



**8TH INTERNATIONAL FOOD LEGUME
RESEARCH CONFERENCE AND
5TH AUSTRALIAN PULSE CONFERENCE**
15 - 19 SEPTEMBER 2025 • PAN PACIFIC PERTH

Leveraging pangenome graphs to accelerate
genetic gain in breeding programs

Kumar Gaurav, John Innes Centre, UK

Types of genetic variation in the genome



Reference 1

ATGTTGTAGATCGTCGTAGCTGTACGATAGCGGGTTTCCCAGATCAGATGA

ATGTTGTAGATCGTCGTAGCTGTACGATAGCGGGTTTCCCAGATCAGATGA

SNP

ATGTTGTAGATCGTCGTAGCTGGCGCTACGATAGCGGGTTTCCCAGATCAGATGA

ATGTTGTAGATCGTCGTAGCTGTACGATAGCGGGTTTCCCAGATCAGAT

Small Indel



Reference 2

ATGTTGTAGATCGTCGTAGCTGTACGATAGTAGCTACGGGGTTTCCCAGATCAGATGA

ATGTTGTAGATCGTCGTAGCCGGGGTTTCCCAGATCAGATGA

Large Indel

ATGTTGTAGATCGTCGTAGCTGTACGATAGCGGGTTTCCCAGATCAGATGA

ATGTTGTAGATCGGCTGTACGATAGCGGGTCGTACAGATCAGATGA

Translocation



Challenges to study structural variations

- No alignment of short reads

```

ATGTTGTAGATCGTCGTAGCTGTACGATAGC-----TTTCCAGATCAGATGAAAGCCGTATGGGTCAAA
TGTTG AGITCG GTACCTG CCATAG          TCGCAGA CAGTTGATA CGTATGGG
ATGTTG GITCGT TACCTGT CATAGC          GCAGATC GATAGCCG GGGTCAAA
      GTAGTTC CGTACCT TACCATAG        TCGCAGA AGTTGATA CGTATGGGT
TGTTGT TTCGTCG CCTGTACC              GCAGATC ATAGCCGTA GGTCAAA
    
```

INDEL

- Tagging SNPs via recombination

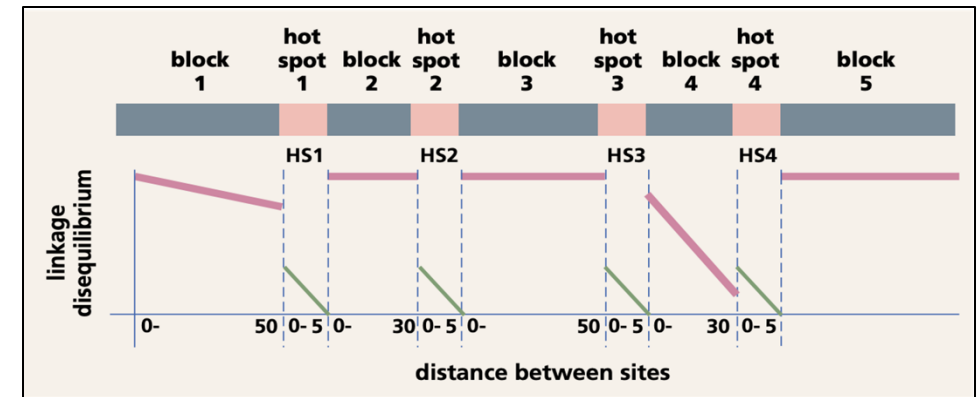


Measure of "linkage disequilibrium":
$$r^2 = \frac{(p_{ab} - p_a p_b)^2}{p_a(1 - p_a)p_b(1 - p_b)}$$

p_a : frequency of allele a

p_b : frequency of allele b

p_{ab} : frequency of co-occurrence of alleles a and b



Goldstein 2001 Nature Genetics

k-mers capture variation without performing alignment

Capturing variation using *k*-mer size 5

⇒ SNP variation

GTACG
TGTAC
CTGTA
GCTGT
AGCTG
ATGTTGTAGATCGTCGTAGCTGTACGATAGCGGGTTTCCAGATCAGATGA
ATGTTGTAGATCGTCGTAGCTGTACGATAGCGGGTTTCCAGATCAGATGA
AGCTC
GCTCT
CTCTA
TCTAC
CTACG

⇒ INDEL variation

CGGTTT
TCGGTT
GTCGGT
AGTCGG
TAGTCG
ATAGTC
GATAGT
CGATAG
ACGATA
TACGAT
GTACGA
TGTACG
CTGTAC
GCTGTA
AGCTG
TAGCT
ATGTTGTAGATCGTCGTAGCTGTACGATAGTCGGGGTTTCCAGATCAGATGA
ATGTTGTAGATCGTCGTAGCGGGTTTCCAGATCAGATGA
TAGCG
AGCGG
GCGGG
CGGGT

Application of *k*-mer GWAS to understand genetic basis of disease resistance in crops

nature plants

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature plants](#) > [brief communications](#) > [article](#)

Brief Communication | [Open Access](#) | [Published: 16 February 2023](#)

A wheat kinase and immune receptor form host-specificity barriers against the blast fungus

[Sanu Arora](#), [Andrew Steed](#), [Rachel Goddard](#), [Kumar Gaurav](#), [Tom O'Hara](#), [Adam Schoen](#), [Nidhi Rawat](#), [Ahmed F. Elkot](#), [Andrey V. Korolev](#), [Catherine Chinoy](#), [Martha H. Nicholson](#), [Soichiro Asuke](#), [Rea Antoniou-Kourounioti](#), [Burkhard Steuernagel](#), [Guotai Yu](#), [Rajani Awal](#), [Macarena Forner-Martínez](#), [Luzie Wingen](#), [Erin Baggs](#), [Jonathan Clarke](#), [Diane G. O. Saunders](#), [Ksenia V. Krasileva](#), [Yukio Tosa](#), [Jonathan D. G. Jones](#), ... [Paul Nicholson](#) ✉

+ Show authors

[Nature Plants](#) **9**, 385–392 (2023) | [Cite this article](#)

6099 Accesses | **159** Altmetric | [Metrics](#)

nature biotechnology

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature biotechnology](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 01 November 2021](#)

Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement

[Kumar Gaurav](#), [Sanu Arora](#), [Paula Silva](#), [Javier Sánchez-Martín](#), [Richard Horsnell](#), [Liangliang Gao](#), [Gurcharn S. Brar](#), [Victoria Widrig](#), [W. John Raupp](#), [Narinder Singh](#), [Shuangye Wu](#), [Sandip M. Kale](#), [Catherine Chinoy](#), [Paul Nicholson](#), [Jesús Quiroz-Chávez](#), [James Simmonds](#), [Sadiye Hayta](#), [Mark A. Smedley](#), [Wendy Harwood](#), [Suzannah Pearce](#), [David Gilbert](#), [Ngonidzashe Kangara](#), [Catherine Gardener](#), [Macarena Forner-Martínez](#), ... [Brande B. H. Wulff](#) ✉

+ Show authors

[Nature Biotechnology](#) **40**, 422–431 (2022) | [Cite this article](#)

23k Accesses | **43** Citations | **242** Altmetric | [Metrics](#)

nature biotechnology

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature biotechnology](#) > [letters](#) > [article](#)

Letter | [Published: 04 February 2019](#)

Resistance gene cloning from a wild crop relative by sequence capture and association genetics

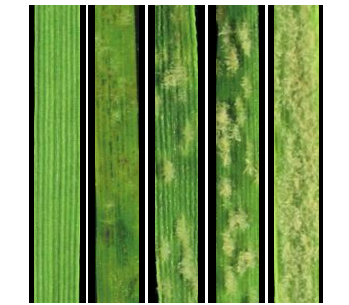
[Sanu Arora](#), [Burkhard Steuernagel](#), [Kumar Gaurav](#), [Sutha Chandramohan](#), [Yunming Long](#), [Qadi Matny](#), [Ryan Johnson](#), [Jacob Enk](#), [Sambasivam Periyannan](#), [Narinder Singh](#), [M. Asyraf Md Hatta](#), [Naveenkumar Athiyannan](#), [Jitender Cheema](#), [Guotai Yu](#), [Ngonidzashe Kangara](#), [Sreya Ghosh](#), [Les J. Szabo](#), [Jesse Poland](#), [Harbans Bariana](#), [Jonathan D. G. Jones](#), [Alison R. Bentley](#), [Mick Ayliffe](#), [Eric Olson](#), [Steven S. Xu](#), ... [Brande B. H. Wulff](#) ✉

+ Show authors

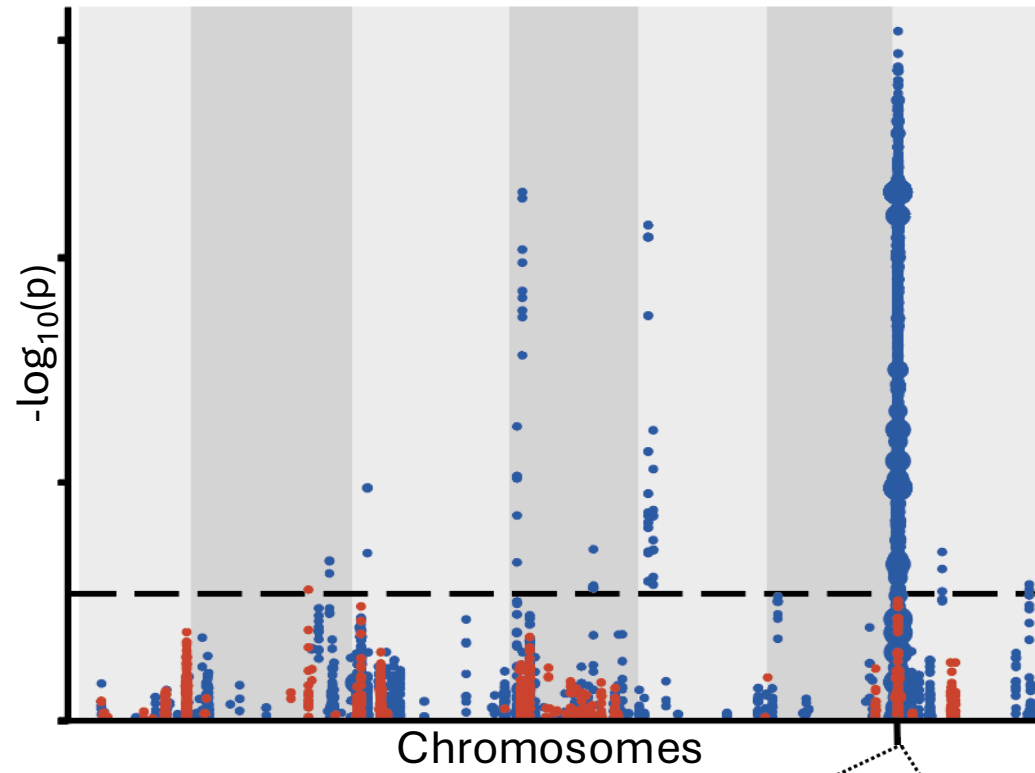
[Nature Biotechnology](#) **37**, 139–143 (2019) | [Cite this article](#)

17k Accesses | **177** Citations | **271** Altmetric | [Metrics](#)

Case study: *k*-mers captured large insertion associated with disease resistance



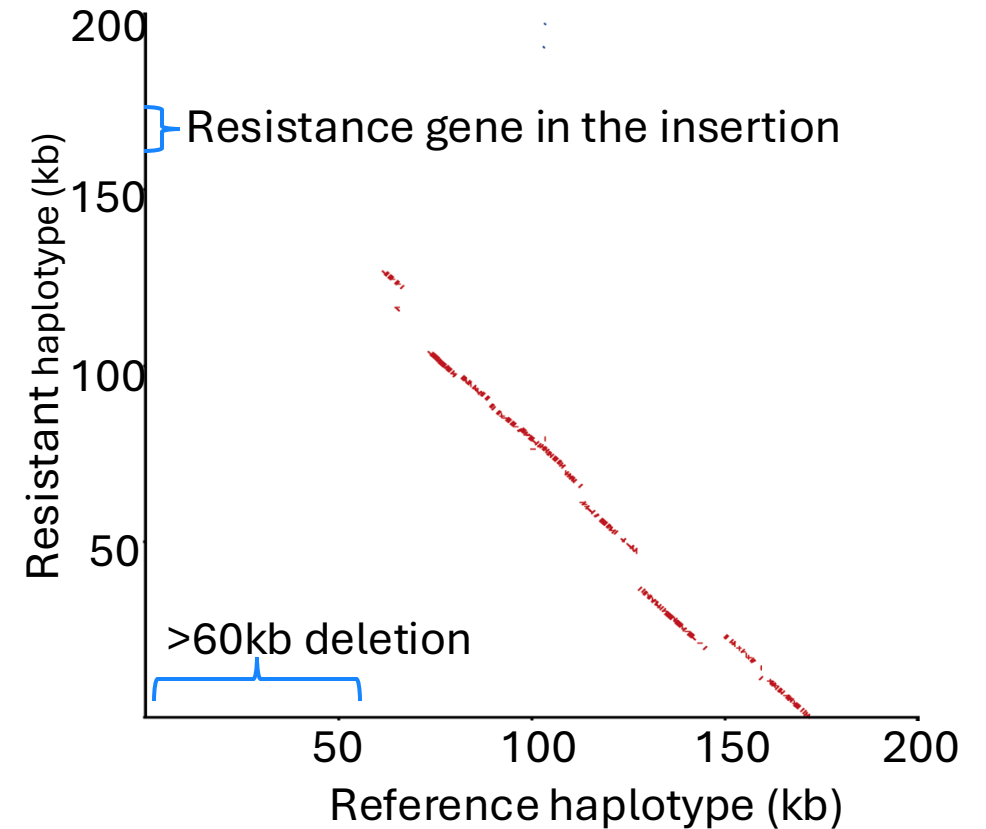
Powdery Mildew



Chromosomes

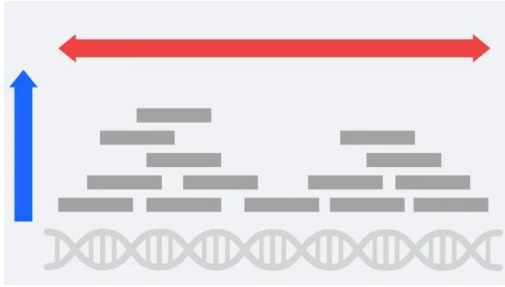
7D: 4.76-5.08 Mb

Resistance gene in the insertion



Gaurav *et al.* **Nature Biotech.** 2022

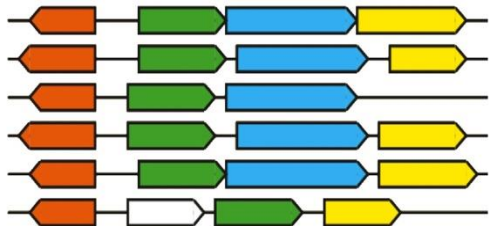
Drawbacks of k -mer based methods



Requires relatively deep sequencing ($>10\times$) to cover k -mer space



Huge number of k -mer variants add computational burden



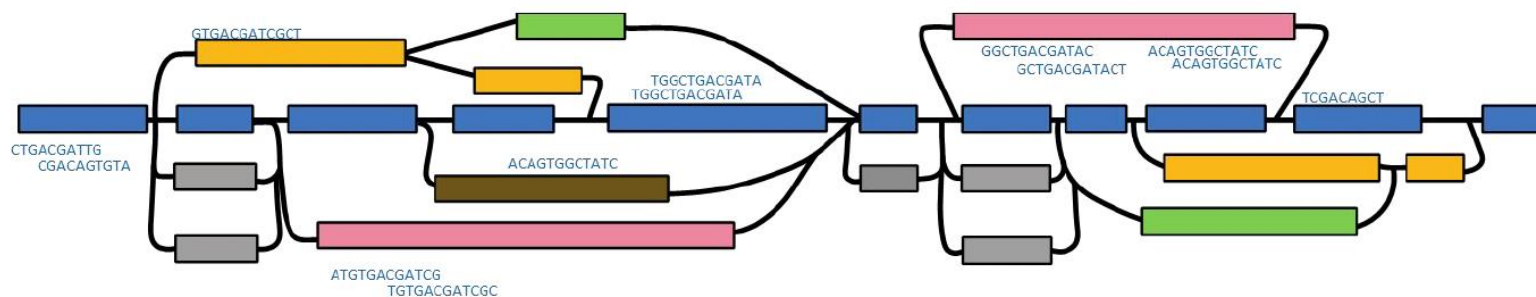
Genomic context of k -mers is not known *a priori*, making it unsuitable for population genetics analysis, genomic prediction, advanced GWAS methods



Graph-based methods for pan-genomic analysis

- Reference 1
- Reference 2
- Reference 3
- Reference 4
- Reference 5
- Reference 6

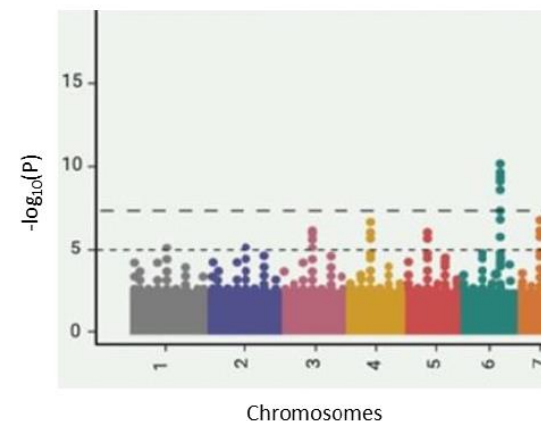
Step 1: Alignment of short reads to pan-genome graph



Step 2: Path inference

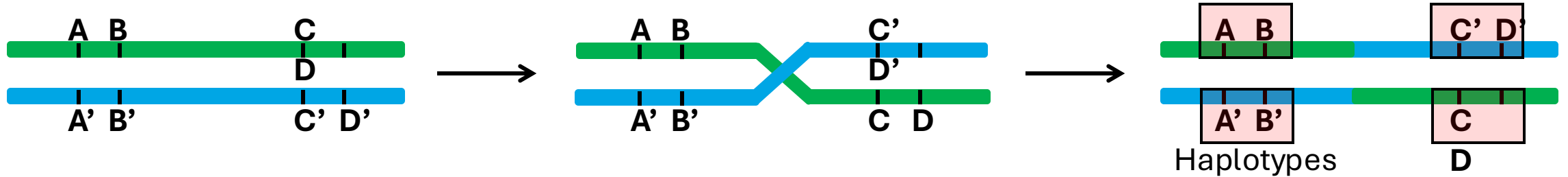
	Graph nodes					
S1	H ₁	H ₂	.	.	.	H ₁
S2	H ₀	H ₃	.			H ₄
S3	H ₂	H ₀		.		H ₃
.	.	.		.		.
.	.	.			.	.
.	.	.			.	.
Sn	H ₁	H ₂	.	.	.	H ₀

Step 3: GWAS

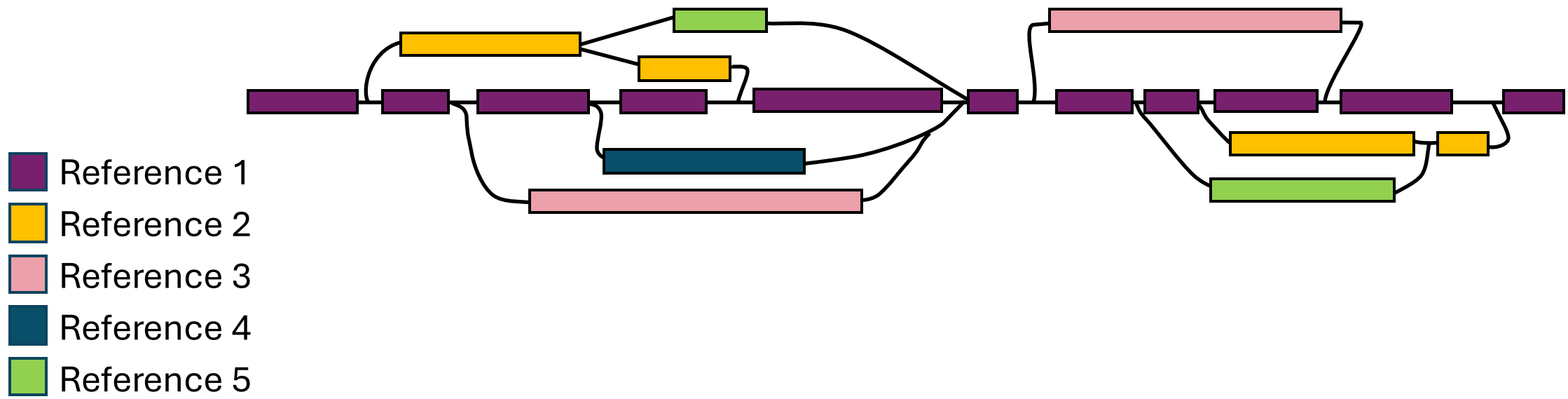


Pan-genome graph captures haplotypic variation

Haplotype → Genomic block inherited as a unit



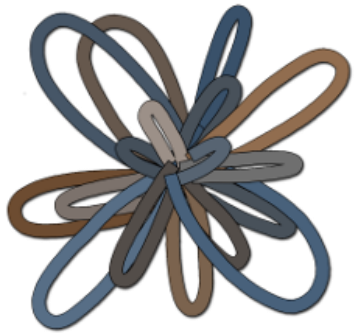
Pan-genome graph captures haplotypic variation



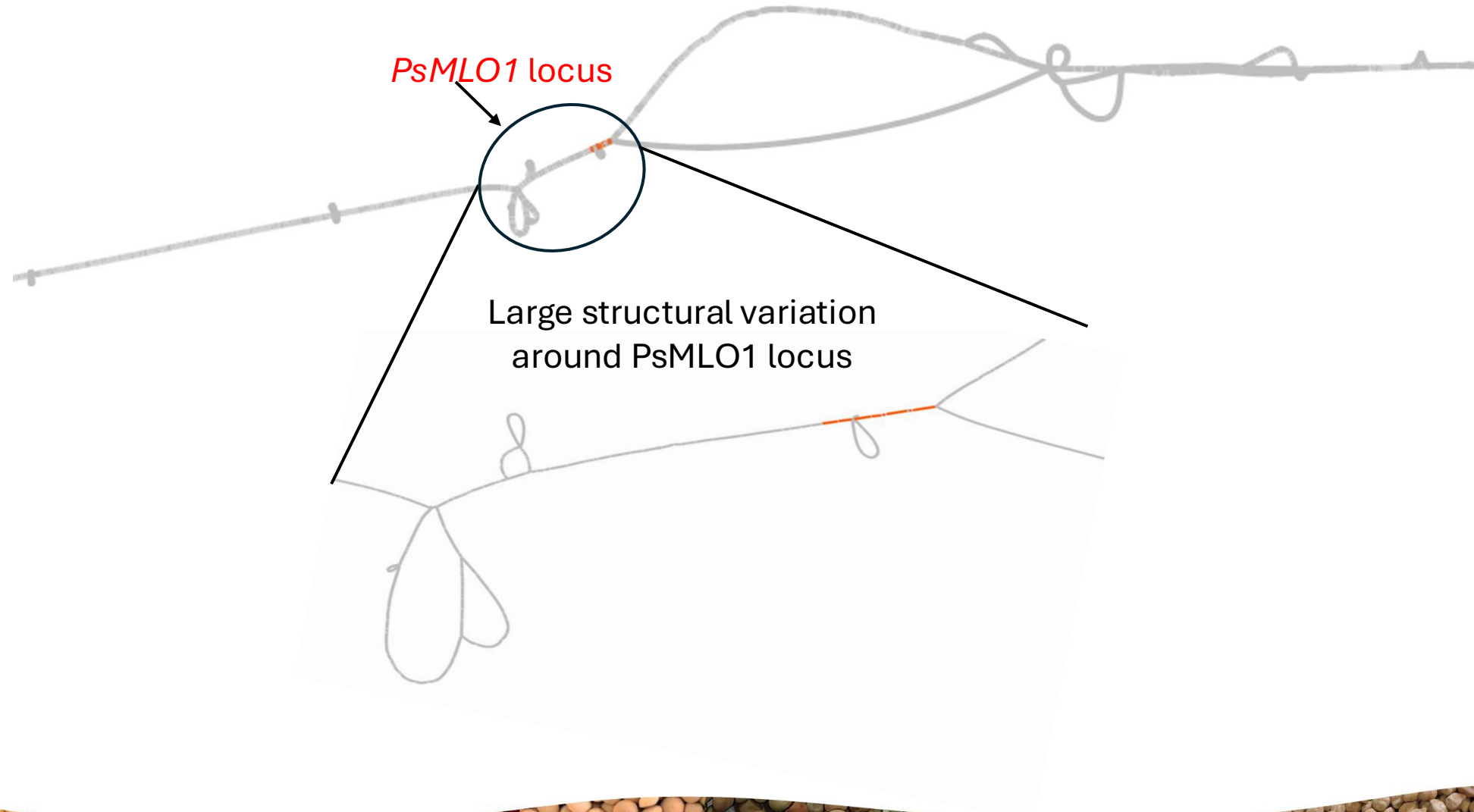
Pan-genome graph captures haplotypic variation: Example in *Pisum* spp.



Graph construction
using PGGB

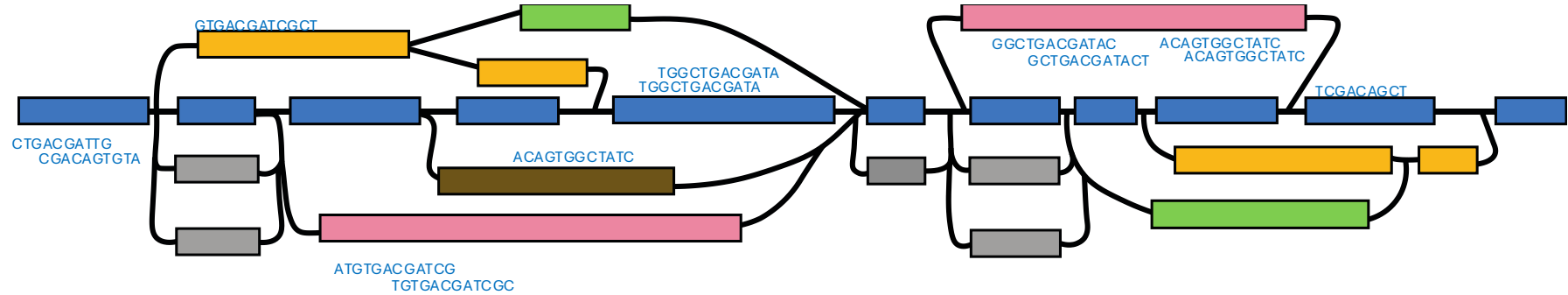


Graph visualization
using Bandage



Node inference based on mapping short reads to graph

- Reference 1
- Reference 2
- Reference 3
- Reference 4
- Reference 5
- Reference 6



Since we have captured most of the species haplotypic variation in the pangenome graph, we can map short reads to the graph and tag each haplotypic variant

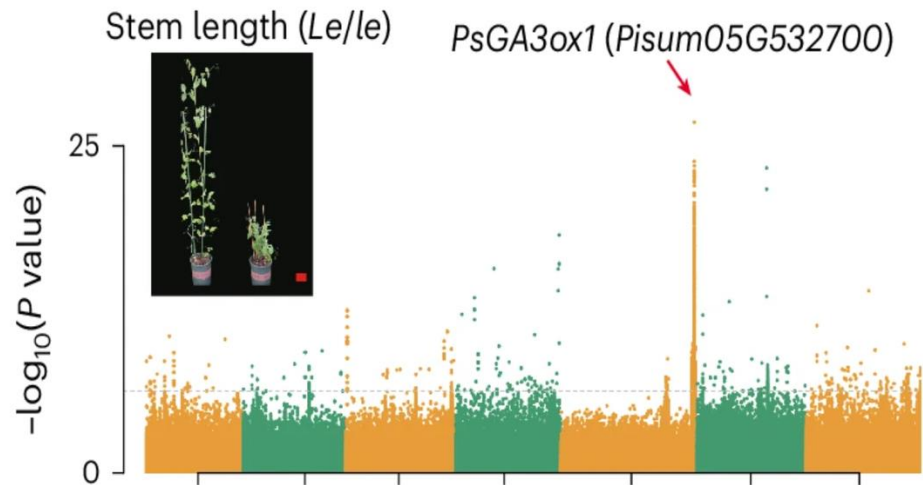
Pangenome GWAS for Mendel's traits – Stem length

nature genetics

Article

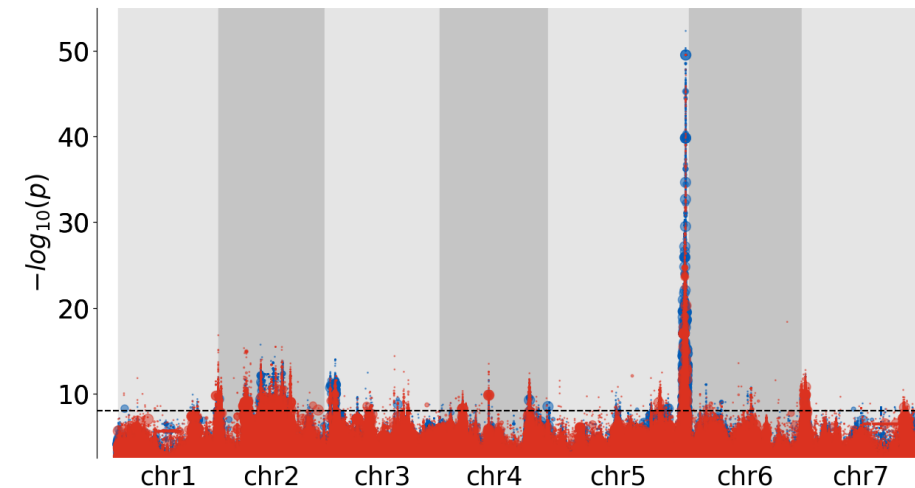
<https://doi.org/10.1038/s41588-024-01867-8>

Reference genome sequence and population genomic analysis of peas provide insights into the genetic basis of Mendelian and other agronomic traits



WGS data of 230 lines with 10x sequencing depth

Pangenome graph-based GWAS



Pangenome GWAS for Mendel's traits – Flower color

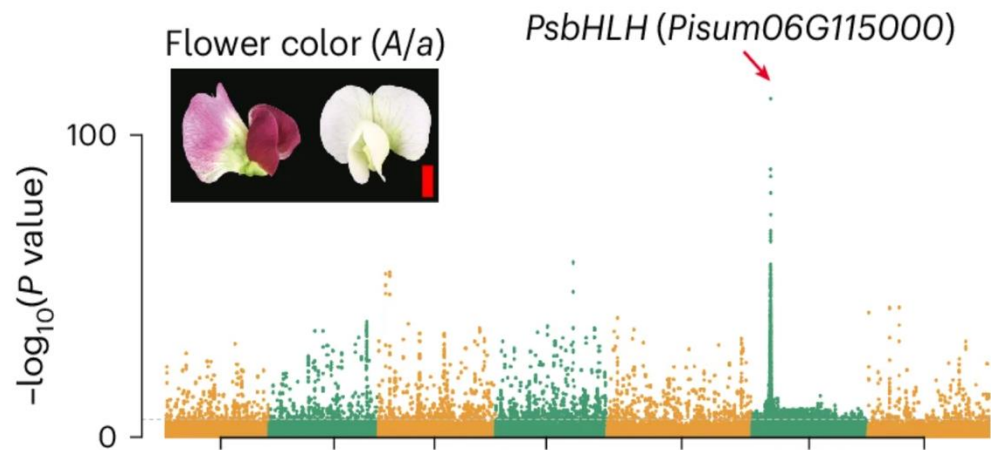
nature genetics

Article

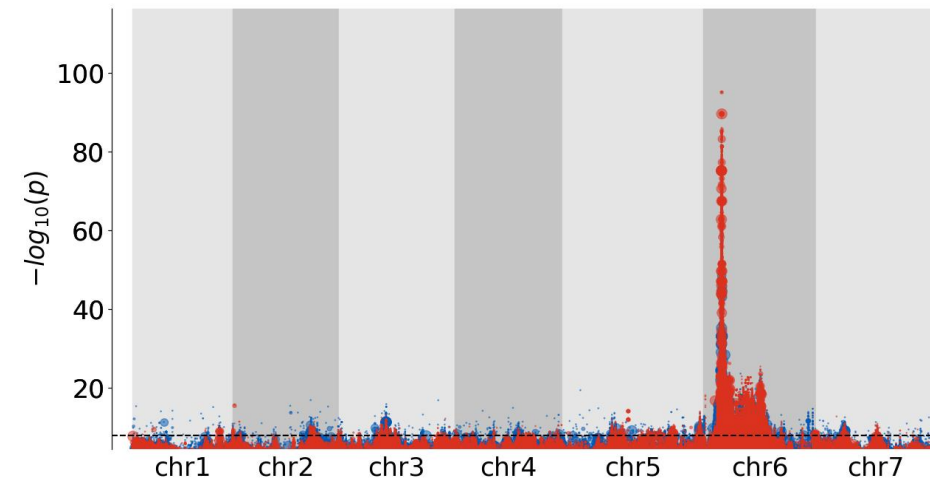
<https://doi.org/10.1038/s41588-024-01867-8>

Reference genome sequence and population genomic analysis of peas provide insights into the genetic basis of Mendelian and other agronomic traits

WGS data of 230 lines with 10x sequencing depth



Pangenome graph-based GWAS



Pea root rot complex

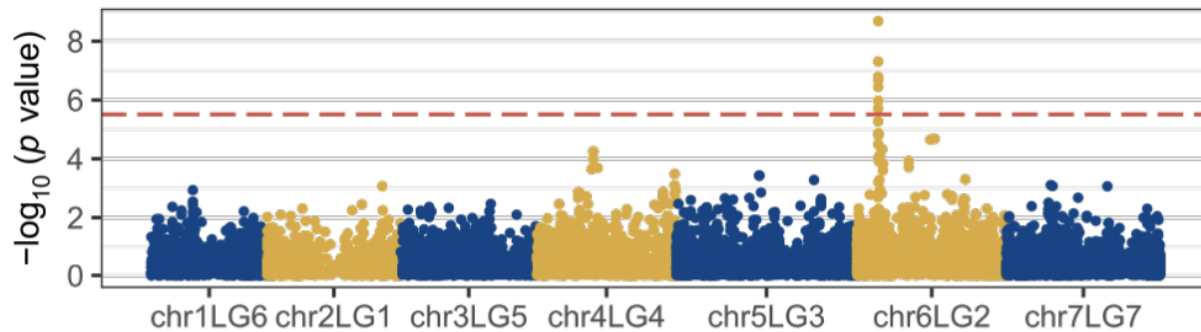
- Disease complex of fungi and oomycetes
 - *Fusarium* ssp.
 - *Aphanomyces euteiches*
 - *Rhizoctonia solani*
 - *Phoma* ssp.
- Yield losses typically 10 % (up to 70 %)
- Pathogen build-up in soil and survival for years
- Limited chemical control available
- No qualitative resistance



Pangenome GWAS for root rot: node inference from GBS reads

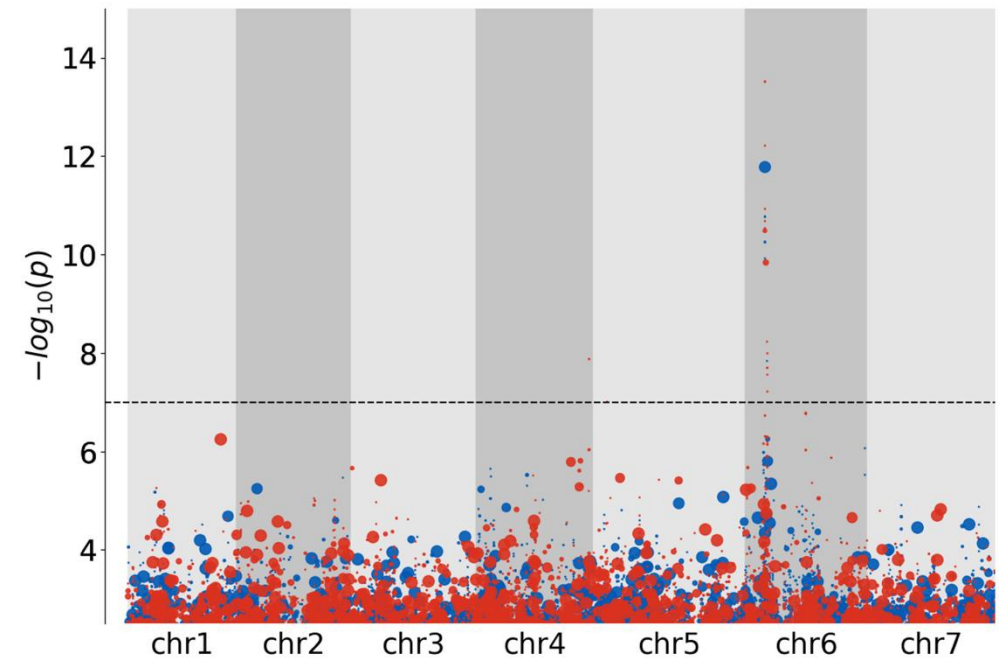
Association genetics and genomic prediction for resistance to root rot in a diverse collection of *Pisum sativum* L.

Daniel Ariza-Suarez¹, Lukas Wille¹, Pierre Hohmann², Valentin Gfeller², Michael Schneider², Matthew Horton², Monika Messmer², and Bruno Studer¹



- GBS data of 250 lines
- GBS reads are mapped to almost 1 million nodes compared to 20000 markers using usual protocol

Pangenome graph-based GWAS



JIC *Pisum* collection

[nature](#) > [articles](#) > article

Article | [Open access](#) | Published: 23 April 2025

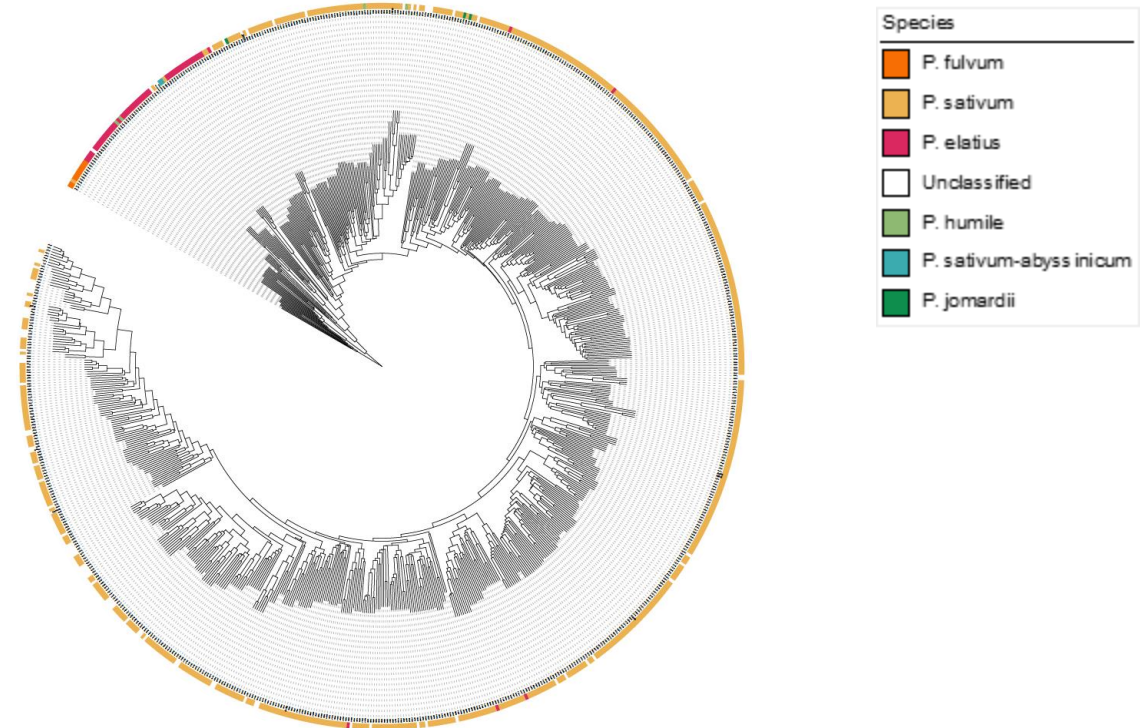
Genomic and genetic insights into Mendel's pea genes

- 700 *Pisum* accessions sequenced with 20x depth
- Artificially lowered sequencing depth to 1x



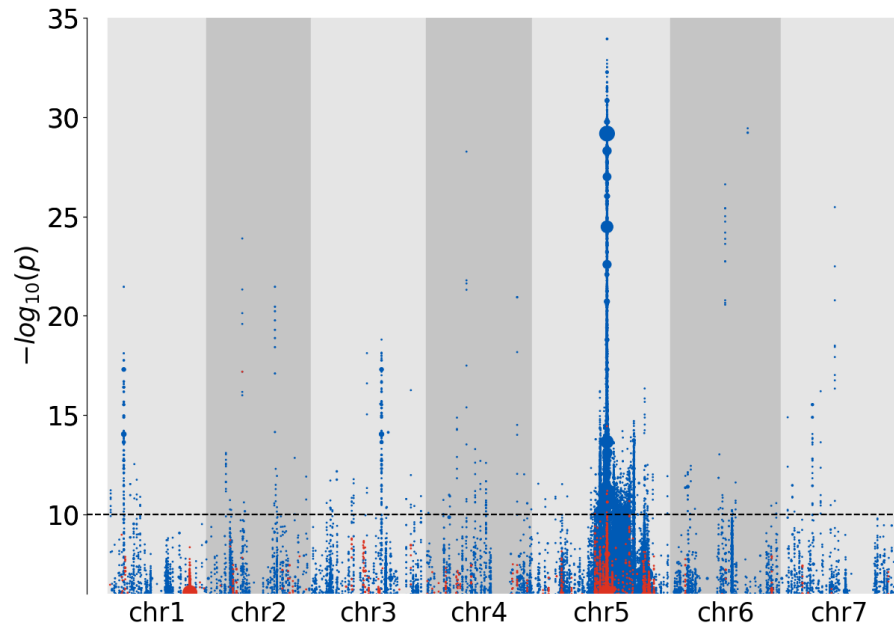
Germplasm Resources Unit

..... a national capability supported by the BBSRC at the John Innes Centre

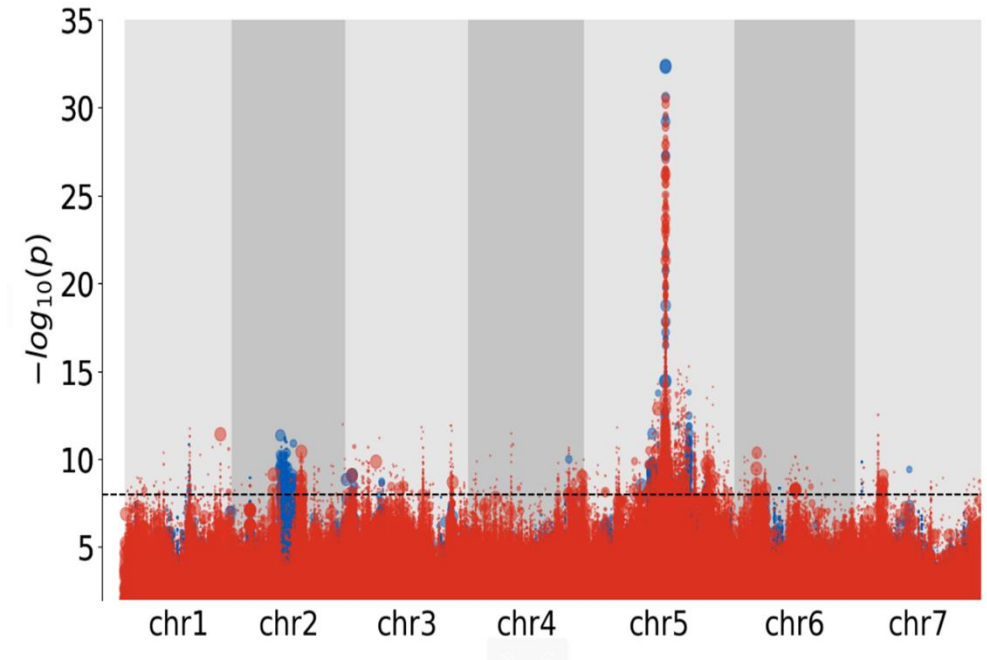


GWAS for a disease trait

***k*-mer GWAS (with 20x sequencing depth)
10 billion *k*-mer variants**



**Pangenome GWAS (~1x sequencing depth+ imputation)
100 million node variants**



Root rot revisited: Phenotyping of a RIL population



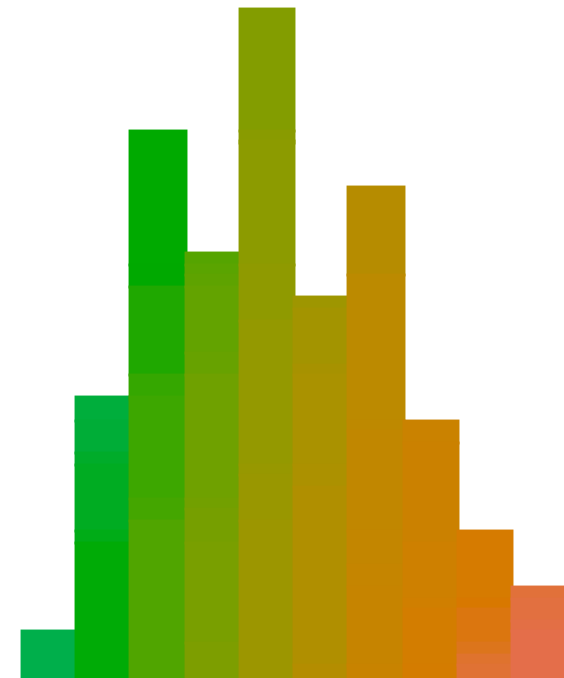
Cultivar

X



Landrace

RIL
Phenotyping



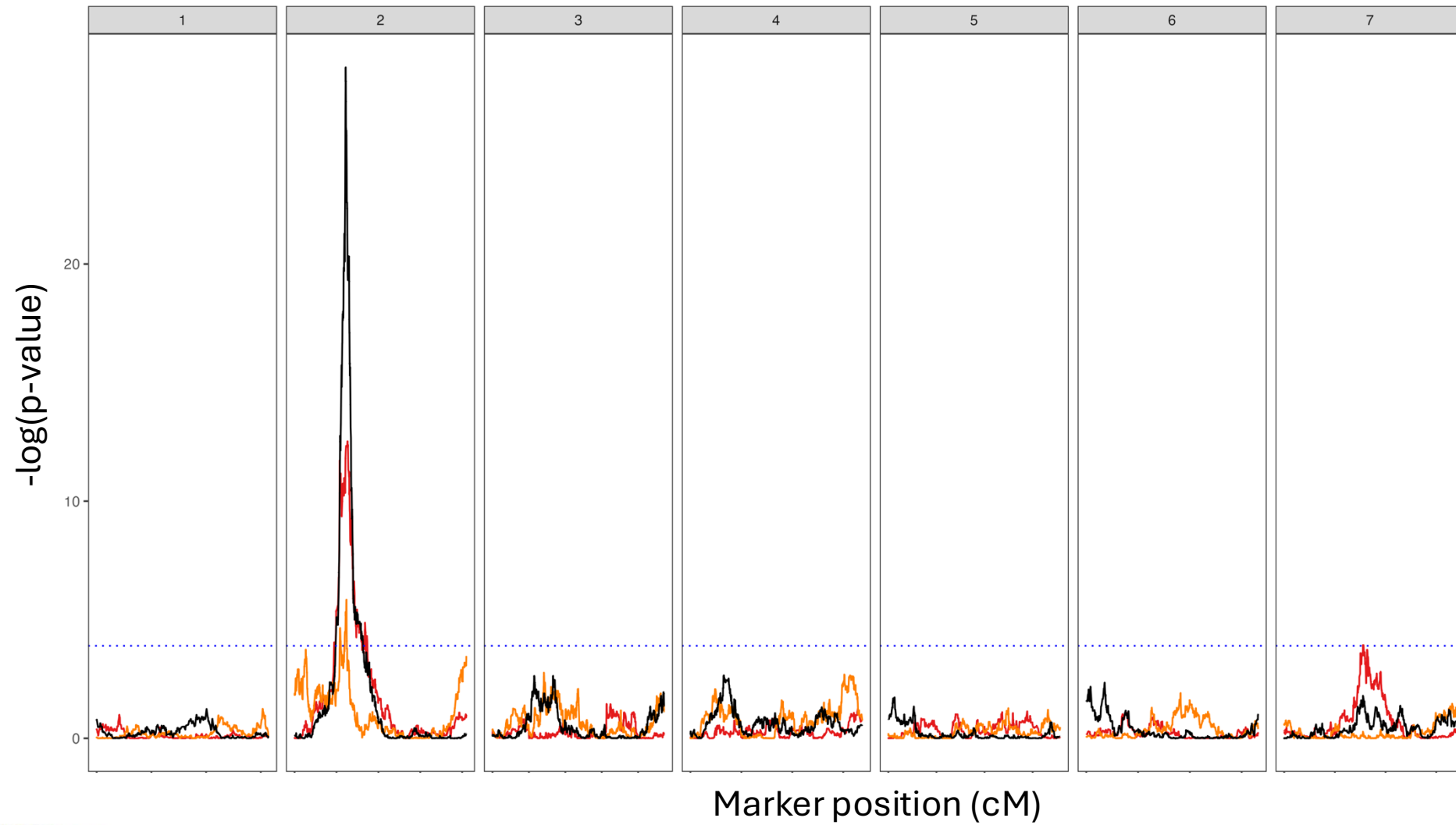
Partially resistant

Susceptible



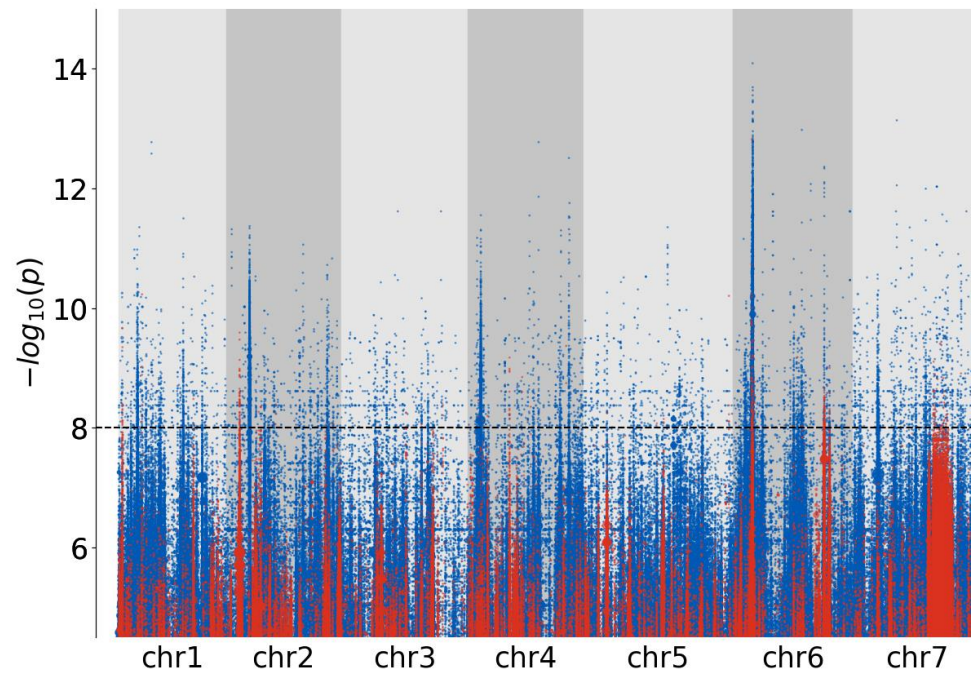
N. Trenk

Genetic mapping for root rot

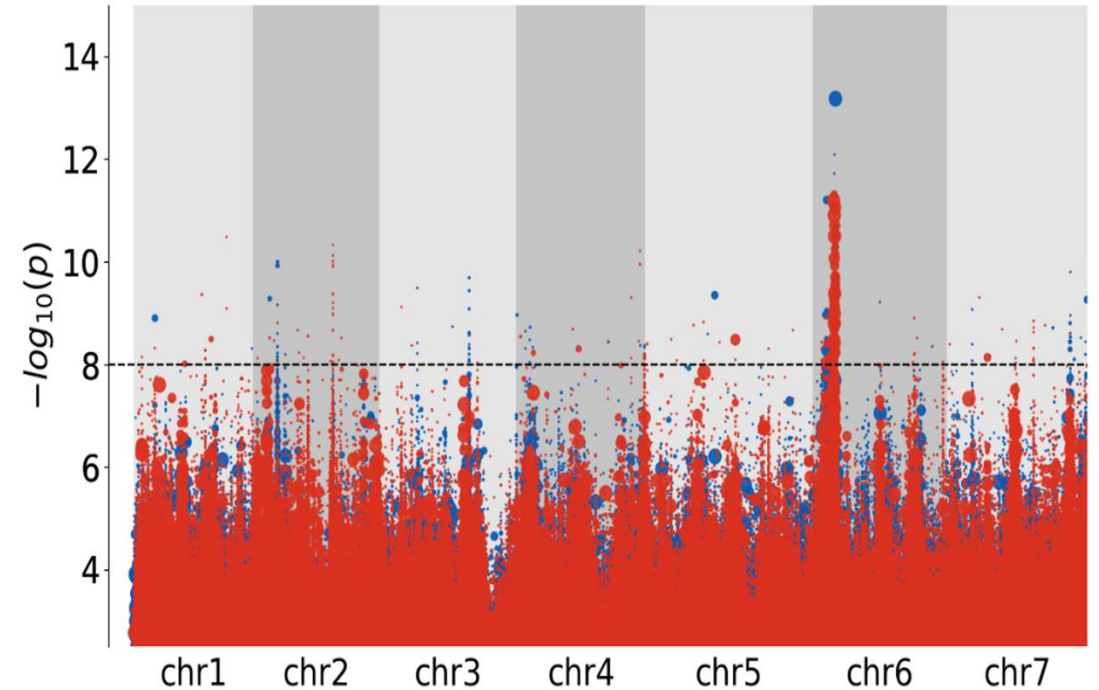


GWAS for root rot

***k*-mer GWAS (with 20x sequencing depth)**
10 billion *k*-mer variants

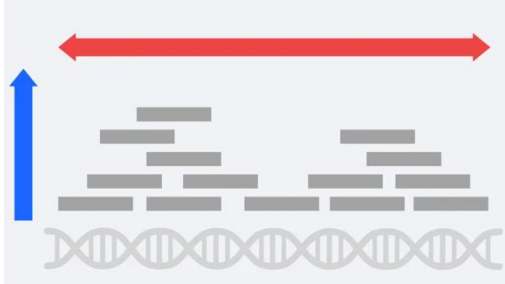


Pangenome GWAS (~1x sequencing depth+ imputation)
100 million node variants



Summary

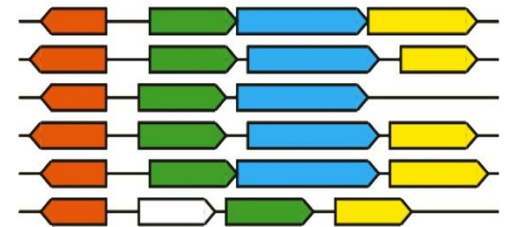
Pangenome graph provides a versatile framework for genomic analysis:



- Allows the sequencing depth to be dramatically reduced



- Huge reduction in computational burden and statistical noise



- Allows genetic analysis requiring genomic context of variants to be known, such as genomic prediction



Acknowledgements

Arora lab members (JIC)

Tony Miller (JIC)

Claire Domoney (JIC)



John Innes Centre

Unlocking Nature's Diversity



Legume Generation Partners



Legume
Generation



Pisum 60+ consortium



GenomeCanada



National Research
Council Canada



UNIVERSITY OF
CALGARY

