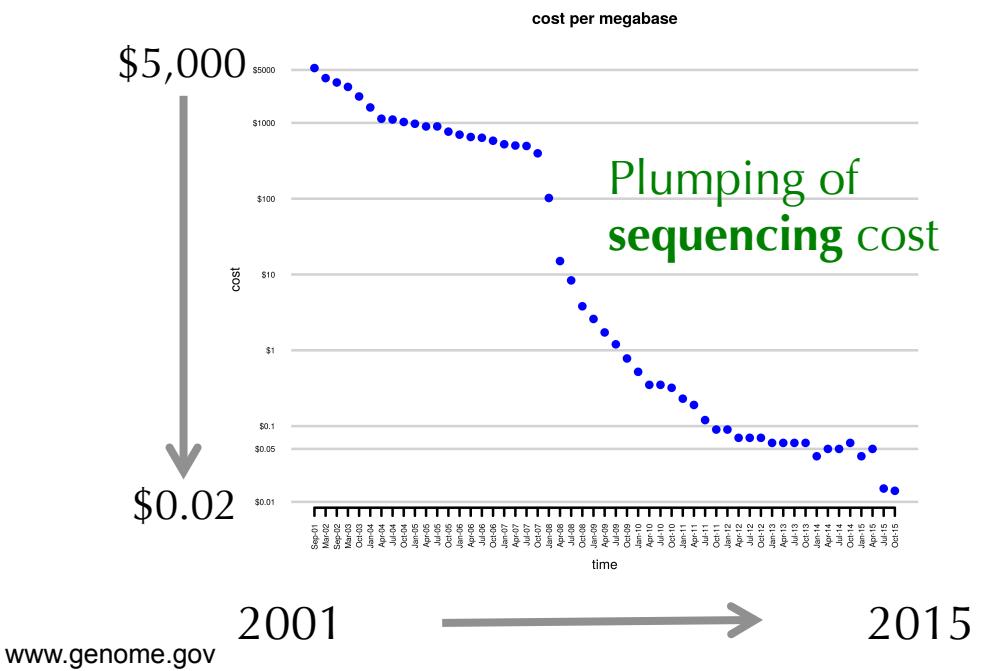


Next-Gen Sequencing (NGS) facilitates gene mapping and cloning



Sequencing data provide

1. sequence (genotype) information
- 2. read counts** (counts of each allele type)

GATCTGCGTCATAACGGAAAT
GATCTGCGTGATAACGGAAAT
GATCTGCGTCATAACGGAAAT
GATCTGCGTGATAACGGAAAT
GATCTGCGTGATAACGGAAAT
GATCTGCGTGATAACGGAAAT
GATCTGCGTGATAACGGAAAT
GATCTGCGTCATAACGGAAAT
GATCTGCGTCATAACGGAAAT
GATCTGCGTGATAACGGAAAT
GATCTGCGTCATAACGGAAAT

-C/G

Outline

Part 1. Genetic mapping via RNA sequencing

Part 2. bacterial genome assemblies

Part 3. k-mer approach to characterize genome difference

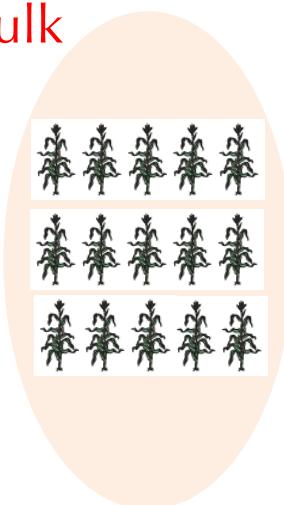
Part 4. Copy number variation in host resistance to Goss's wilt

Genetic mapping via sequencing bulks (I)

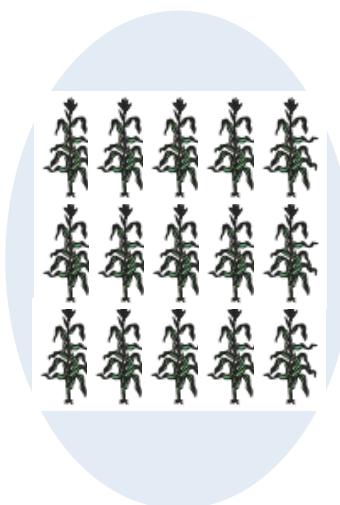
Traditional method



bulk



bulk 1



bulk 2

Maize photo from Candela *et al.* 2008, Nat Rev Genet 9:192--2030

Bulked Segregant Analysis (BSA)

Objective:

Rapidly genetically map mutant genes or QTLs

First BSA paper:

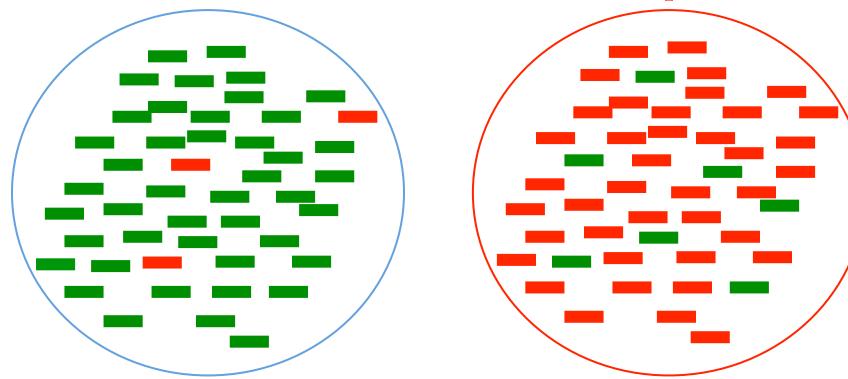
MICHELMORE, RW *et al.*, 1991, PNAS 88: 9828-9832.

Genetic mapping via sequencing bulks (II)

— tall allele 
— short allele 

a bulk of
tall plants

a bulk of
short plants



sequencing: “randomly” samples alleles from each pool

Read counts

— 20
— 1

— 2
— 18

Green/red alleles are respectively enriched in the **tall/short bulks**. Therefore the gene is somehow associated with the phenotype (tall/short)

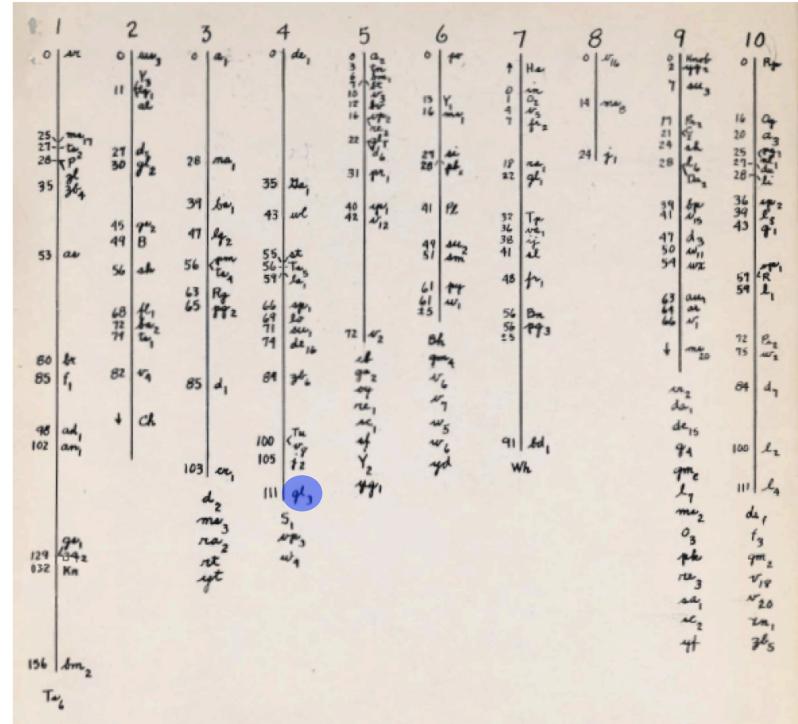
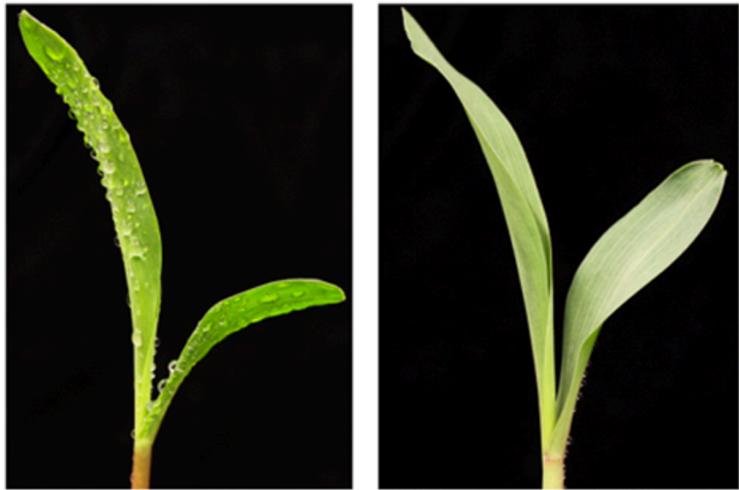
Sequencing methods for BSA

- Restriction-site associated DNA (RAD) tags sequencing (Baird et al. 2008: Fungus)
- ShoreMap (Schneeberger et al. 2009: Arabidopsis)
- BSA in yeast (Wenger et al. 2010, Ehrenreich 2010)
- NGM (Austin et al. 2011: Arabidopsis)
- MutMap (Abe et al. 2011: Rice)

All are DNA-based (DNA sequencing + BSA analysis).

Bulked Segregant RNA-Seq (BSR-Seq)

BSR-Seq to map *glossy3* (*gl3*)

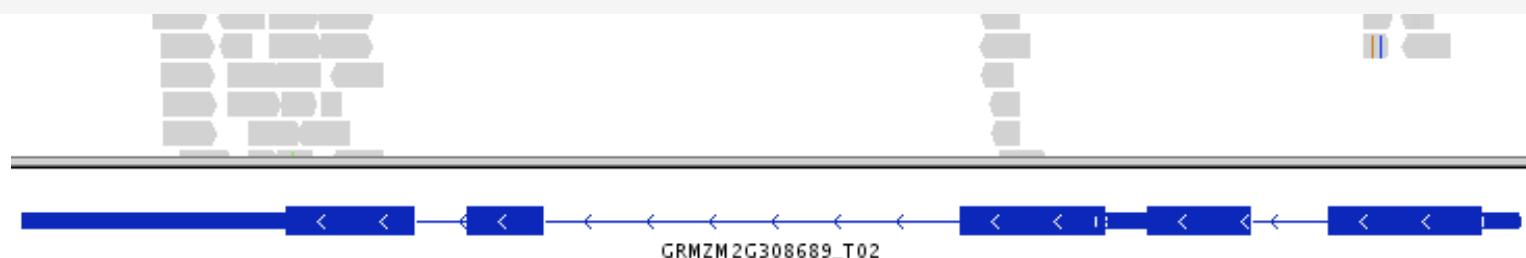


An early maize genetic map (1937)
– image from maizeGDB.org

RNA-Seq



- RNA-Seq generated ~13 millions of reads per sample
- 64,852 SNPs were identified and quantified in both the mutant pool and the non-mutant pool

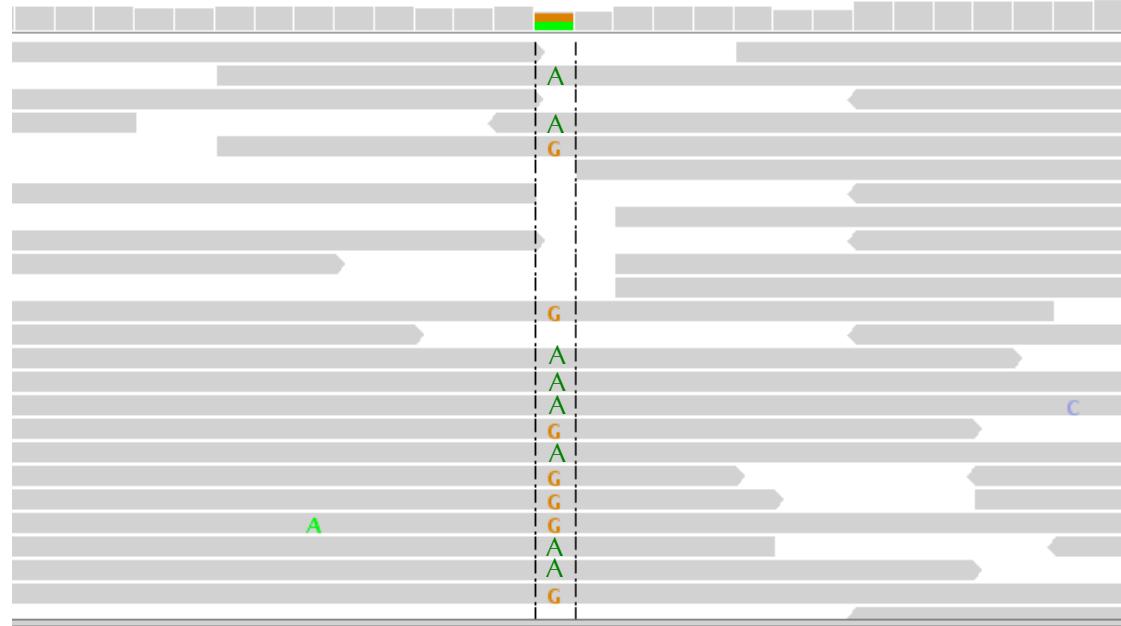


Example of a SNP **UNLINKED** to a mutant gene

Reference

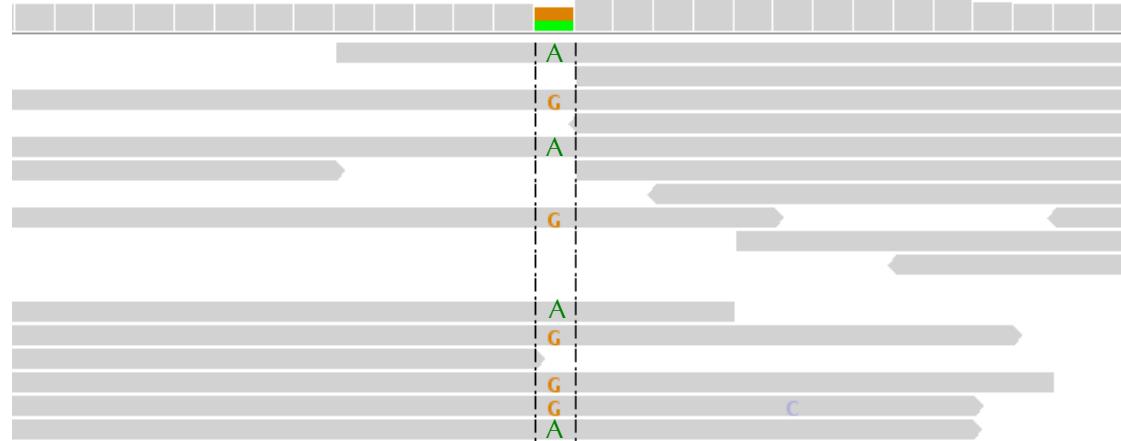
G G C T G C C G C G C A C C G C A A C C C G C T G
GRMZM2G089783_T01

Reads
from
Mutants



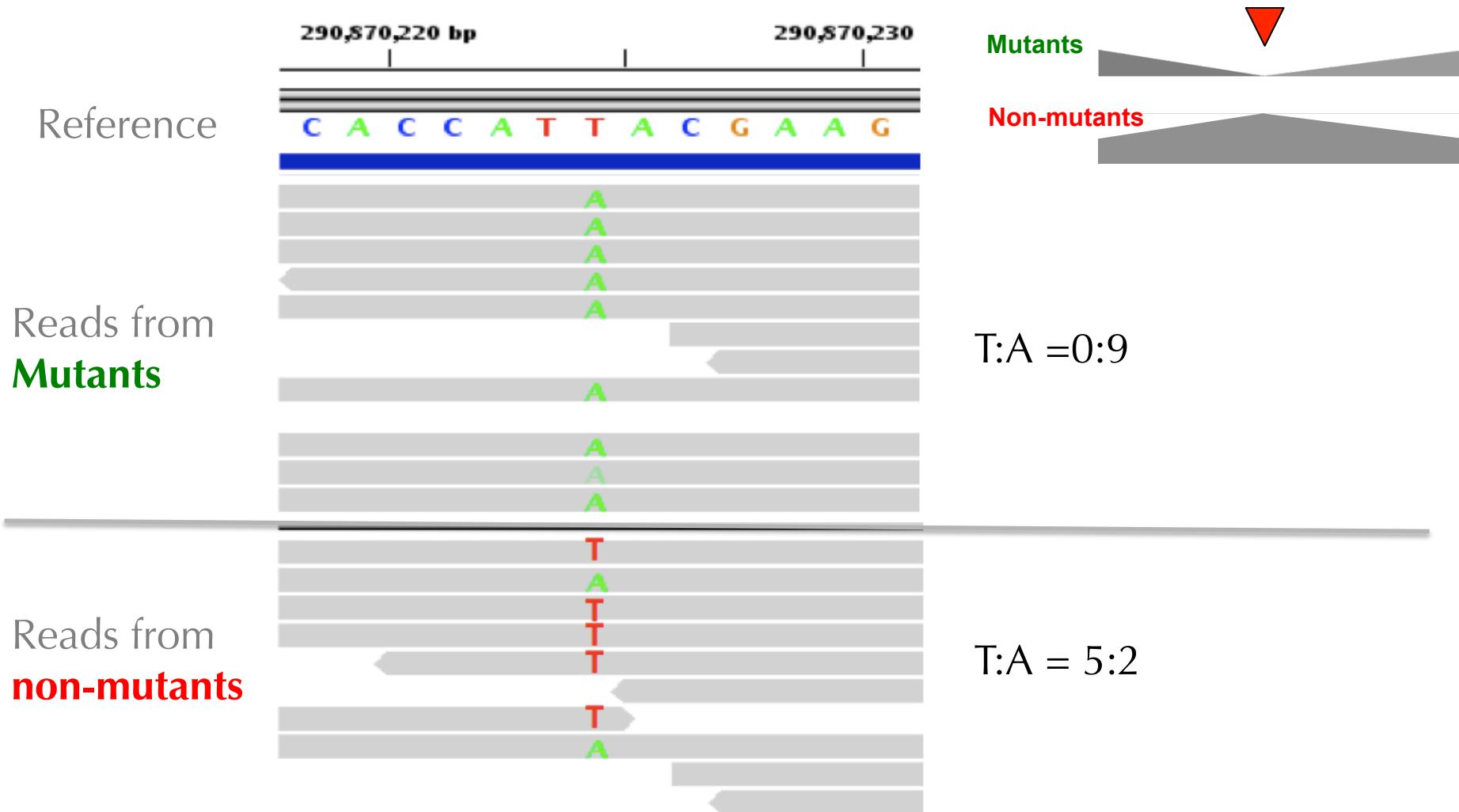
$$A:G = 8:7$$

Reads
from non-
mutants



$$A:G = 4:5$$

Example of a SNP Completely LINKED to a mutant gene

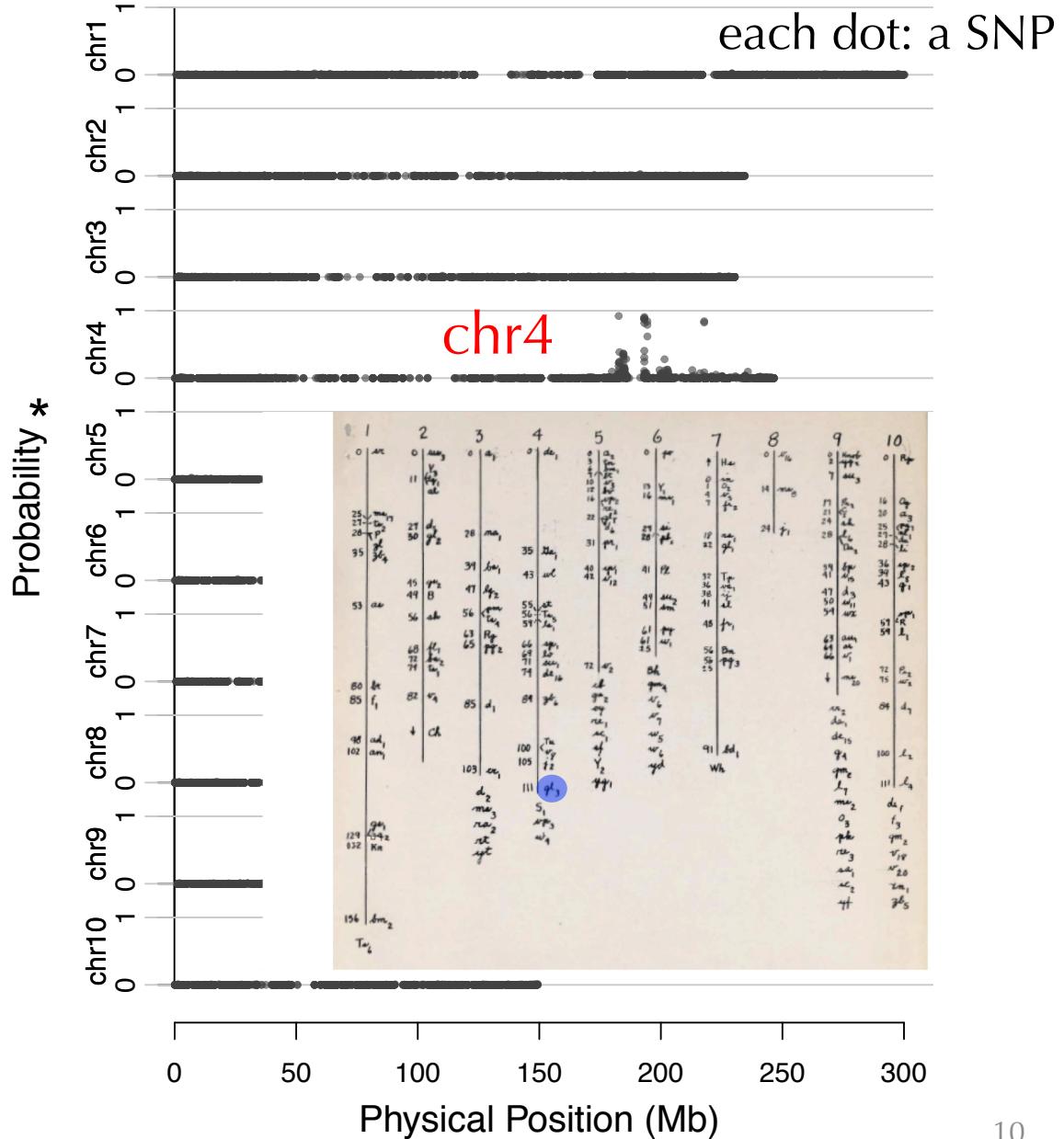


A Bayesian approach was developed to calculate **the probability of complete linkage** between the SNP site and the mutant gene.

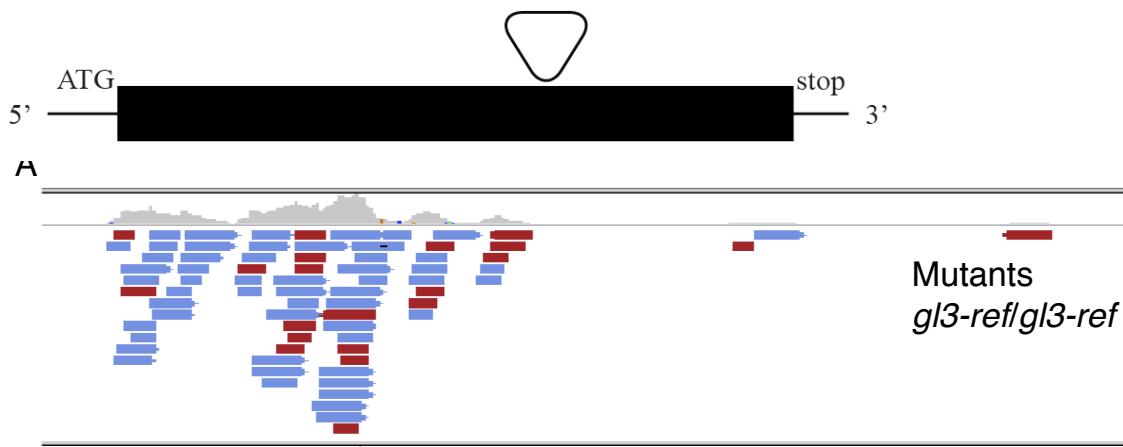
BSR-Seq mapping result

* Probability of complete linkage between the SNP and the *gl3* gene

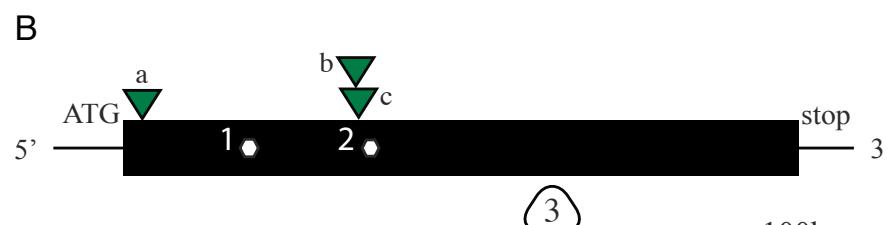
gl3_RNA-seq_BSA



gl3 cloning



Down-regulated
Partially expressed

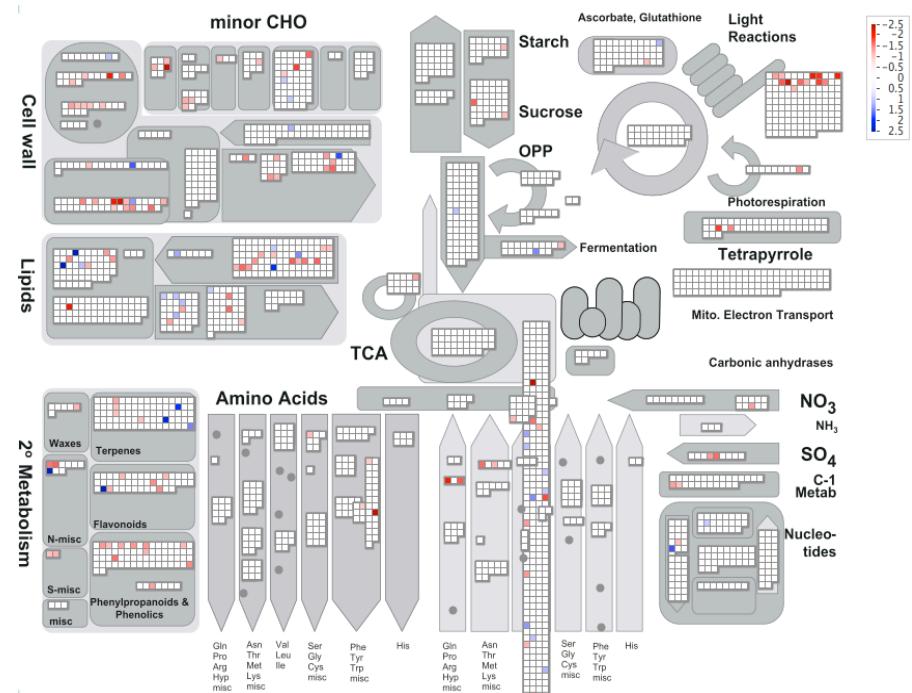
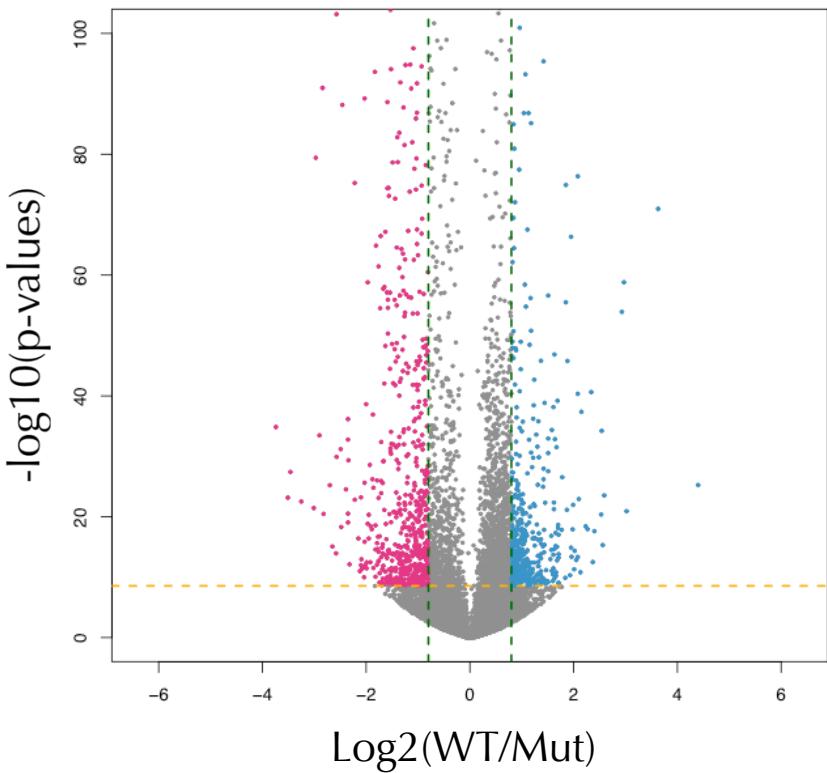


GRMZM2G162434, an R2R3 type *myb* transcription factor

▼ *Mu* insertions □ nonsense mutations ○ unknown insertions in the allele of *gl3-ref*

Figure 2

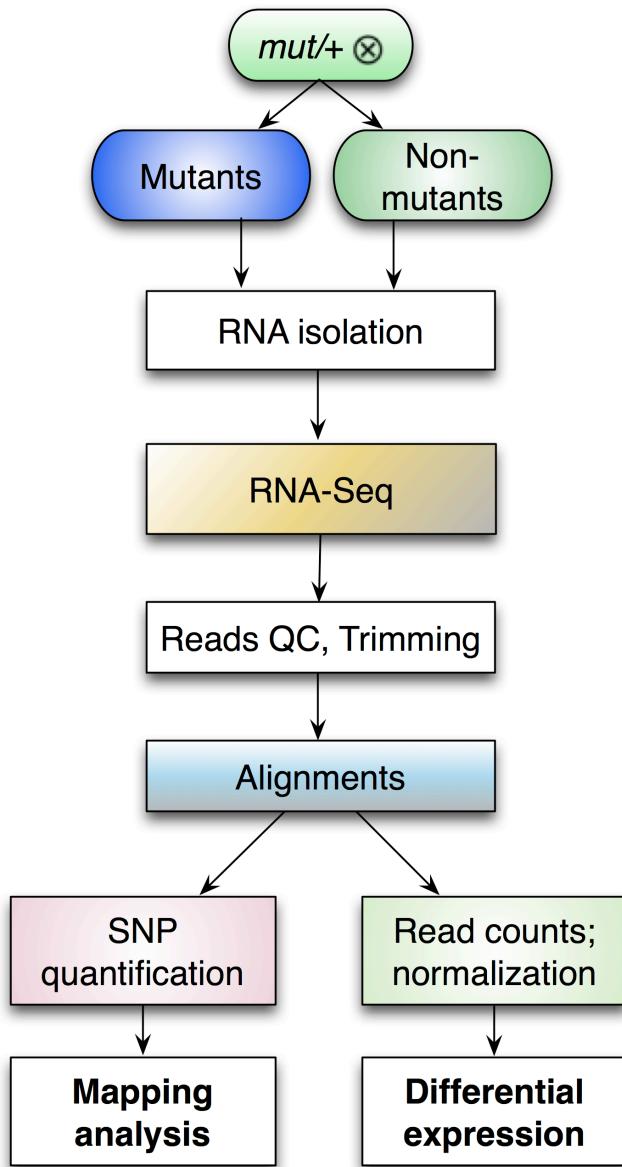
Differential expression analysis of *gl3* using the SAME RNA-Seq data used for BSR-Seq



MapMan view

1,095 genes with significantly differential expression

Summary (BSR-Seq)



1. Define genetic markers
2. Map the causal gene or identify trait-associated genetic markers
3. Genome-wide gene expression

Liu, S et al., 2012 PLoS ONE, 7(5): e36406.

Outline

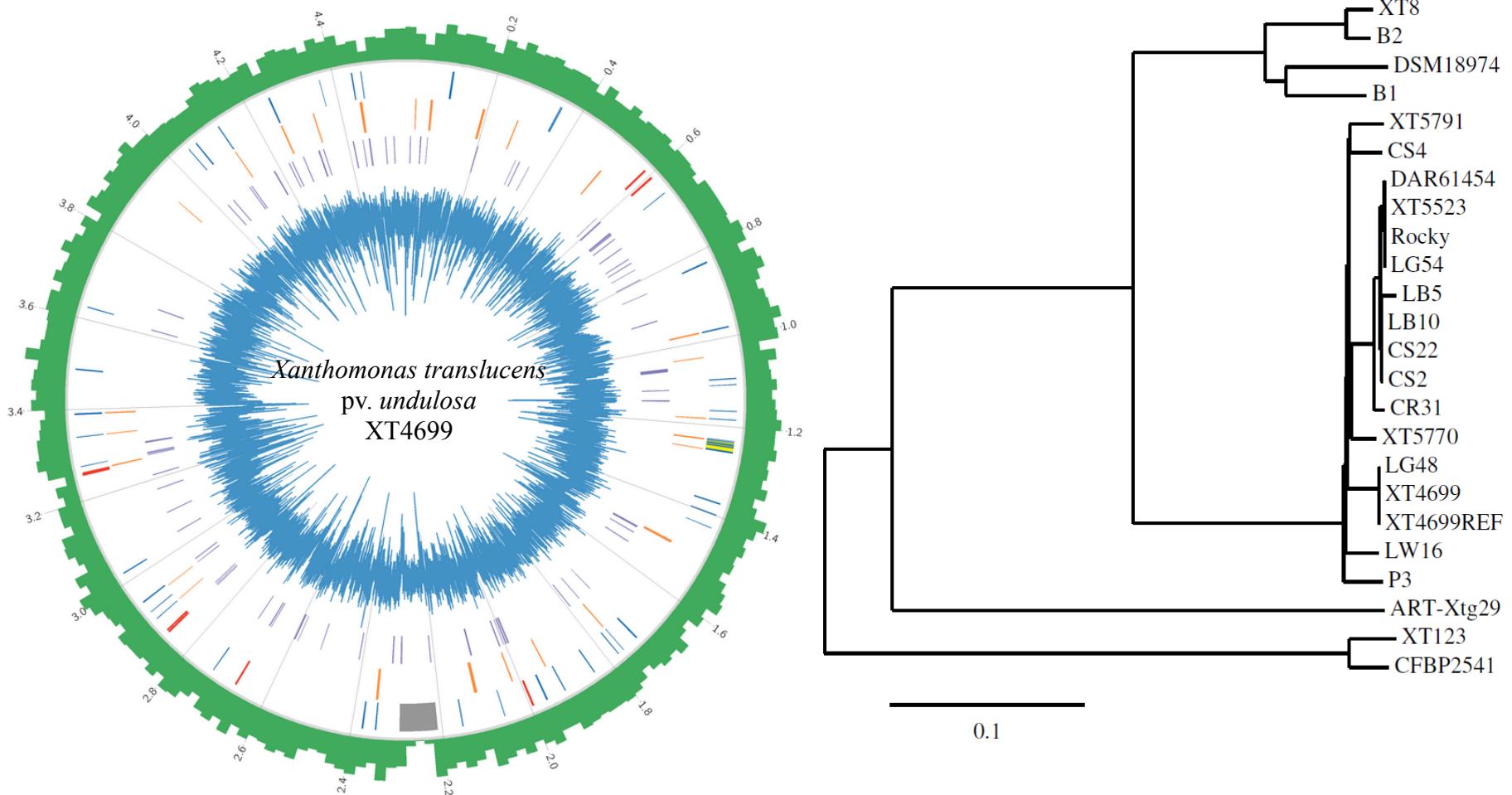
Part 1. Genetic mapping via RNA sequencing

Part 2. bacterial genome assemblies

Part 3. k-mer approach to characterize genome difference

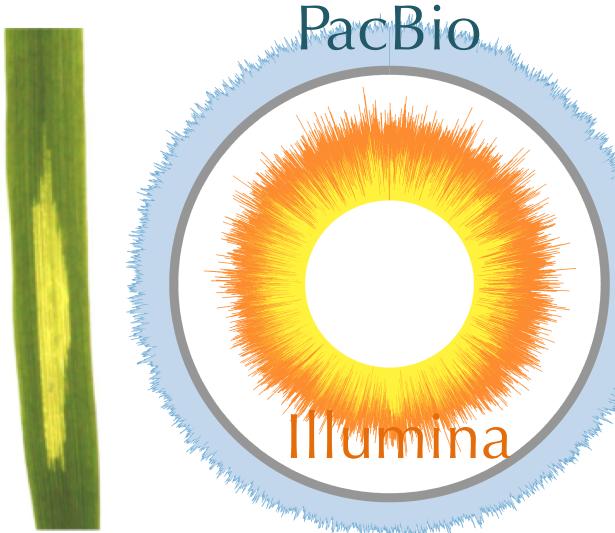
Part 4. Copy number variation in host resistance to Goss's wilt

De novo assembly of a *Xanthomonas translucens* strain

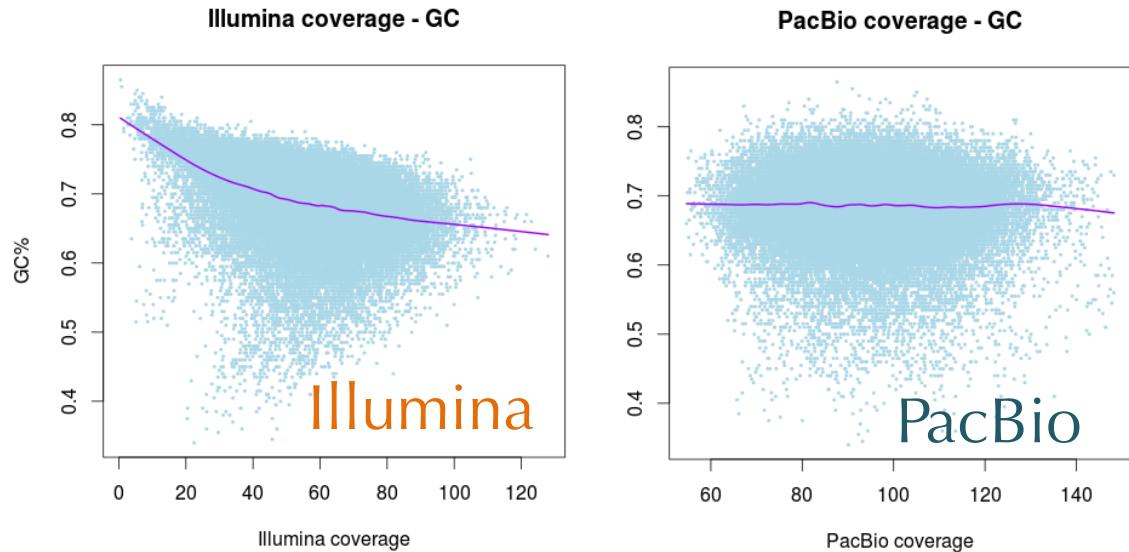


PacBio **long reads** greatly help to obtain a finished bacterial assembly

- PacBio long reads help resolve the assembly at repeats
- PacBio exhibits more uniform sequencing coverage
- Illumina exhibits sequencing bias related to GC contents



Xanthomonas translucens on wheat (4.5 Mb)



Sequencing bias at
extreme GC% regions

Genome sequences of bacterial strains

- PacBio + Illumina PE data

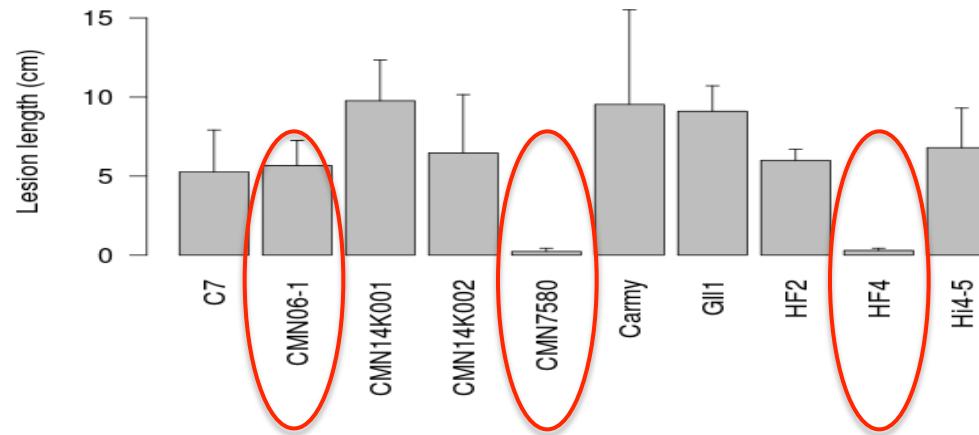


PacBio RS II

+

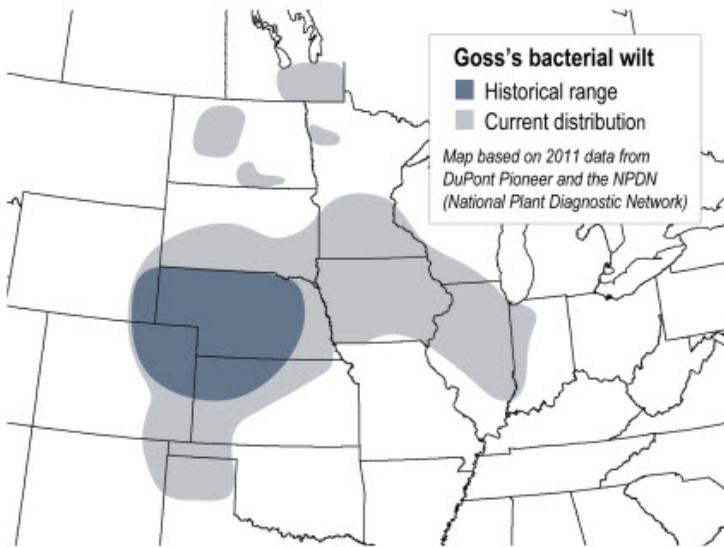


MiSeq

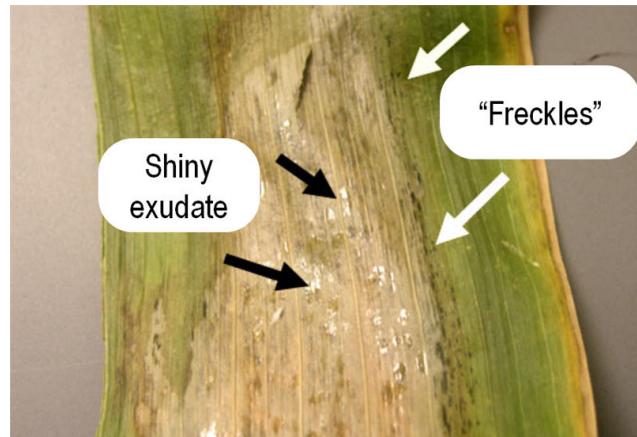


Comparison	SNPs/INDELs
CMN7580 vs. CMN06-1	5/4
HF4 vs. CMN06-1	1509/561

Goss's bacterial wilt and leaf blight (Goss's wilt)



- Goss' wilt is a bacterial disease, caused by the Gram positive bacterium, *Clavibacter michiganensis* subsp. *Nebraskense* (**Cmn**).
- Historically, it was confined to the Great Plains. Now it has spread to multiple states in the US.
- Plant wounds from wind, sandblasting, and hail provide openings for bacteria.

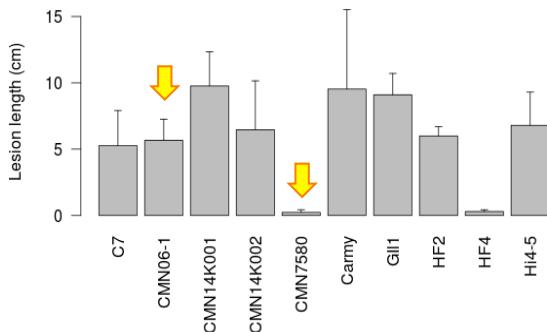


* figures from www.pioneer.com

courtesy of Drs. Tamara Jackson-Ziems & Doug Jardine

A candidate mutation in Cmn7580

CMN06-1 CMN7580



secreted peroxidase

Score	Expect	Method	Identities	Positives	Gaps
872 bits(2252)	0.0	Compositional matrix adjust.	444/445(99%)	444/445(99%)	1/445(0%)
Query 1	MTDHDSTAAGDTAAAAAAA MTDHDSTAAGDTAAAAAAA	PAFAGISRRGILGLLGAGALGGGLVGSAGGVMDRAFAG PAFAGISRRGILGLLGAGALGGGLVGSAGGVMDRAFAG			60
Sbjct 1	MTDHDSTAAGDTAAAAAAA	PAFAGISRRGILGLLGAGALGGGLVGSAGGVMDRAFAG			59
Query 61	ARQAAGGATYAFGHQAHQAGITTPAQDRLHFAAFDVADIDRAGLVSLLQDW	WSAAAARMTAG			120
Sbjct 60	ARQAAGGATYAFGHQAHQAGITTPAQDRLHFAAFDVADIDRAGLVSLLQDW	WSAAAARMTAG			119
Query 121	GSAGALGAVDGPYDSPPDDTGEALDLPPAGLTITFGLGPSLFTTADGVDRFGIADRRPAA				180
Sbjct 120	GSAGALGAVDGPYDSPPDDTGEALDLPPAGLTITFGLGPSLFTTADGVDRFGIADRRPAA				179
Query 181	LVDLPRFPGEALVAQATGGDLCVQACSDDPQVAVHAIRNL	SRIA FGRASIRWSQLGFGRT			240
Sbjct 180	LVDLPRFPGEALVAQATGGDLCVQACSDDPQVAVHAIRNL	SRIA FGRASIRWSQLGFGRT			239
Query 241	SSTSRAQVTPRNLFGFKDGTANIKSEDTRQVEDHWWADAGSSPAEAWMQGGSYLVARRIR				300
Sbjct 240	SSTSRAQVTPRNLFGFKDGTANIKSEDTRQVEDHWWADAGSSPAEAWMQGGSYLVARRIR				299
Query 301	MTIETWDRSSLREQERVGRTKGSGAPLSGGQEHTAPDFSATGRAGAPLIEPASHVRLAH				360
Sbjct 300	MTIETWDRSSLREQERVGRTKGSGAPLSGGQEHTAPDFSATGRAGAPLIEPASHVRLAH				359
Query 361	PDANGGAVALRRGYNFVDGNDLGLRNAGLFFLAQRDPRQFIPIQRSLARDAMNEYLR				420
Sbjct 360	PDANGGAVALRRGYNFVDGNDLGLRNAGLFFLAQRDPRQFIPIQRSLARDAMNEYLR				419
Query 421	HVGSGIWA	VPPGASRAGYVGETLFA	445		
Sbjct 420	HVGSGIWA	VPPGASRAGYVGETLFA	444		

Comparison	SNPs/ INDELS
CMN7580 vs. CMN06-1	5/4

Highly abundant **fungal peroxidase** was found in xylem, and demonstrated to be required to promote fungal colonization in hosts. - Flajsman et al. MPMI, 2016, 29 (5), 362-373.

Outline

Part 1. Genetic mapping via RNA sequencing

Part 2. bacterial genome assemblies

Part 3. k-mer approach to characterize genome difference

Part 4. Copy number variation in host resistance to Goss's wilt

Hybrid vigor from two inbred lines

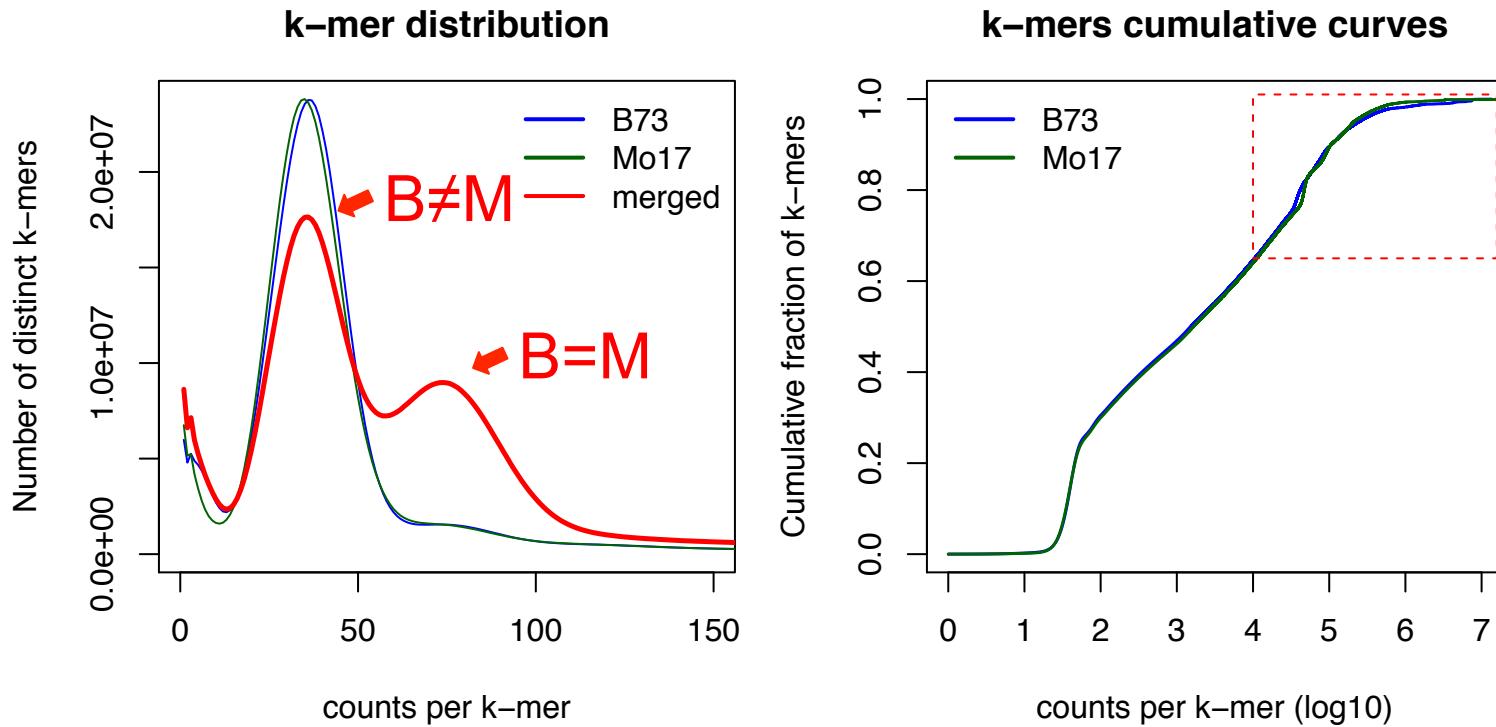


B73

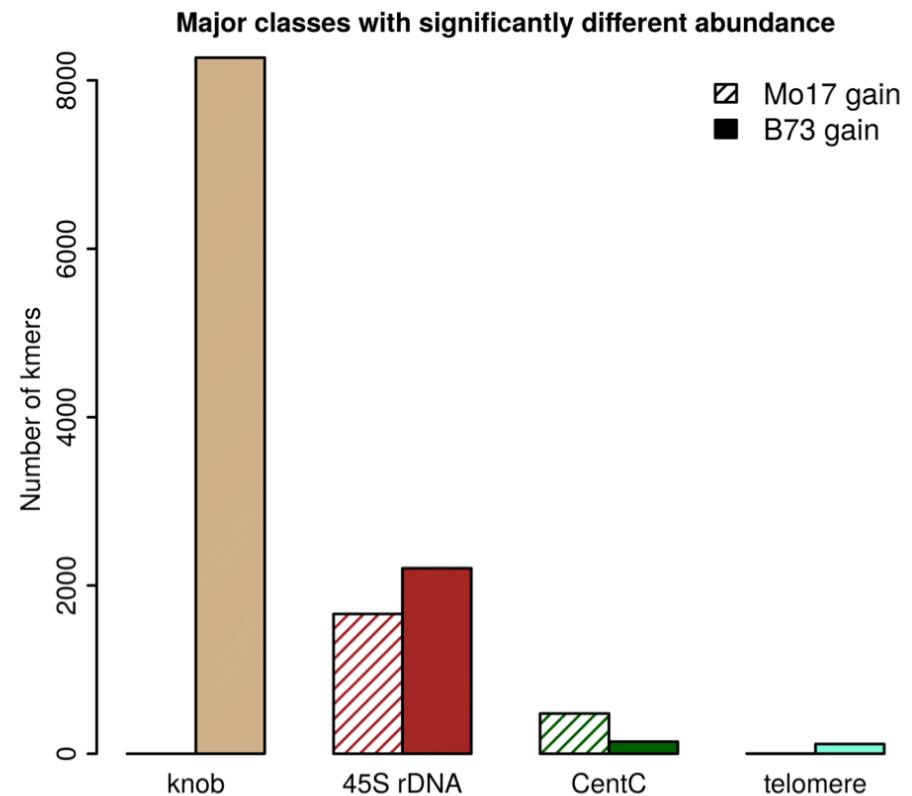
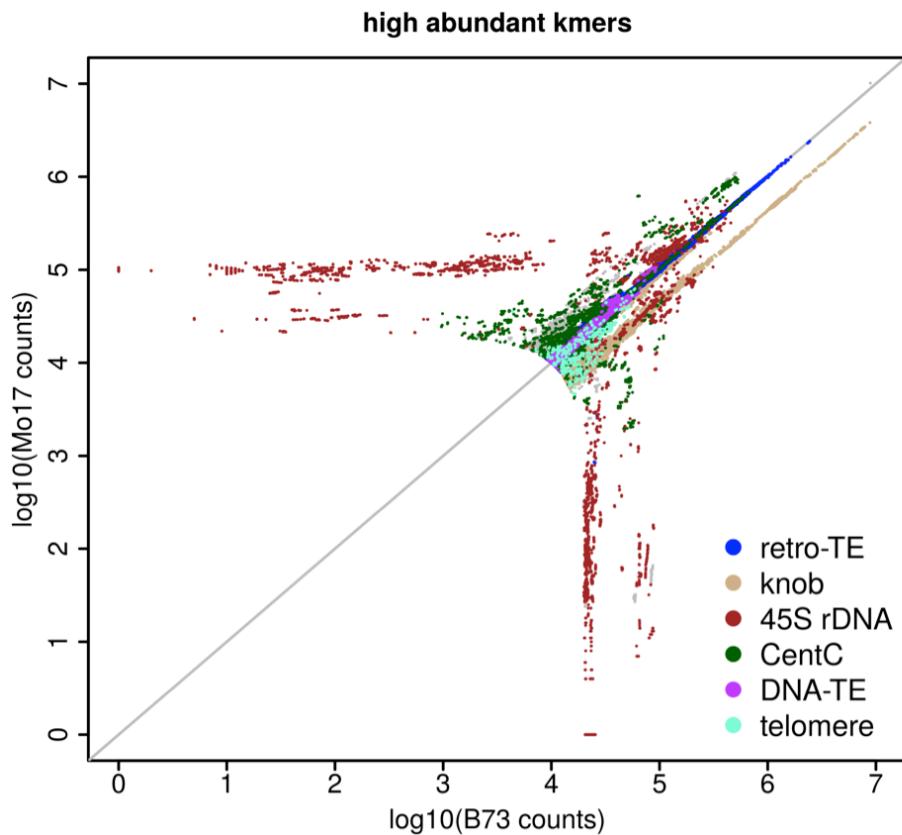
F1

Mo17

Genome comparison via plotting k-mer abundance distributions

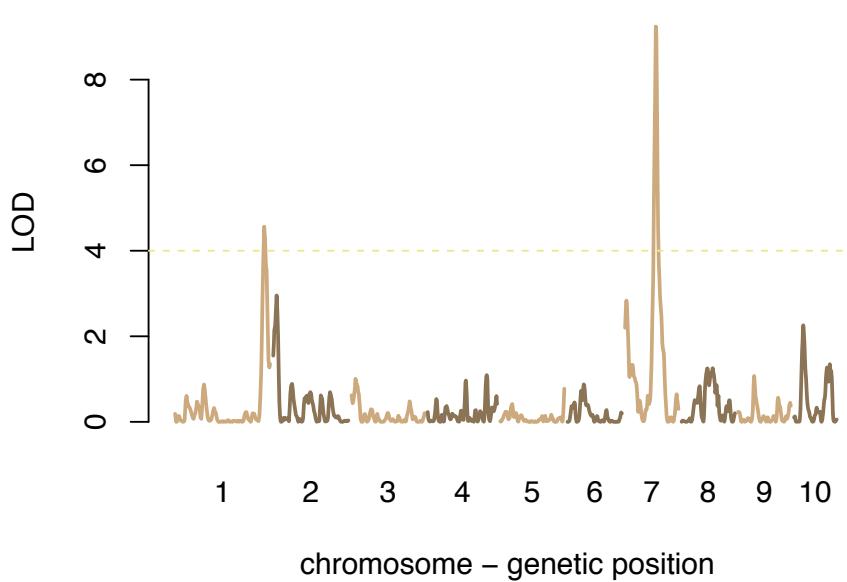


Abundance difference at repetitive regions

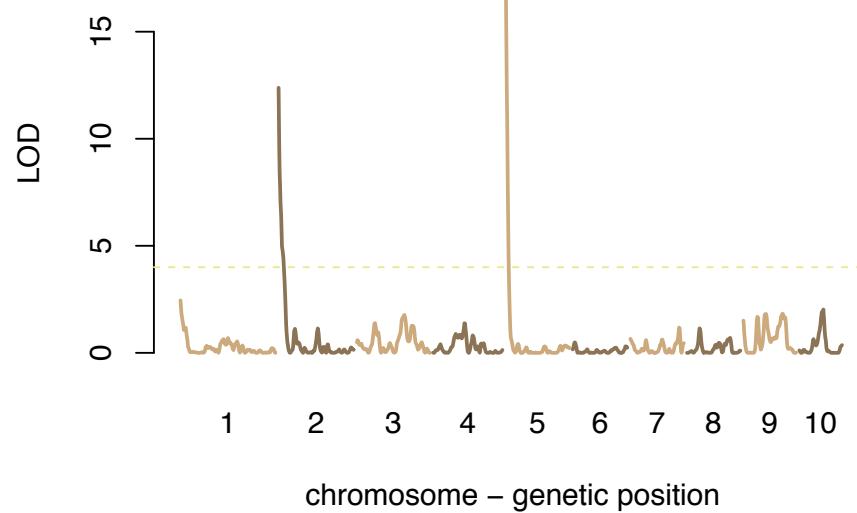


Genetic mapping of loci affecting abundance difference

AAACATATGTGGGGTTAGGTGTATG
knob–B73gain



CCATTAAAGTACGTGTTCCACCA
telomere–B73gain



Outline

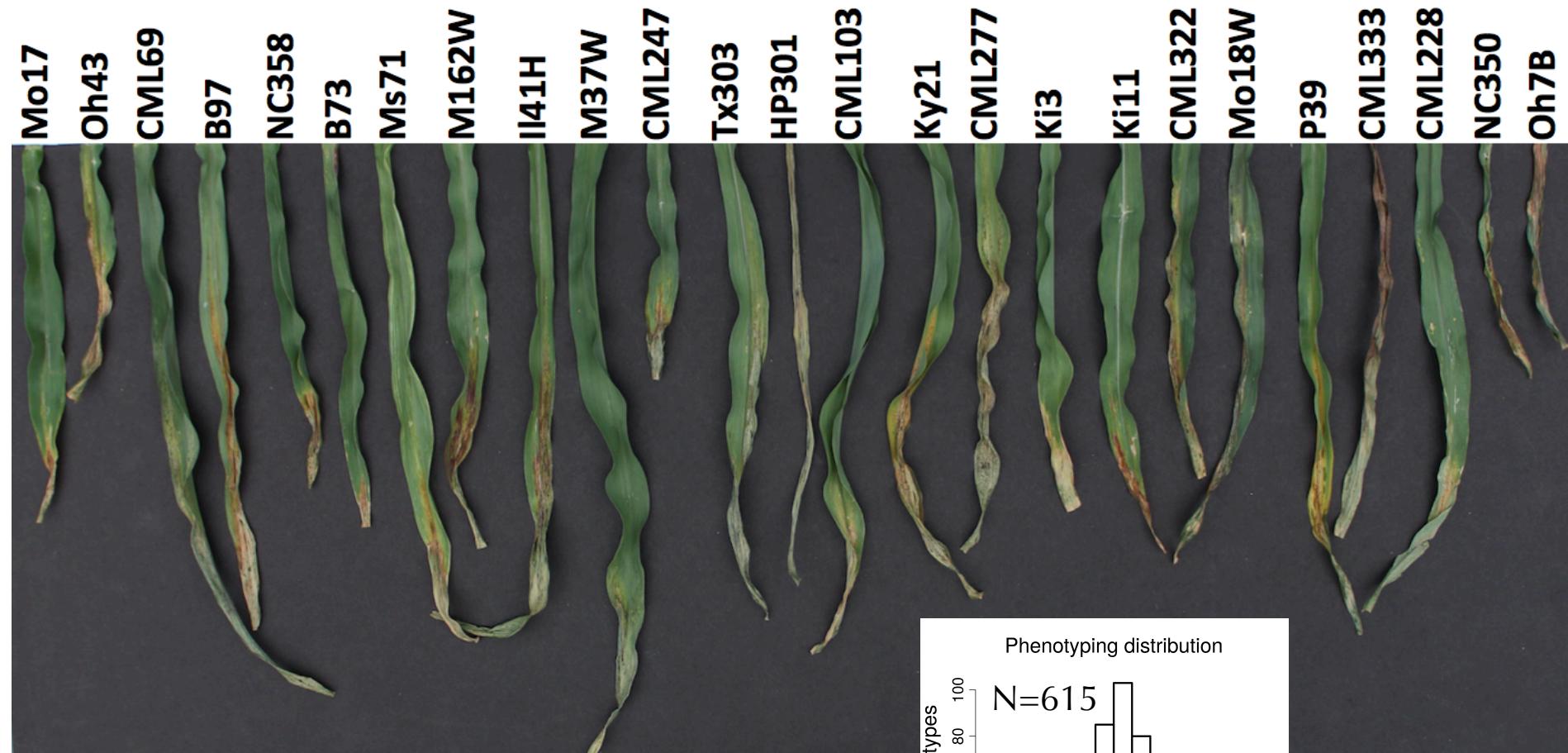
Part 1. Genetic mapping via RNA sequencing

Part 2. bacterial genome assemblies

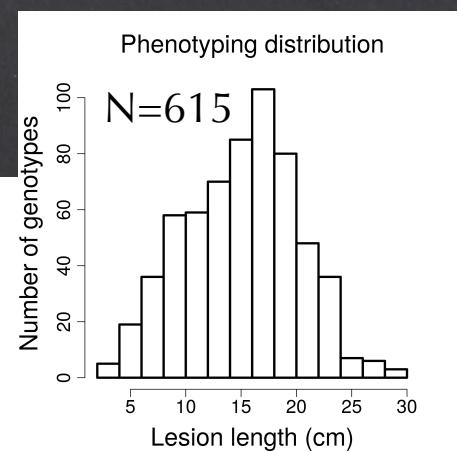
Part 3. k-mer approach to characterize genome difference

Part 4. **Copy number variation** in host resistance to Goss's wilt

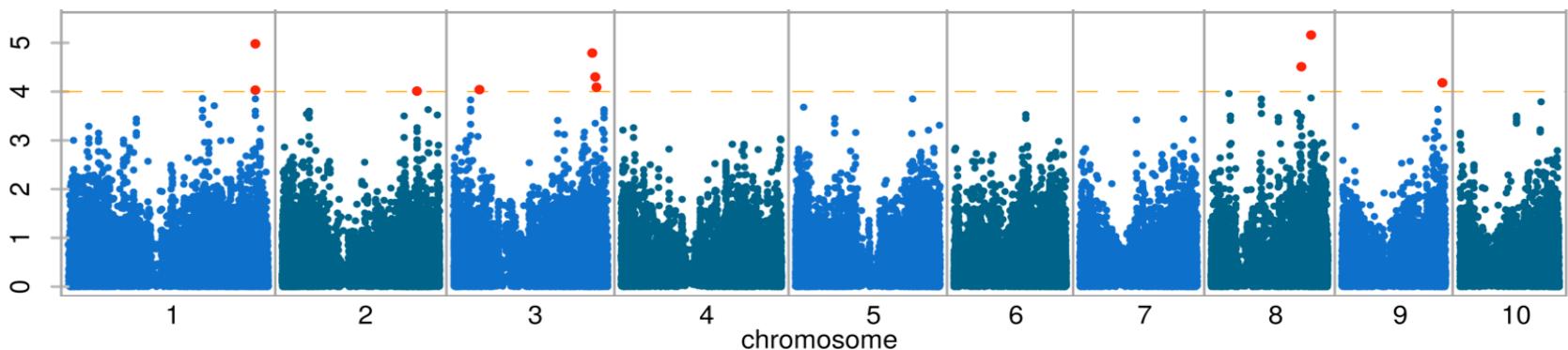
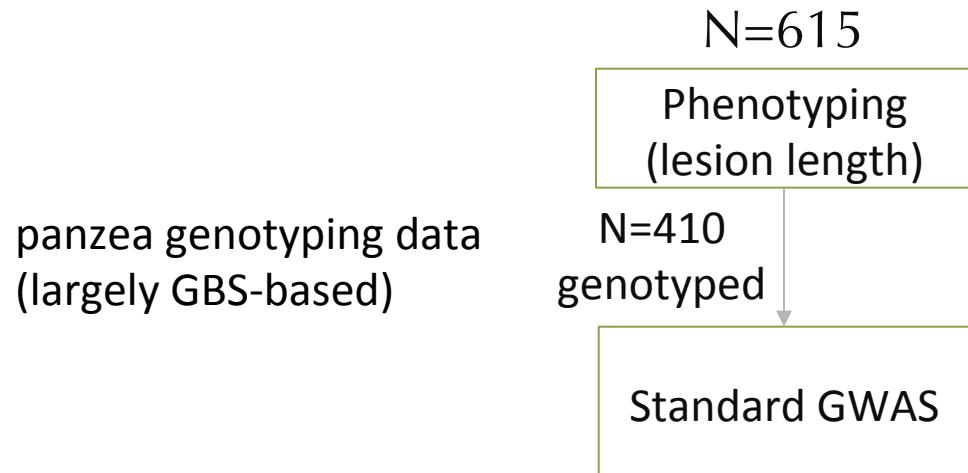
Phenotype maize lines through the clipping method



3 - 8 replicates (most)



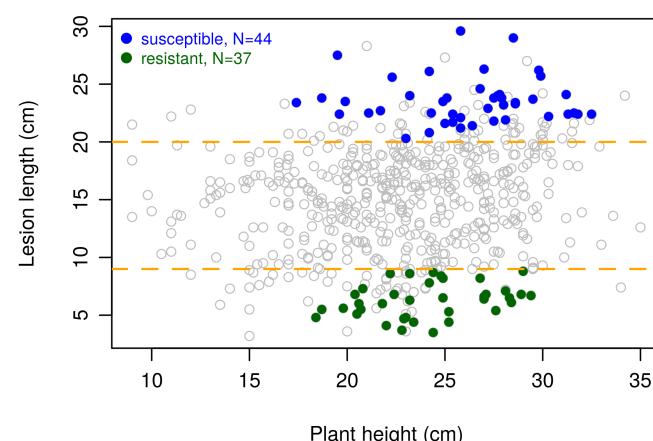
Standard GWAS



low-density markers

Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel

(a) Pooling based on phenotype
 Jinliang Yang^{1,†}, Haiying Patrick S. Schnable^{1,4,*}
¹Department of Agronomy,
²Technology Innovation,
³Department of Statistics
⁴Center for Plant Genom



XP-GWAS

N=615

Phenotyping
(lesion length)

N=37

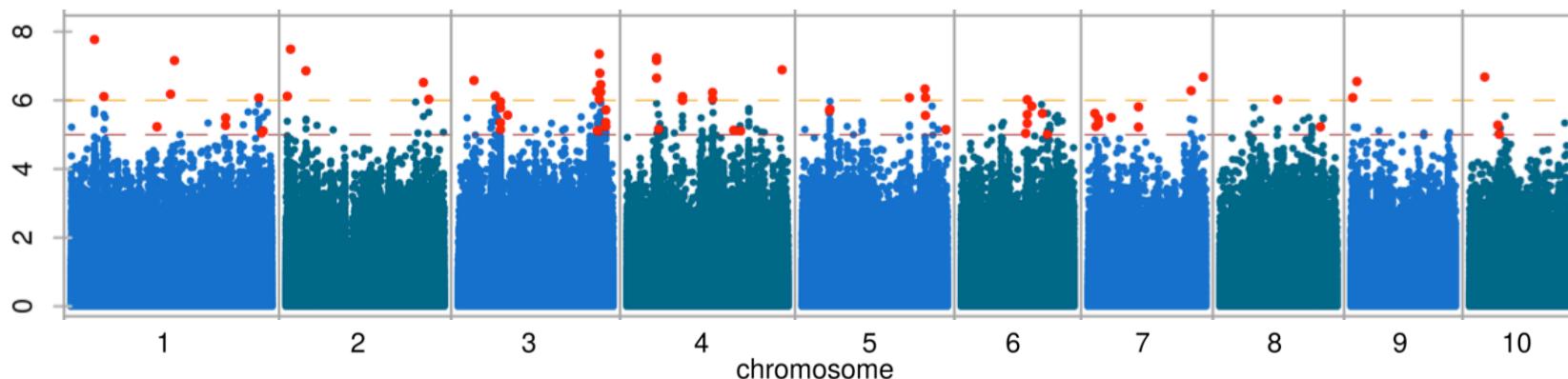
N=44

Highly R lines

Highly S lines

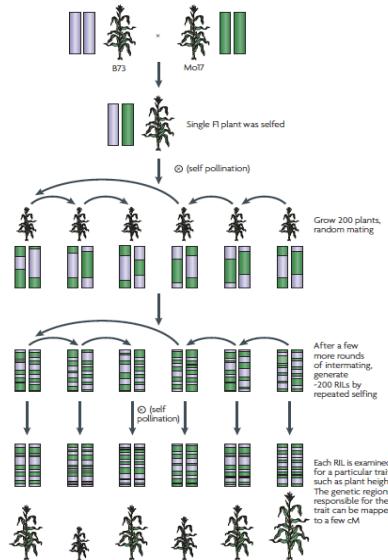
Whole genome sequencing

XP-GWAS

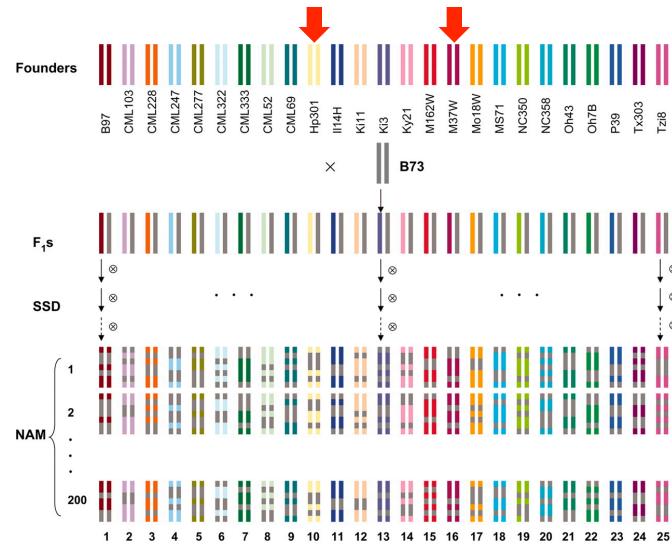
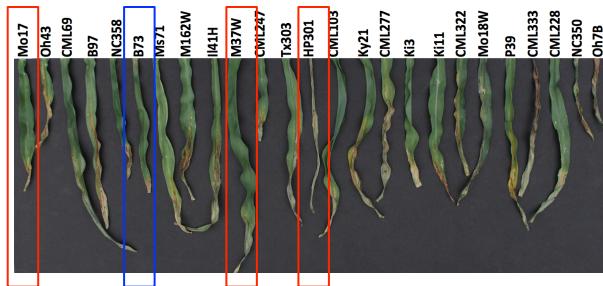


QTLs from three RIL populations

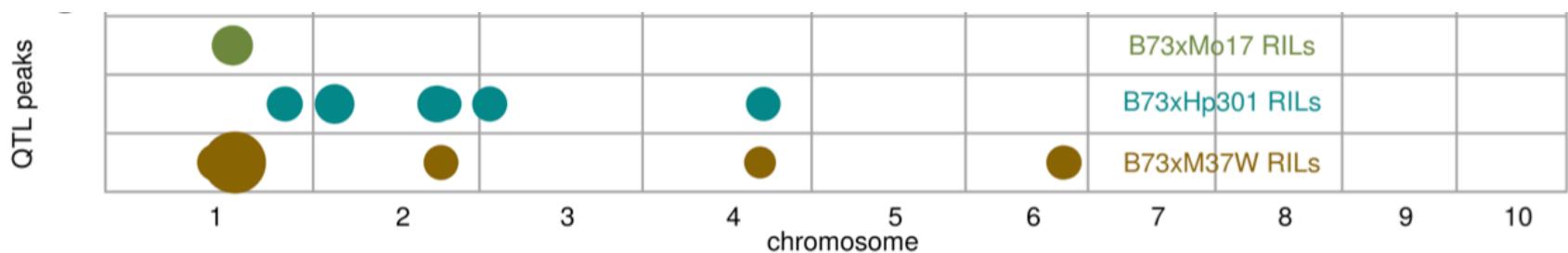
Intermated B73xMo17 RILs



Candela et al. 2008, Nat Rev Genet 9:192--203

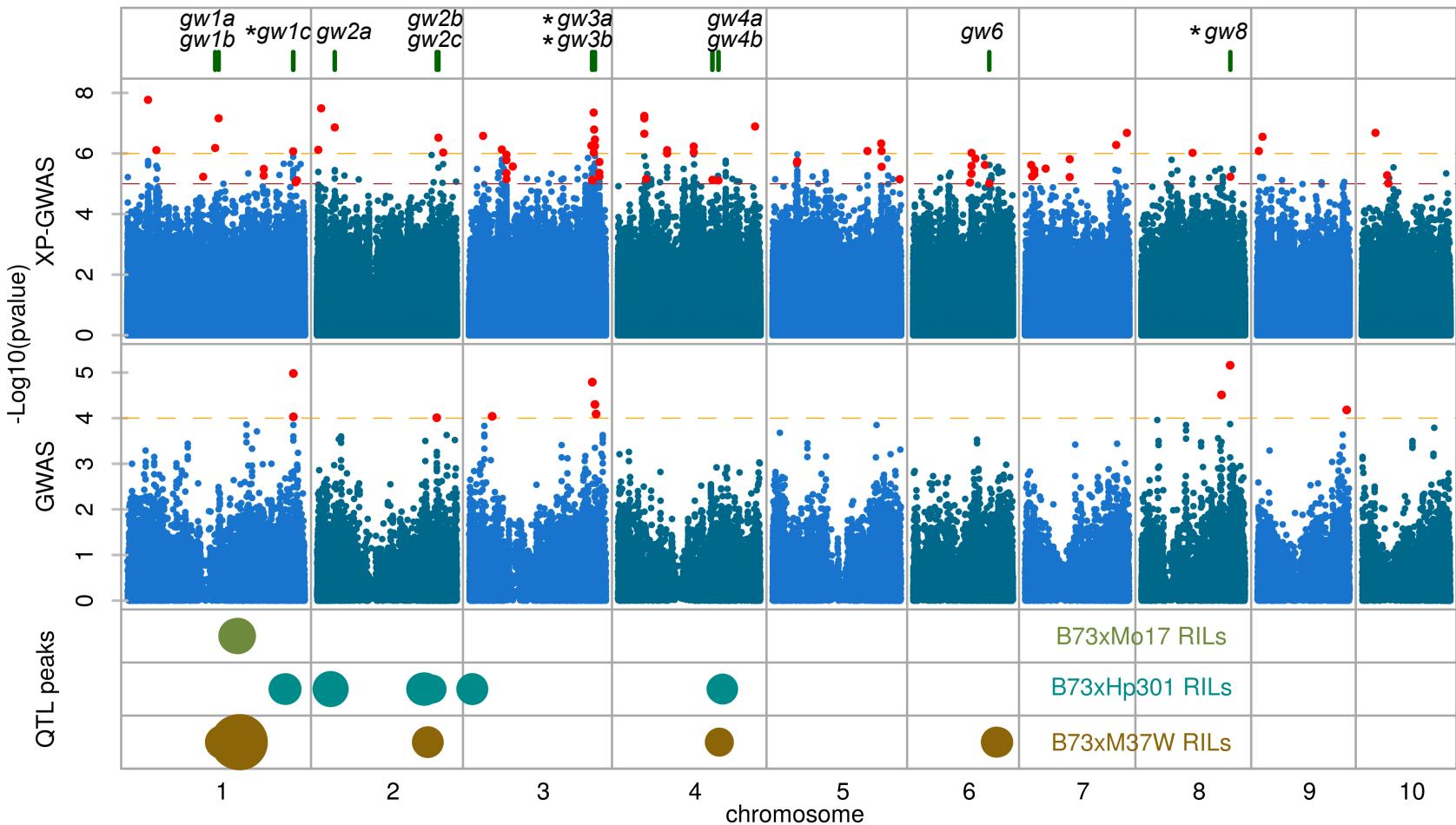


Yu et al., 2008, Genetics 178:539-551



Disease associated QTLs were identified at chromosomes 1, 2, 3, 4, 6

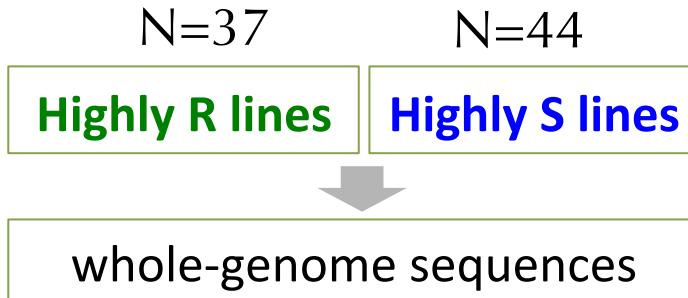
Summary of QTL, GWAS and XP-WGS analysis



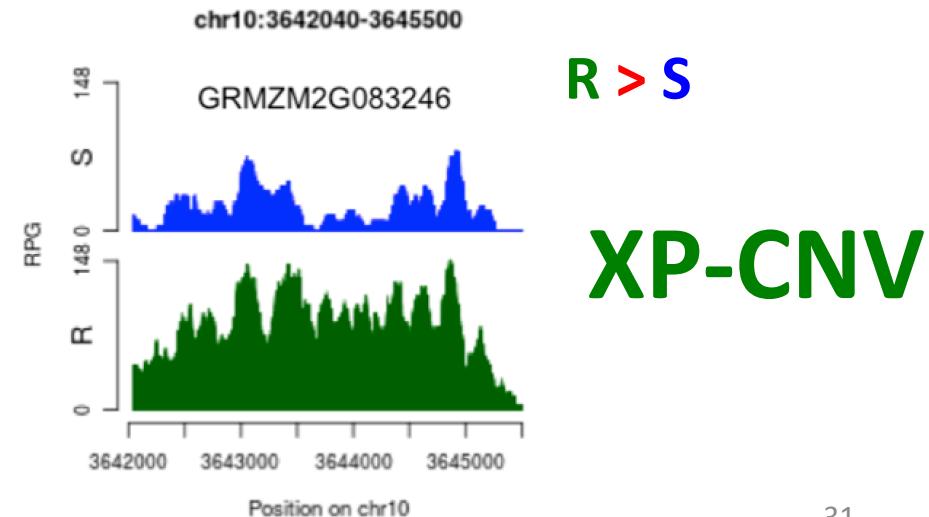
Twelve disease associated loci were identified

Disease associated Copy Number Variation (CNV)

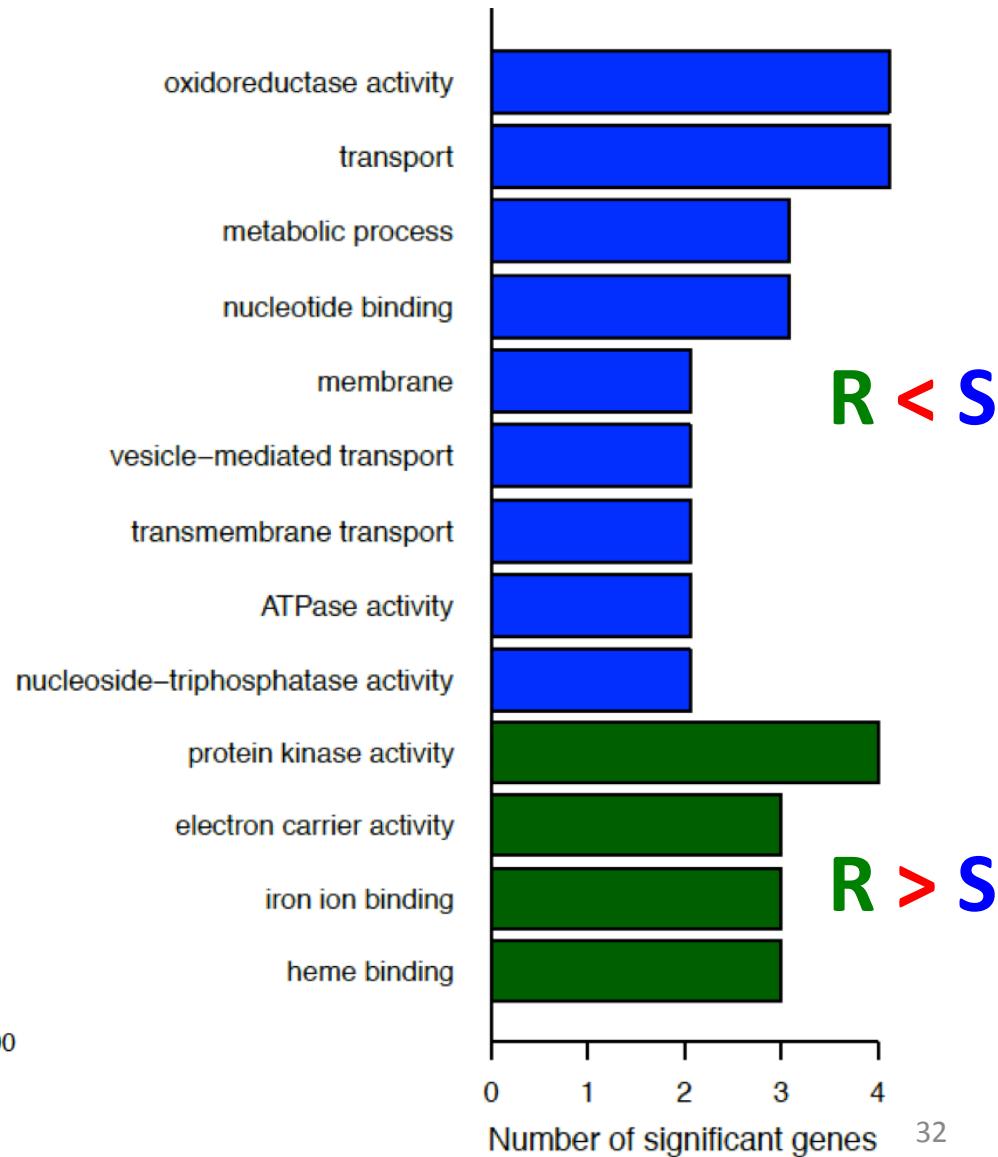
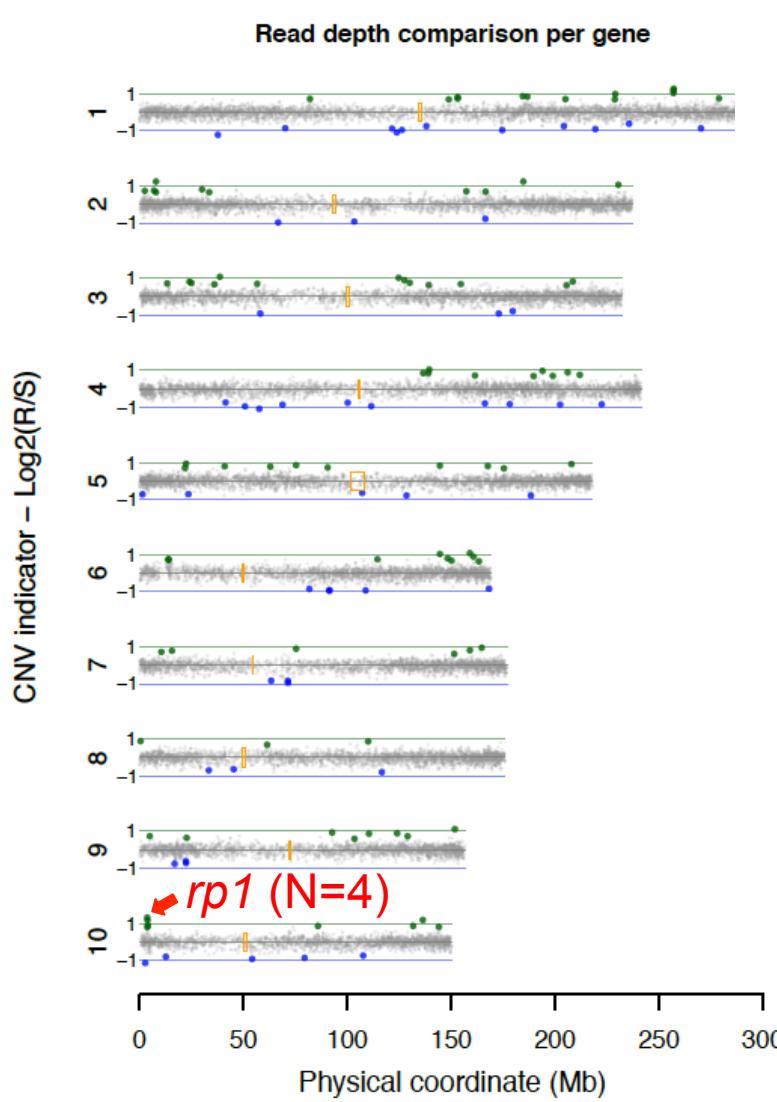
- Plant defense genes are highly polymorphic
(dynamically co-evolve with pathogens)
- CNV has been identified for plant defense genes in many species.
(e.g., higher copy number of *Rhg1* confers stronger resistance to soybean cyst nematode)
- **Maize CNV genes are enriched in the pathways of stress responses.**



Compare sequencing depth
(read counts) between R and S



141 CNV genes identified via XP-CNV



Disease resistance was found in multiple *Rp1* accessions

Multiple *Rp1* alleles were introgressed into a susceptible line H95 for 3-8 generations

