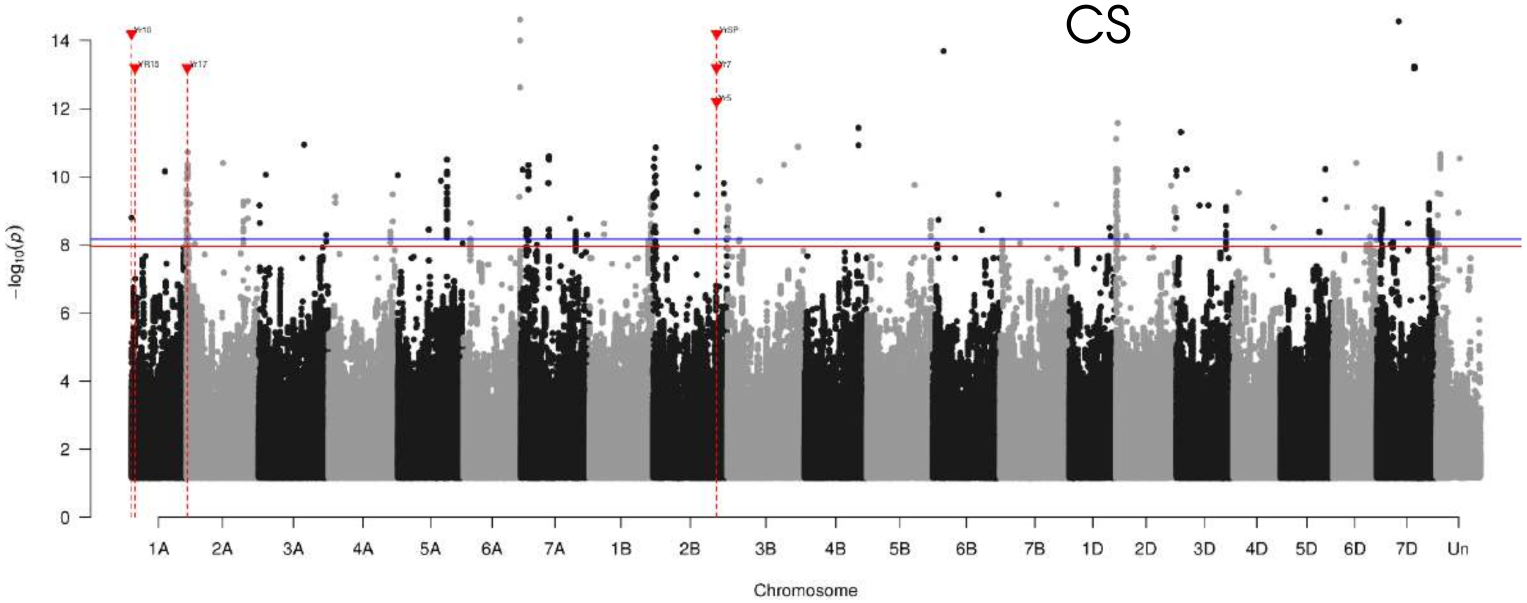
Kmers passing permutation based threshold

N permutations

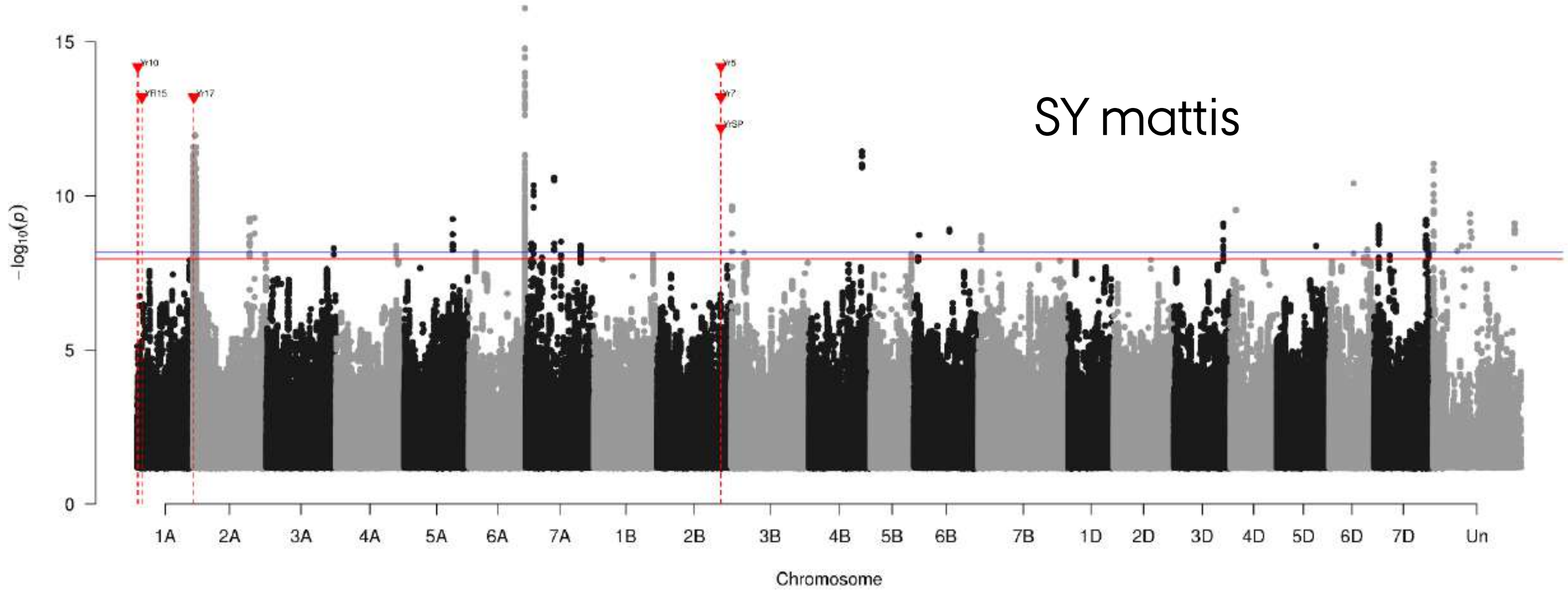
Mapping on Reference/denovo assembly

kmersGWAS for YR : Genome assembly is important!

- Number of genotypes: 4,635
- Kinship: RD (using SNPs)
- Number of significant kmers: 21,537

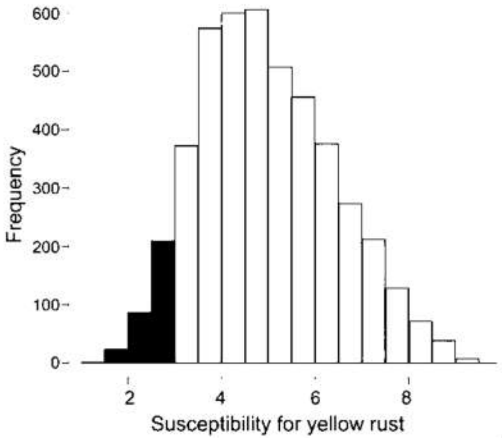


CS

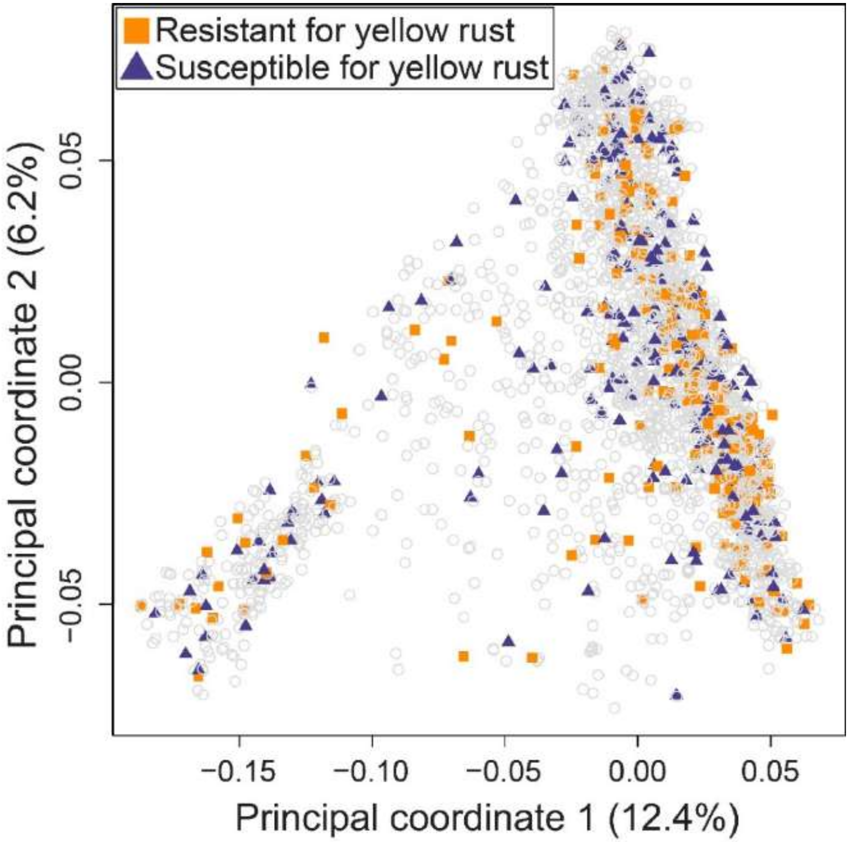


SY mattis

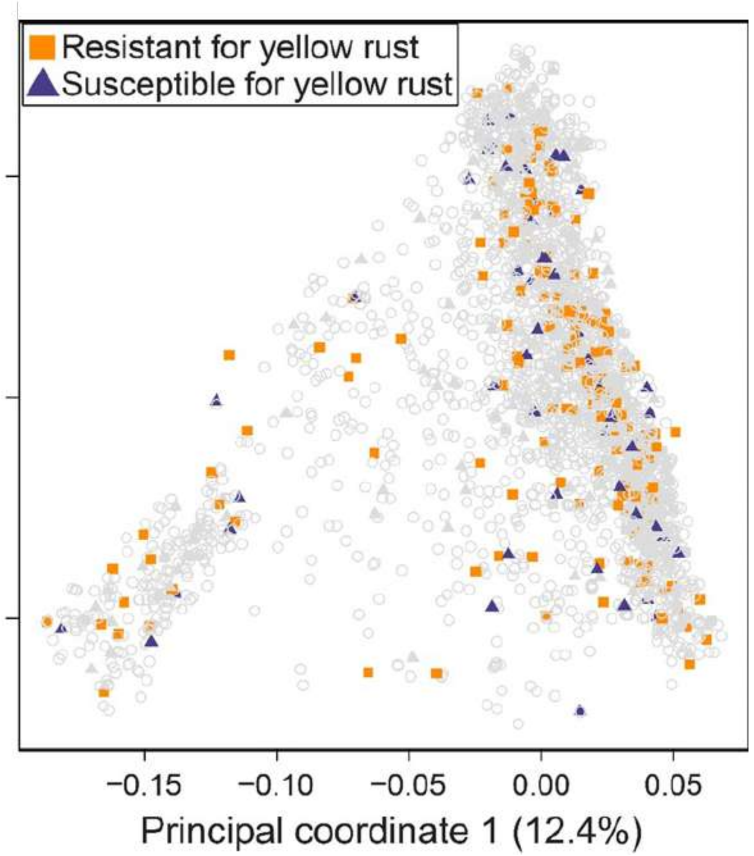
Trait-customized CC (TCCC): YR Example



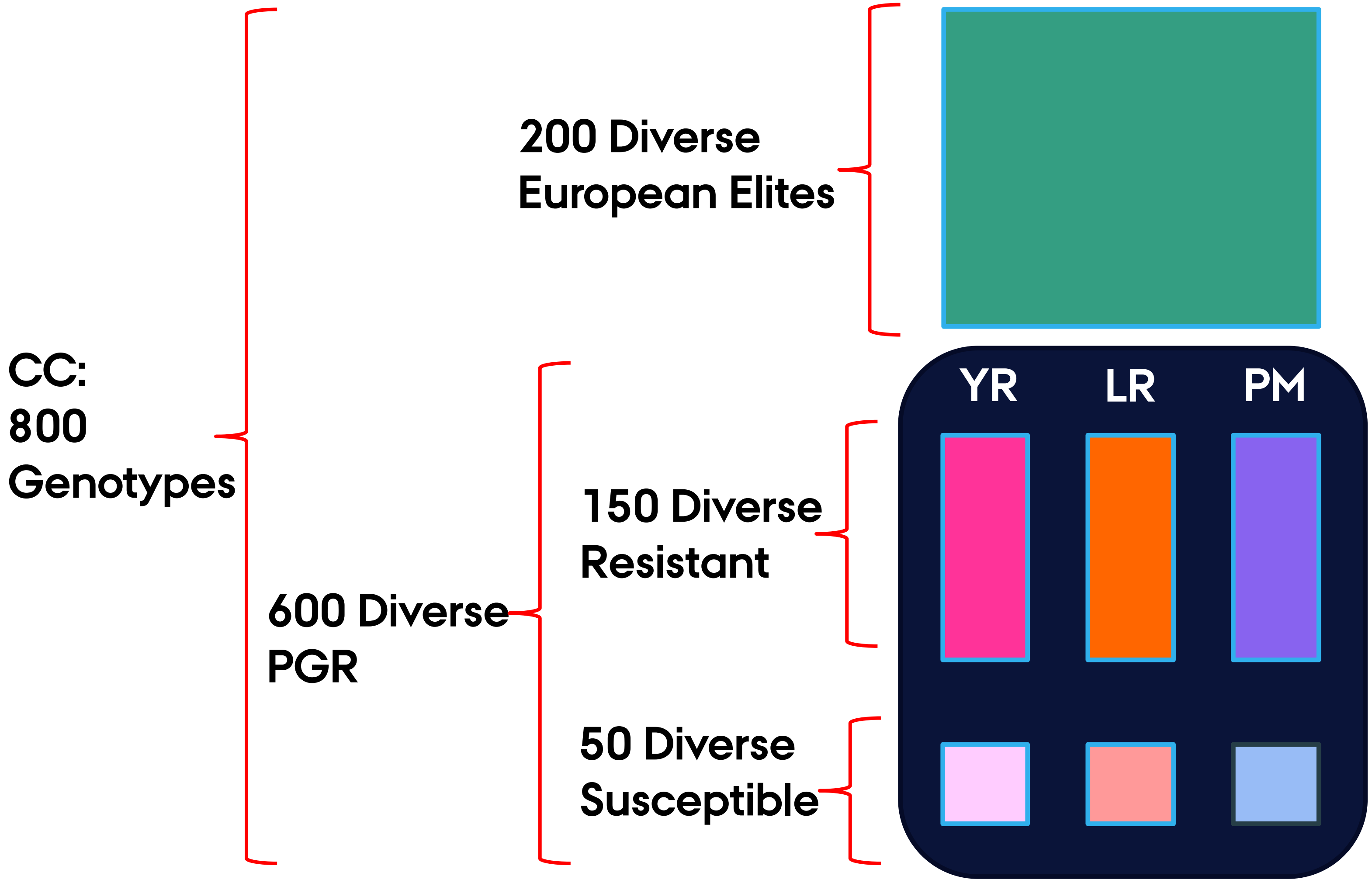
2-tailed selection



Core-collection



TCCC: Overview



CC:
800
Genotypes

200 Diverse
European Elites

600 Diverse
PGR

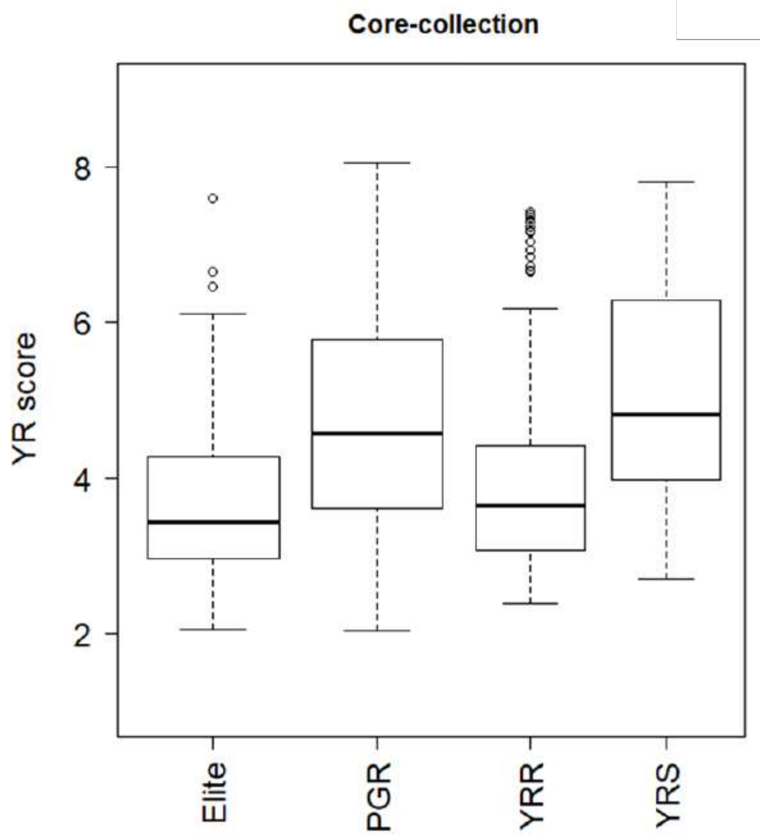
150 Diverse
Resistant

50 Diverse
Susceptible

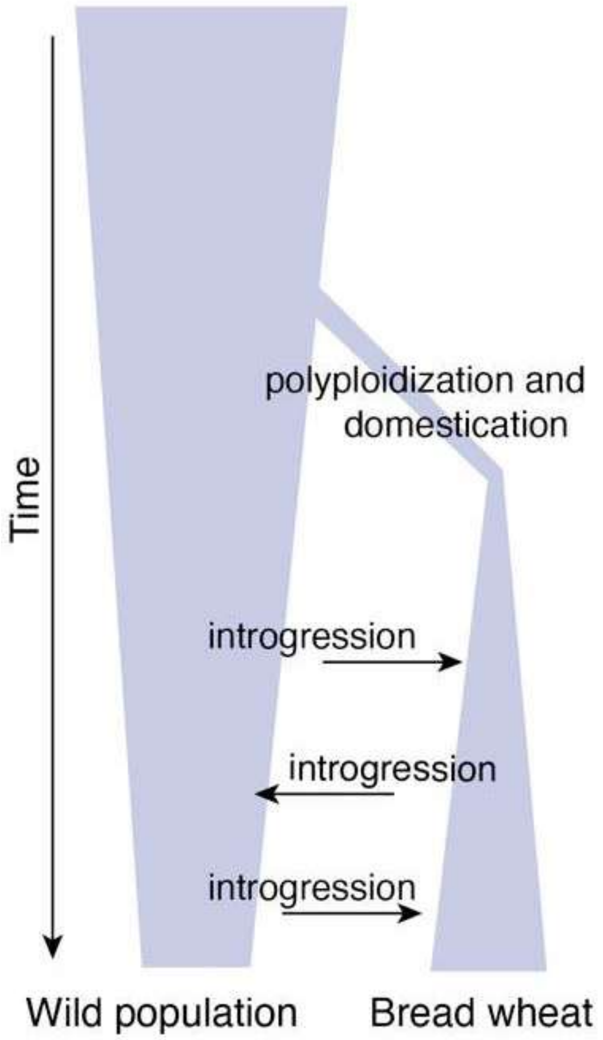
YR LR PM

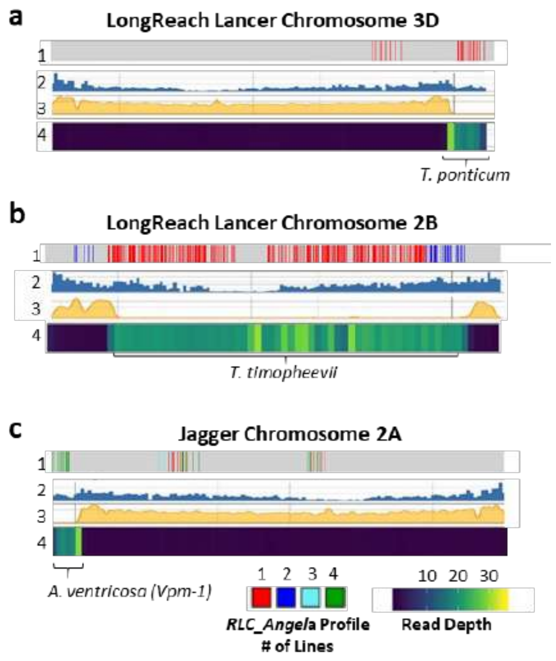
Balanced replicated field trials

Feature	YR
Env	7
h^2	0.89



Identification of alien Introgressions





Walkowiak et al., (10+ wheat genome)

1B/1R rye introgression

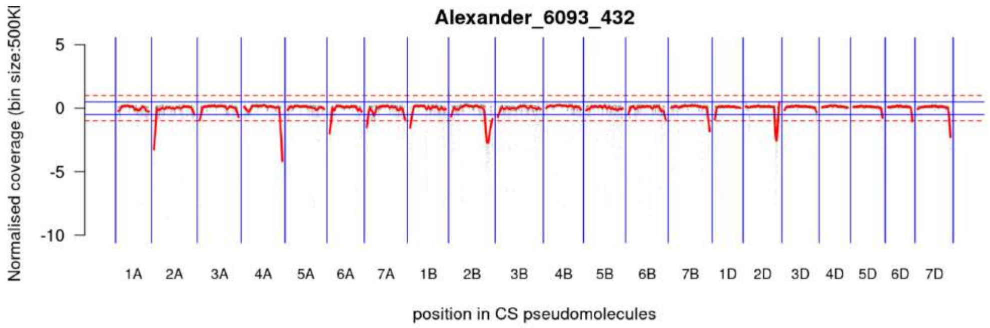
Map trimmed reads Against CS RefV1.0

Count no of reads in 500Kb windows

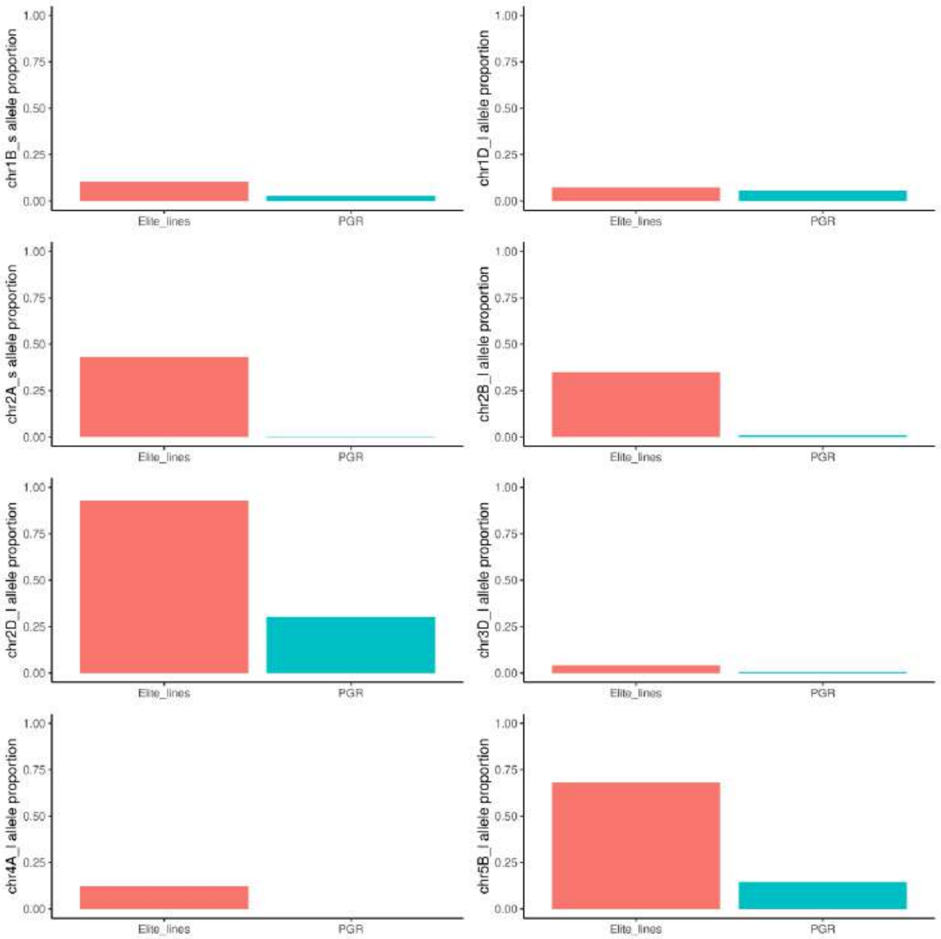
Normalise for depth

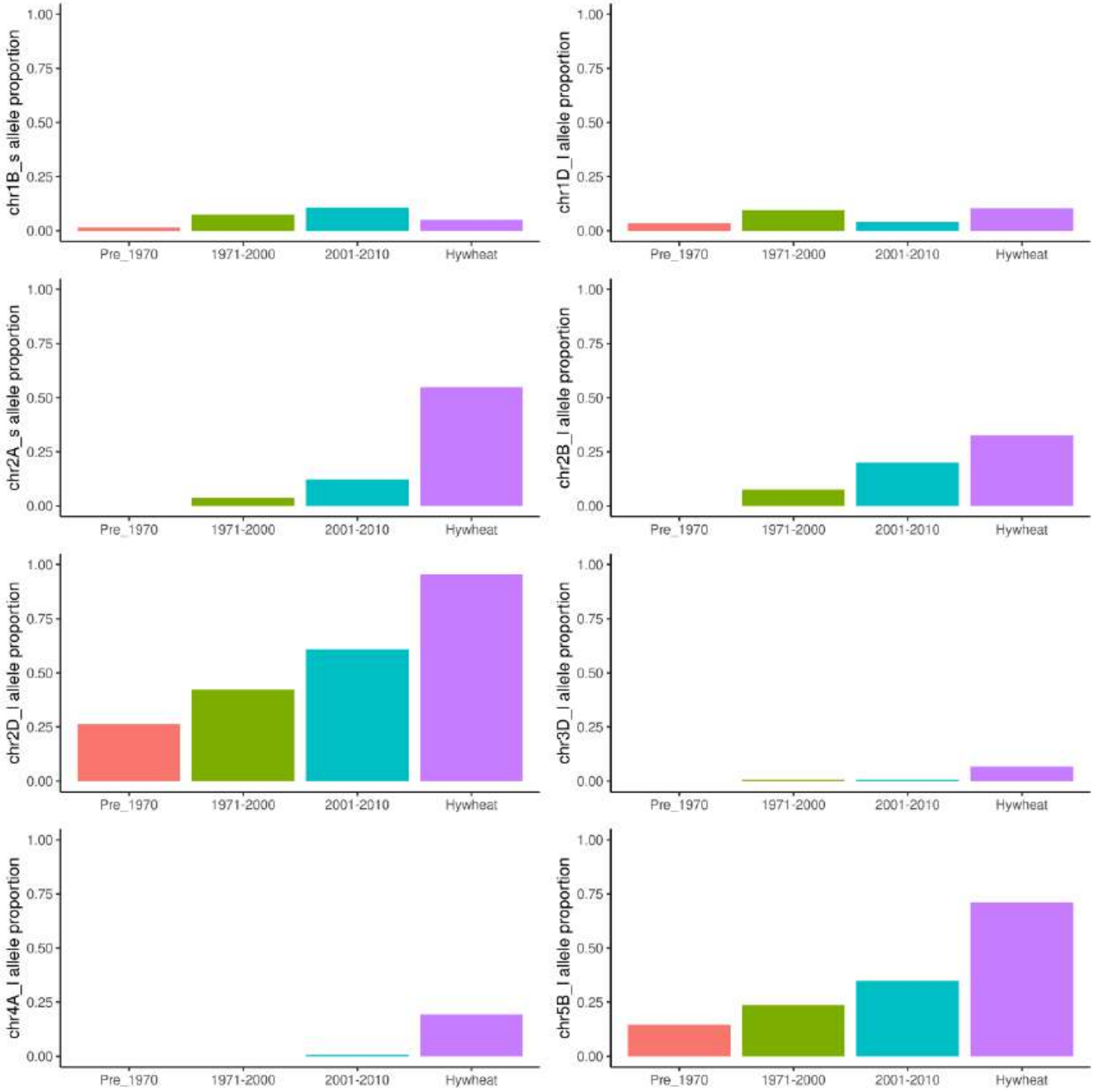
Normalise with CS read count For same window

Downstream analysis

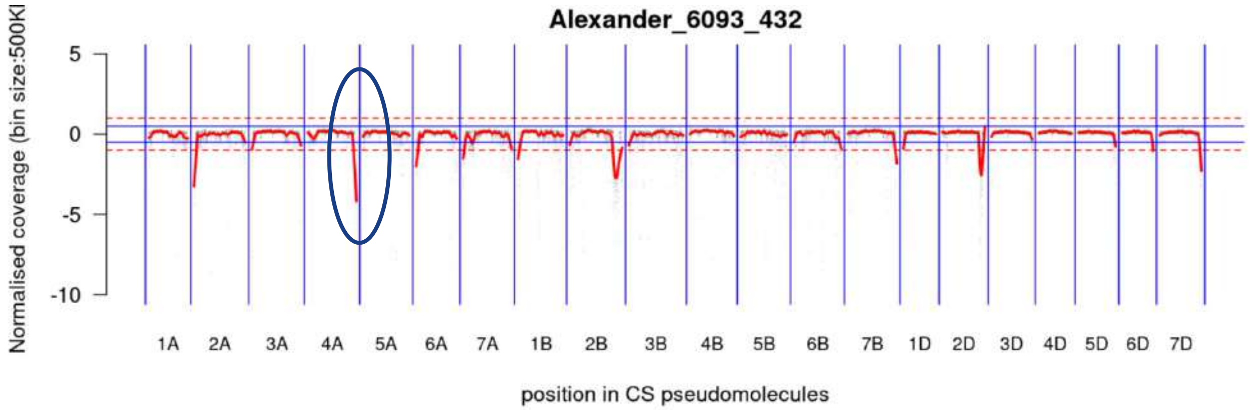


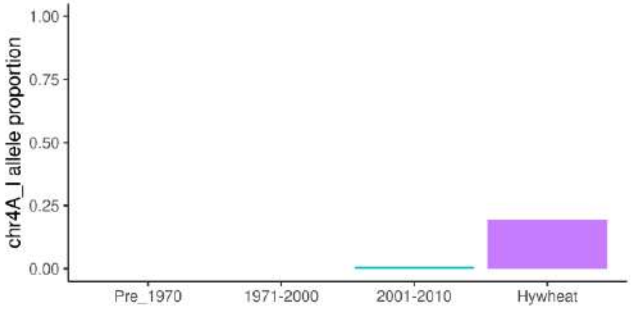
Tracing the introgression history



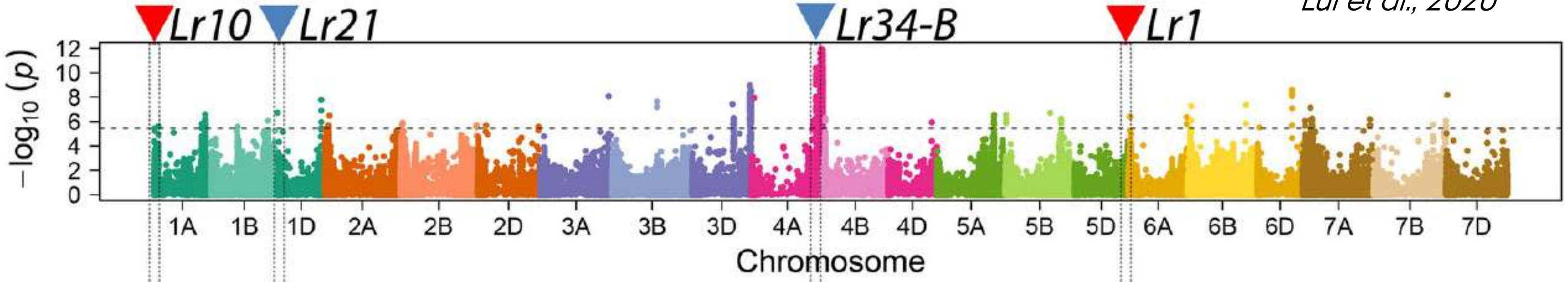


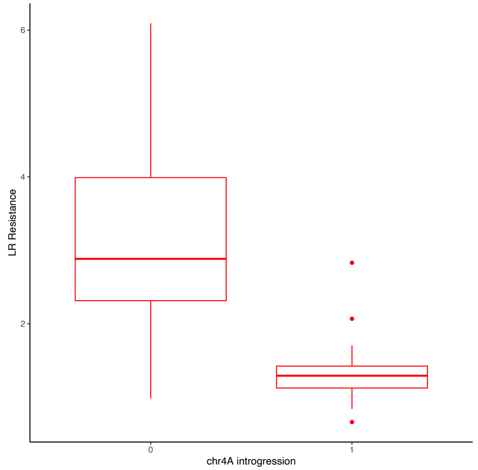
4A introgression and LR



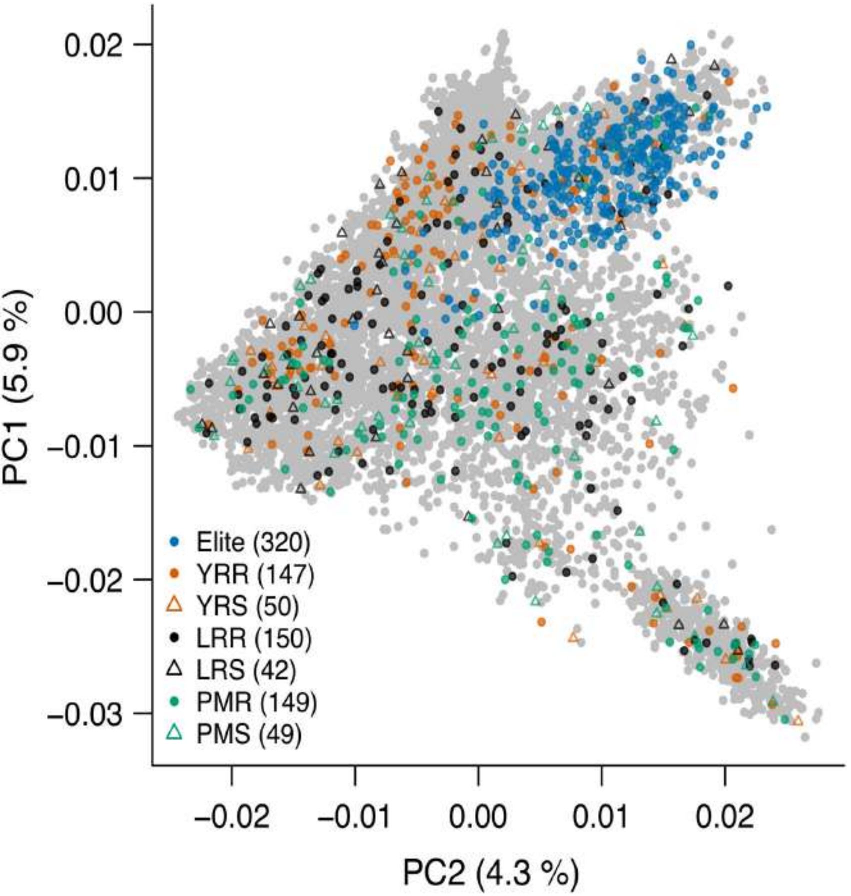


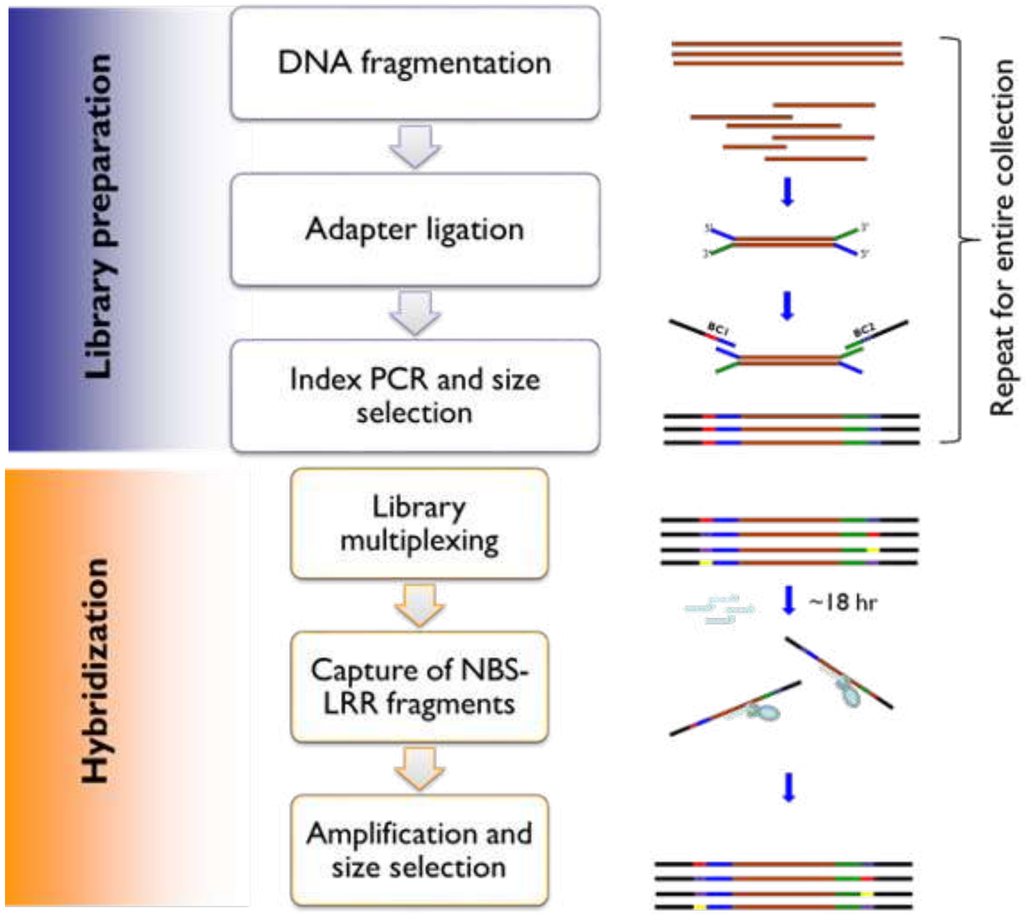
Lui et al., 2020





Resistance gene enrichment sequencing (RenSeq)



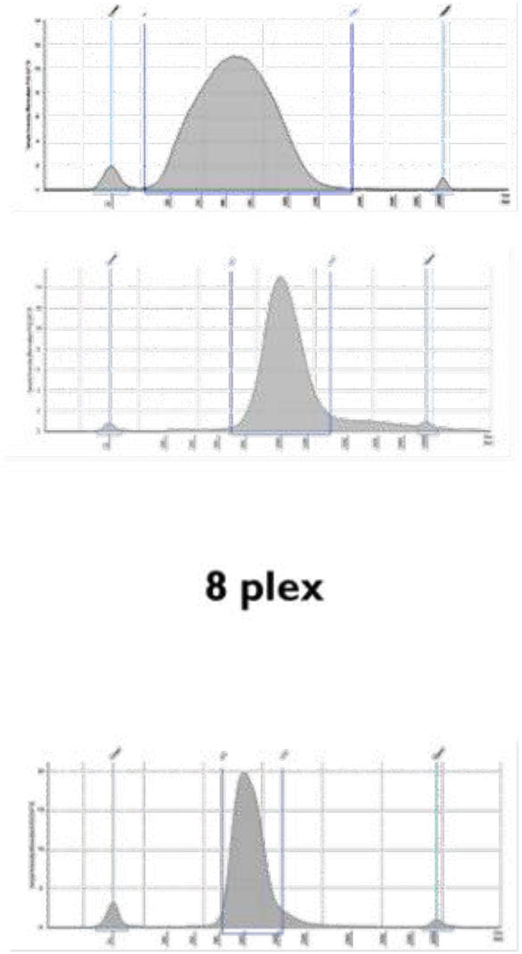


Library preparation

- DNA fragmentation
- Adapter ligation
- Index PCR and size selection

Hybridization

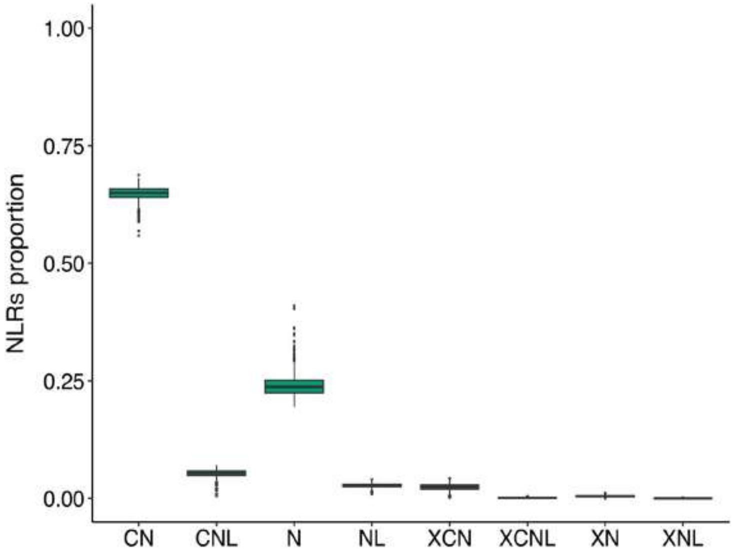
- Library multiplexing
- Capture of NBS-LRR fragments
- Amplification and size selection

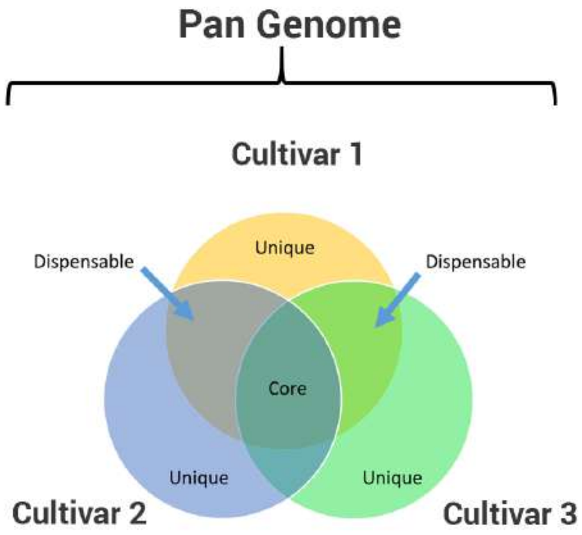


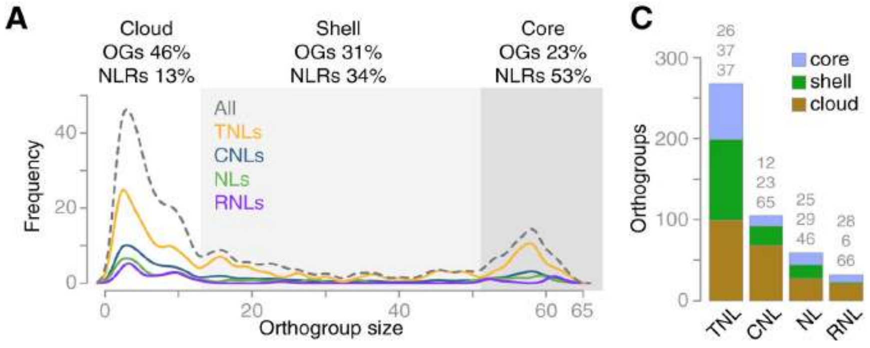
8 plex

RenSeq

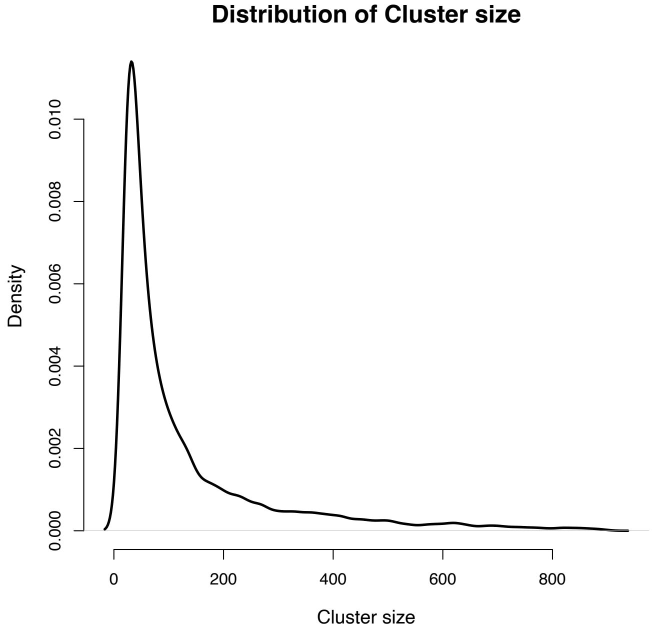
PanNLRome in hexaploid wheat





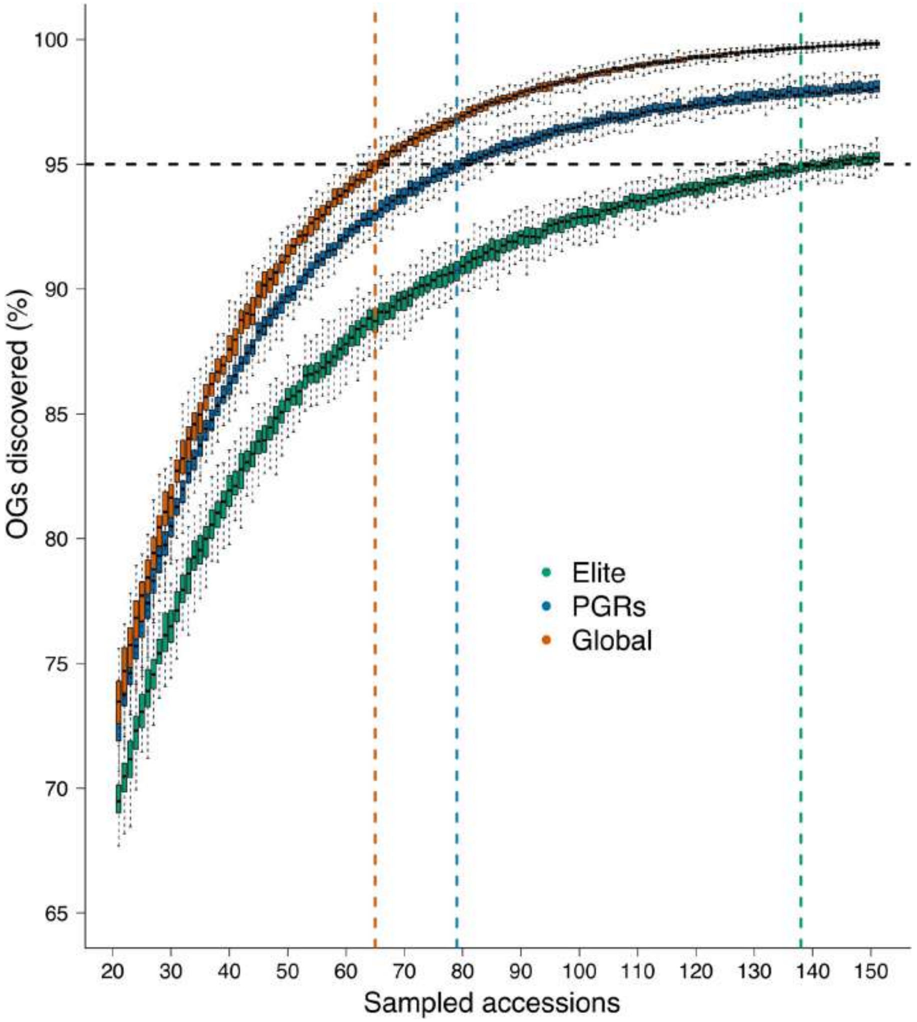


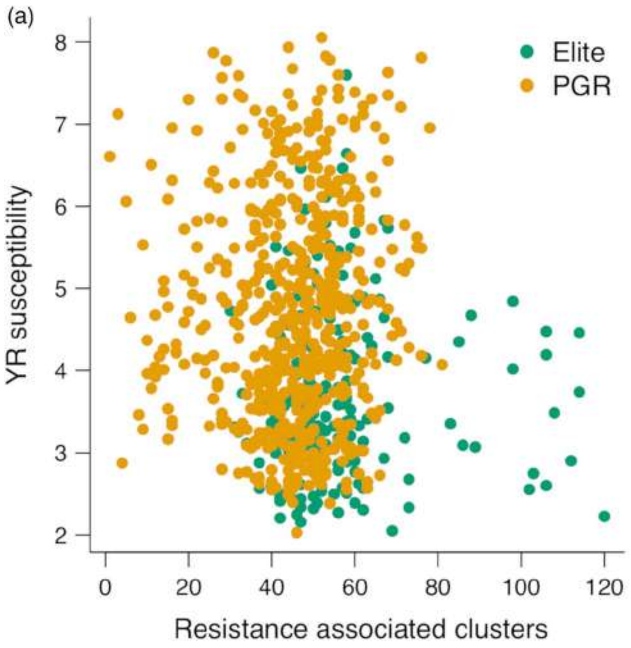
Arabidopsis panNLRome



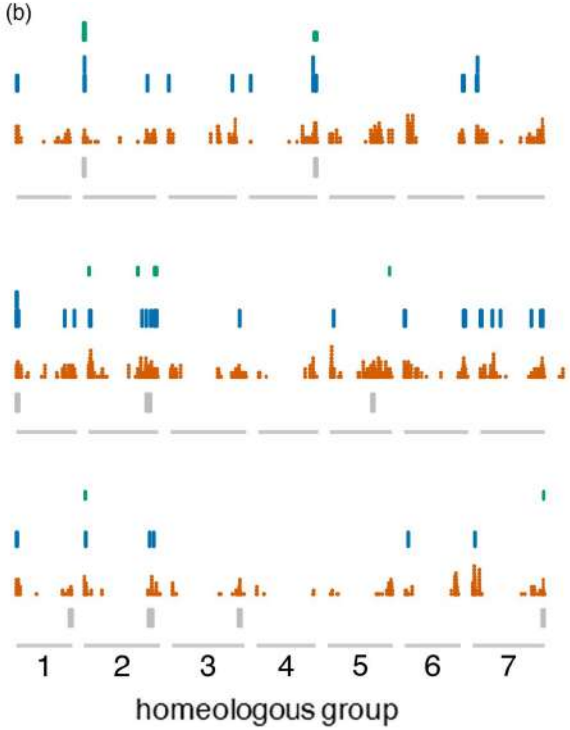
- Total number of predicted genes: 1,537,765
- Total number of clsuters (size >=2 & <20): 29002 (153734; 10%)
- Total number of clsuters (size >=20): 10298 (1,318,462; 85.74%)
- Singletons: 65569 (4.26%)

PanNLRome in hexaploid wheat

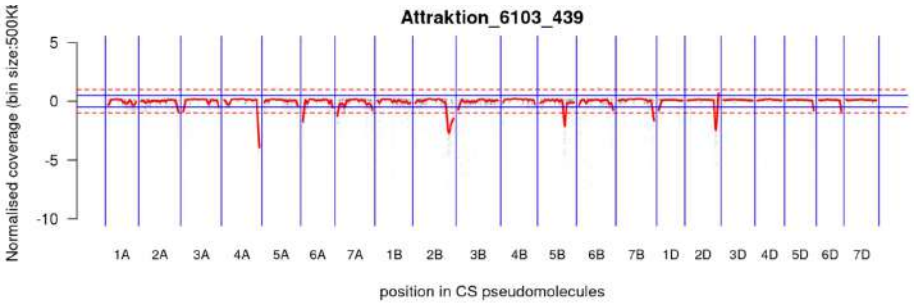




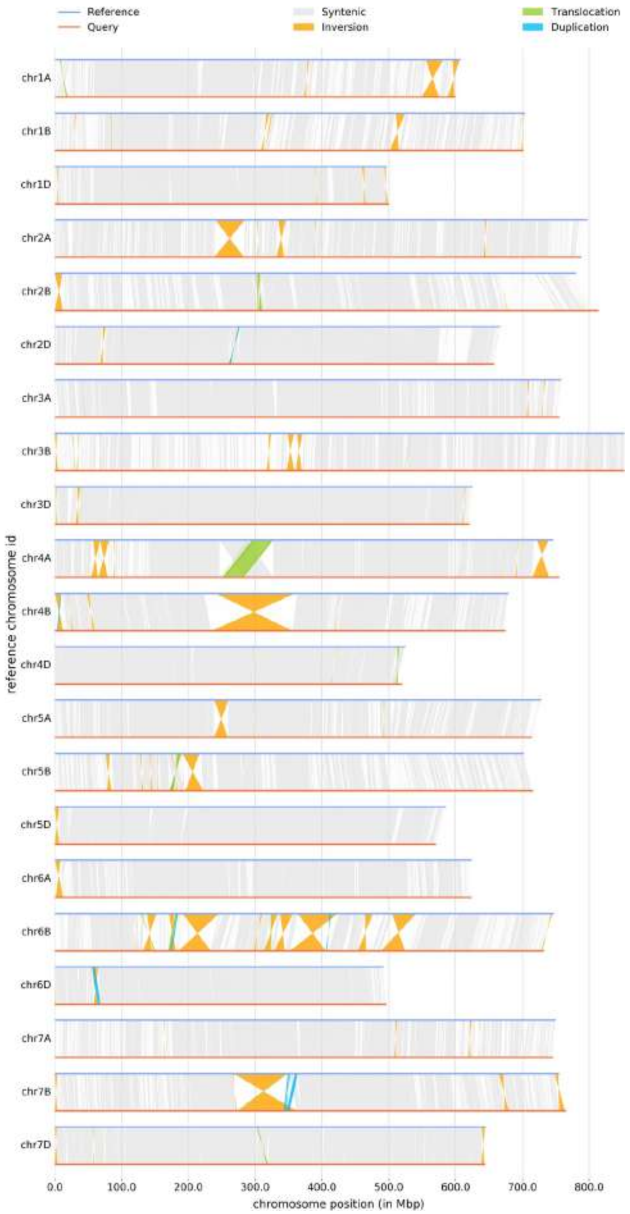
elite-specific (OGs, green),
PGR-specific OGs (blue),
elite lines and PGRs (orange)



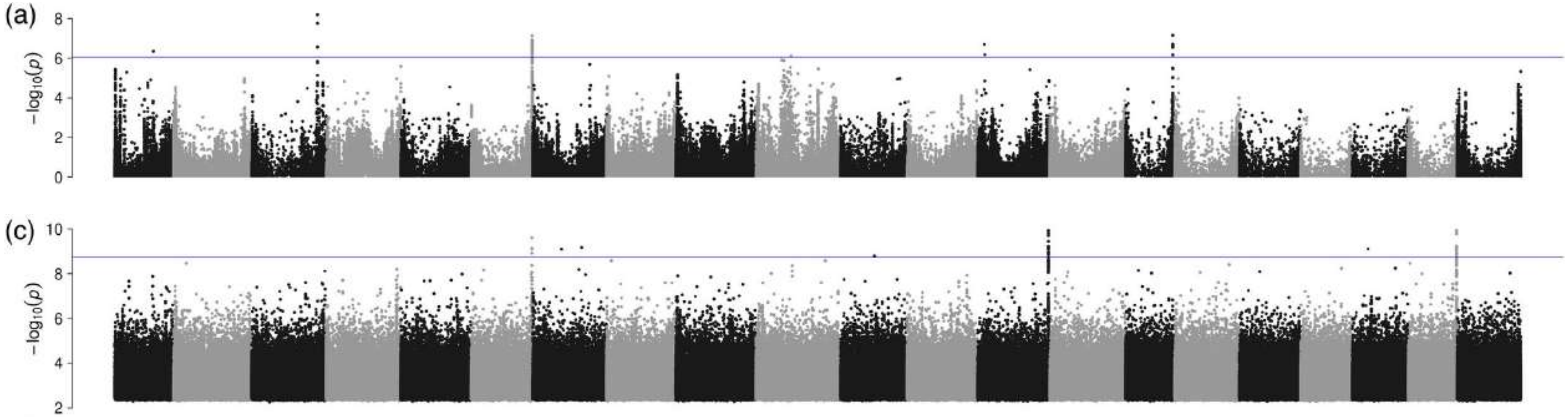
Genome Assembly of “Attraktion”



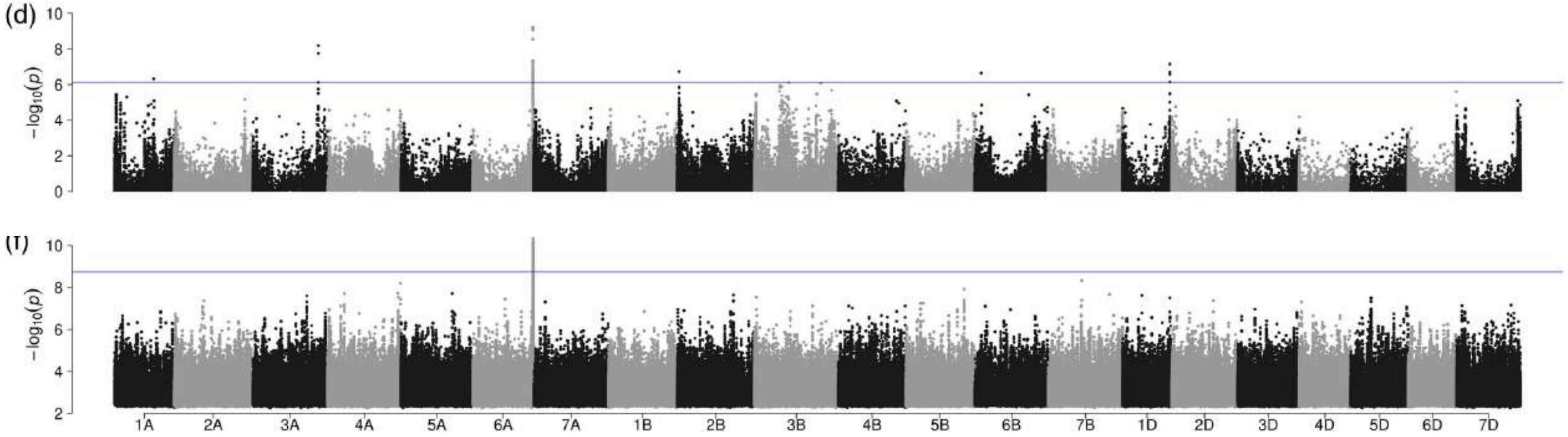
Assembly size	14.7 Gb
Number of contigs	4953
Contig N50	17.3 Mb
Contig N90	4.1 Mb
Pseudomolecule size	14.3 Gb
Number of contigs in pseudomolecules	1553
Complete BUSCOs	1584 (98.2%)



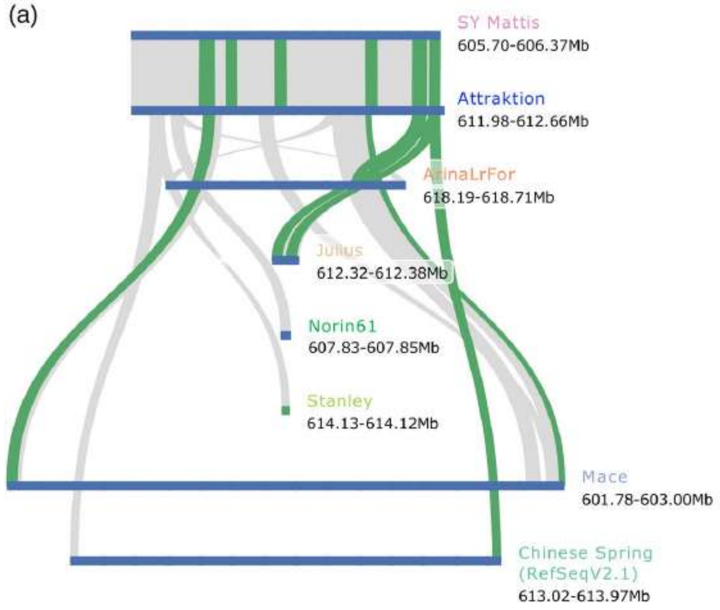
GWAS: YR

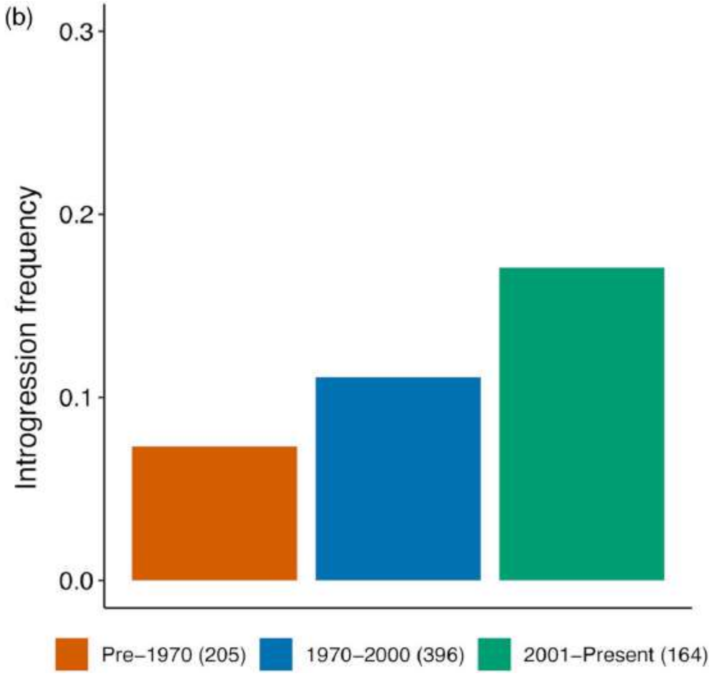


RefSeq v1.0

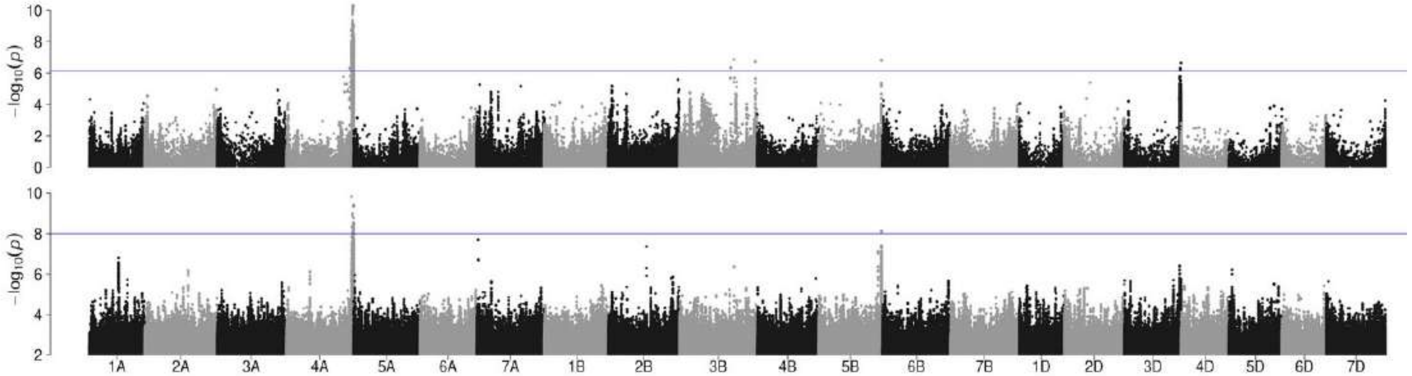


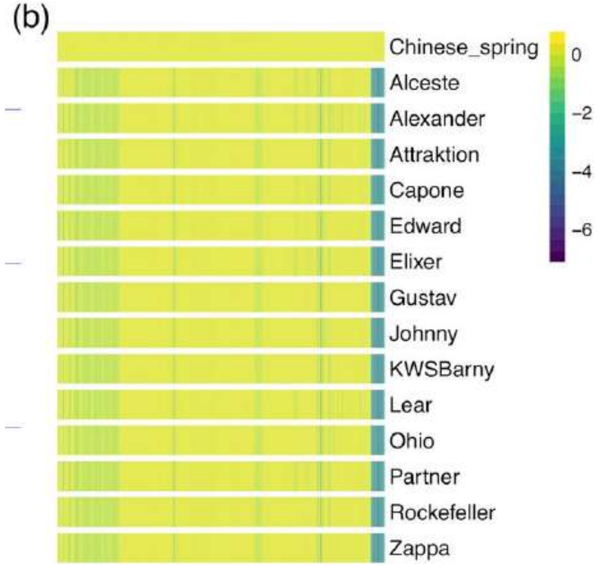
Attraktion

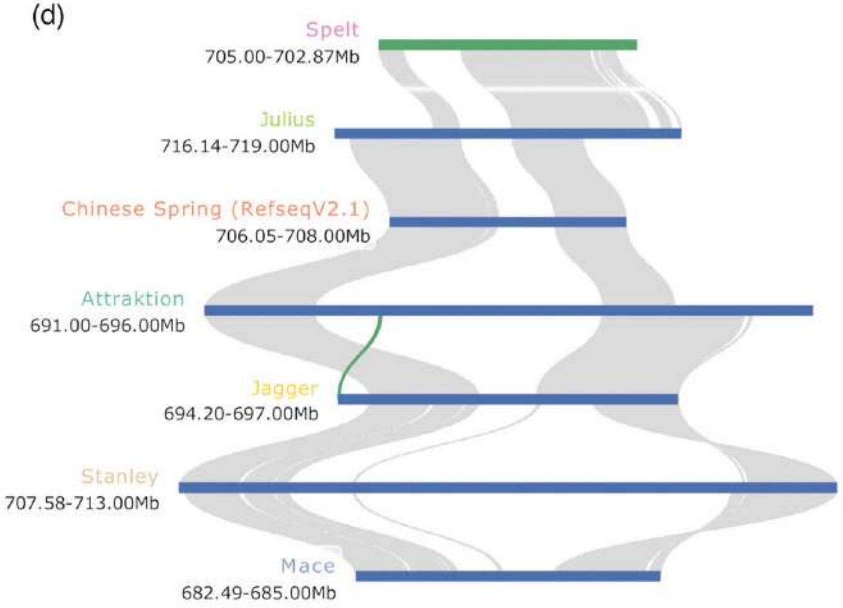


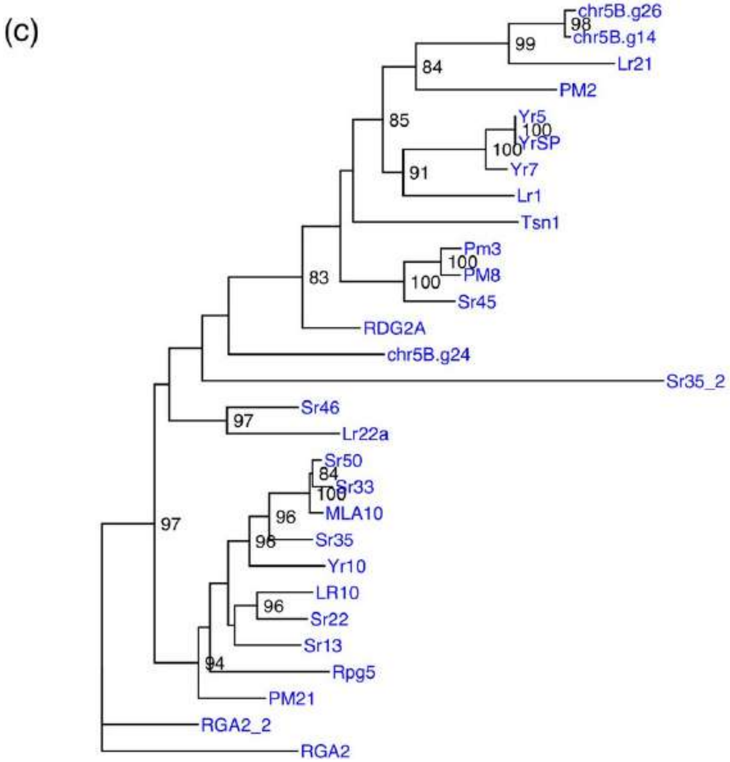


GWAS: LR









AARHUS
UNIVERSITY
DEPARTMENT OF AGROECOLOGY

Summary

- ✓ Activated EU largest winter wheat collection.
- ✓ Compared diversity between Genebanks.
- ✓ Identified novel introgressions and studied their involvement in wheat breeding
- ✓ RenSeq: attempted panNLRome construction
- ✓ Identified novel sources of resistance against YR and LR diseases

Thank You