# Reverse vs. forward genetics

- There are basically two ways to link the sequence and function of a specific gene: forward and reverse genetics
- Reverse approaches rely upon sequence information as retrieved from genome and transcript profiling projects and tries to gain insight into the underlying function by selecting for mutation
- Forward genetics aims to identify the sequence change that underlies a specific mutant phenotype. The starting point is an already available or a specifically searched for and predicted phenotypic mutant of interest

(Peters et al., 2003)

# QTL mapping using a bi-parental mapping population

- Locate the gene in a broad chromosomal region
  1) Create a **segregating population** for the interested trait
  2) Genotype the population for **molecular markers** and construct **linkage map**
  3) **Phenotype the population** for the interested trait
  4) Perform **marker-trait statistical analysis** to find markers linked to the causal gene
- Fine mapping
- Candidate gene validation

# Outline

- QTL mapping (linkage mapping) and map-based clone
- Genome wide association mapping (linkage disequilibrium mapping)
- Mapping-by-sequencing using next-generation sequencing and induced mutations

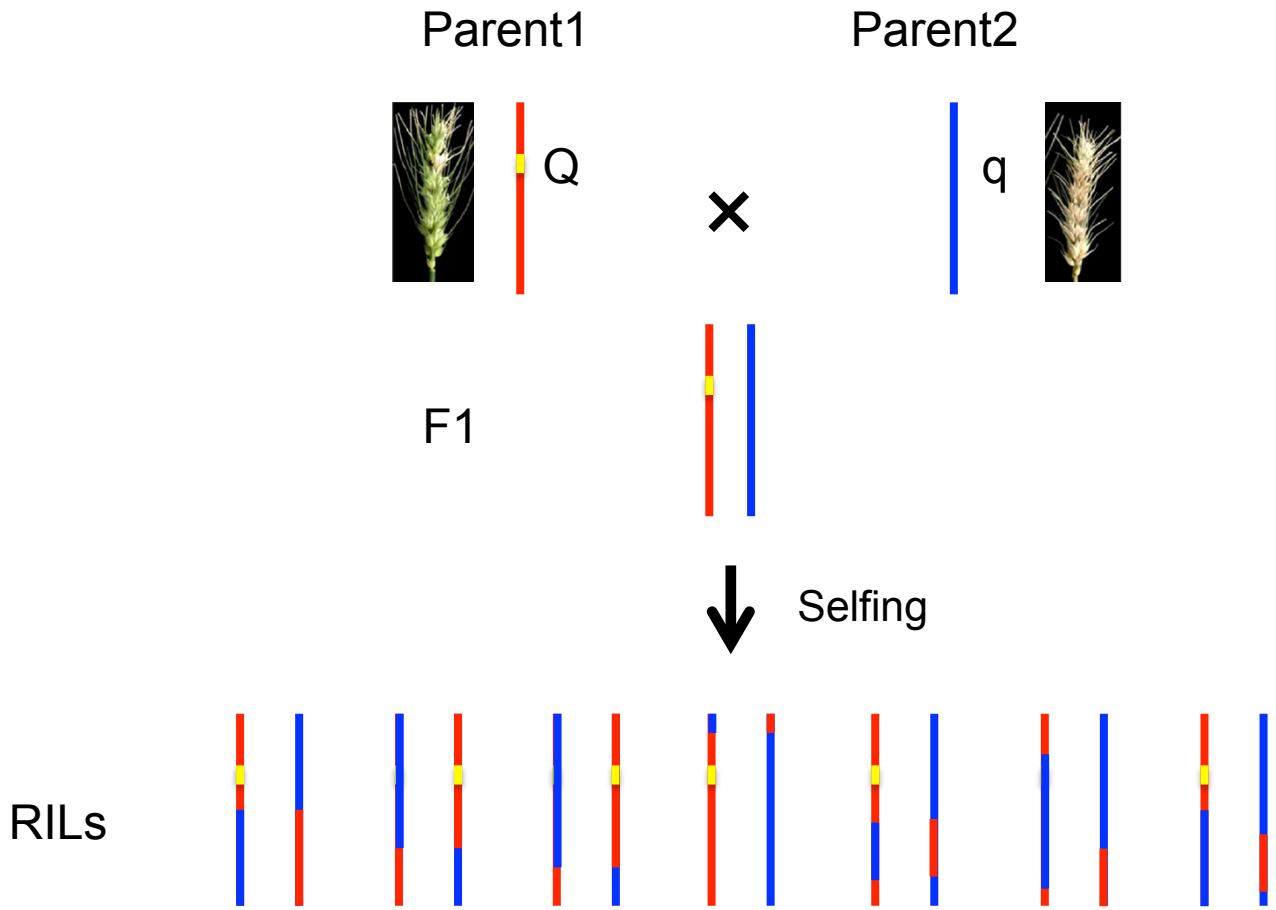
- ✓ In 1990s, DNA molecular markers are available, which make gene mapping feasible
- ✓ Reference genome and development of saturating marker technologies further facilitate QTL mapping and map-based clone

# Suggested papers to read

1. Bergelson, J. and Roux, F., 2010. Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature Reviews Genetics*, *11*(12), p.867.
2. Borevitz, J.O. and Chory, J., 2004. Genomics tools for QTL analysis and gene discovery. *Current Opinion in Plant Biology*, *7*(2), pp. 132-136.
3. Mackay, I. and Powell, W., 2007. Methods for linkage disequilibrium mapping in crops. *Trends in plant science*, *12*(2), pp.57-63.
4. **Peters, J.L., Cnudde, F. and Gerats, T., 2003. Forward genetics and map-based cloning approaches. *Trends in plant science*, *8*(10), pp.484-491.**
5. **Schneeberger, K., 2014. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature Reviews Genetics*, *15*(10), p.662.**

# Linkage mapping

## --Create a mapping population



# Molecular marker

- There are millions of mutations or variations between two parents
- Molecular marker can be used to locate causal variation for an interested trait
- In genetics, a **molecular marker** (or genetic marker) is a fragment of DNA that is associated with a certain location within the genome

……ATCTTCGCCATAAAGATGAAGTT……  Sample1

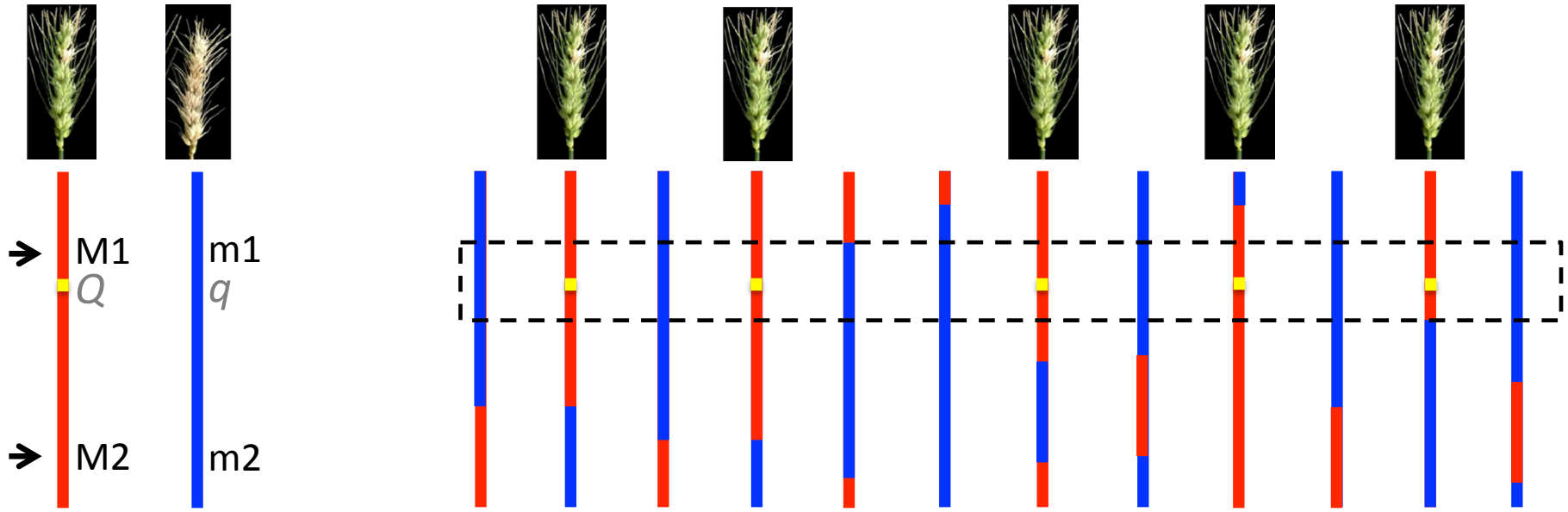
……ATCTTCGCCAAGATGAAGTT……  Sample2  3 bp deletion

……ATCTTCGCCATAAAGATGAAATT……  Sample3  SNP

# Molecular marker

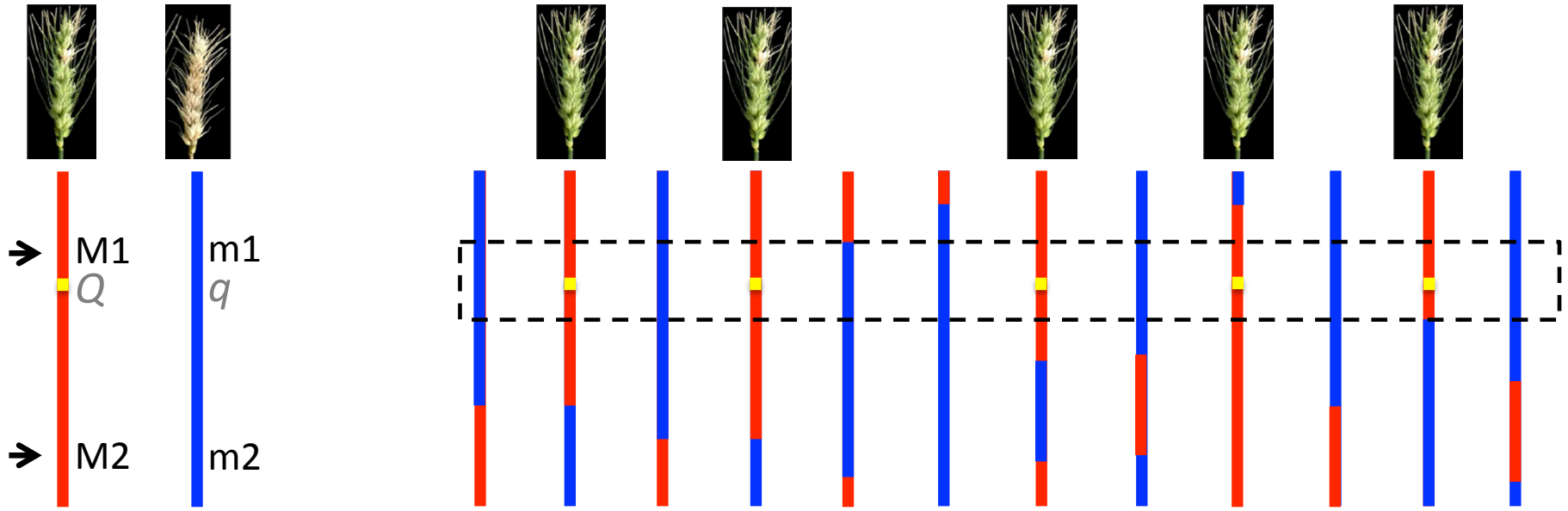
- Types of molecular markers
  - RFLP, AFLP, RAPD, SSR, STR, CAPS, SCAR, SNP, etc.
- Suggested paper to read
  - Peters, J.L., Cnudde, F. and Gerats, T., 2003. Forward genetics and map-based cloning approaches. *Trends in plant science*, *8*(10), pp. 484-491.

# Linkage mapping




| SNPs | SNP1 | | SNP2 | |
|---|---|---|---|---|
| Alleles | M1 (100) (100 Q) | m1 (100) (100 q) | M2 (100) (50 Q and 50 q) | m2 (100) (50 Q and 50 q) |
| Mean resistance | 20% | 70% | 45% | 48% |
| *p*-value | 0.0001 | | 0.2 | |

# Linkage mapping



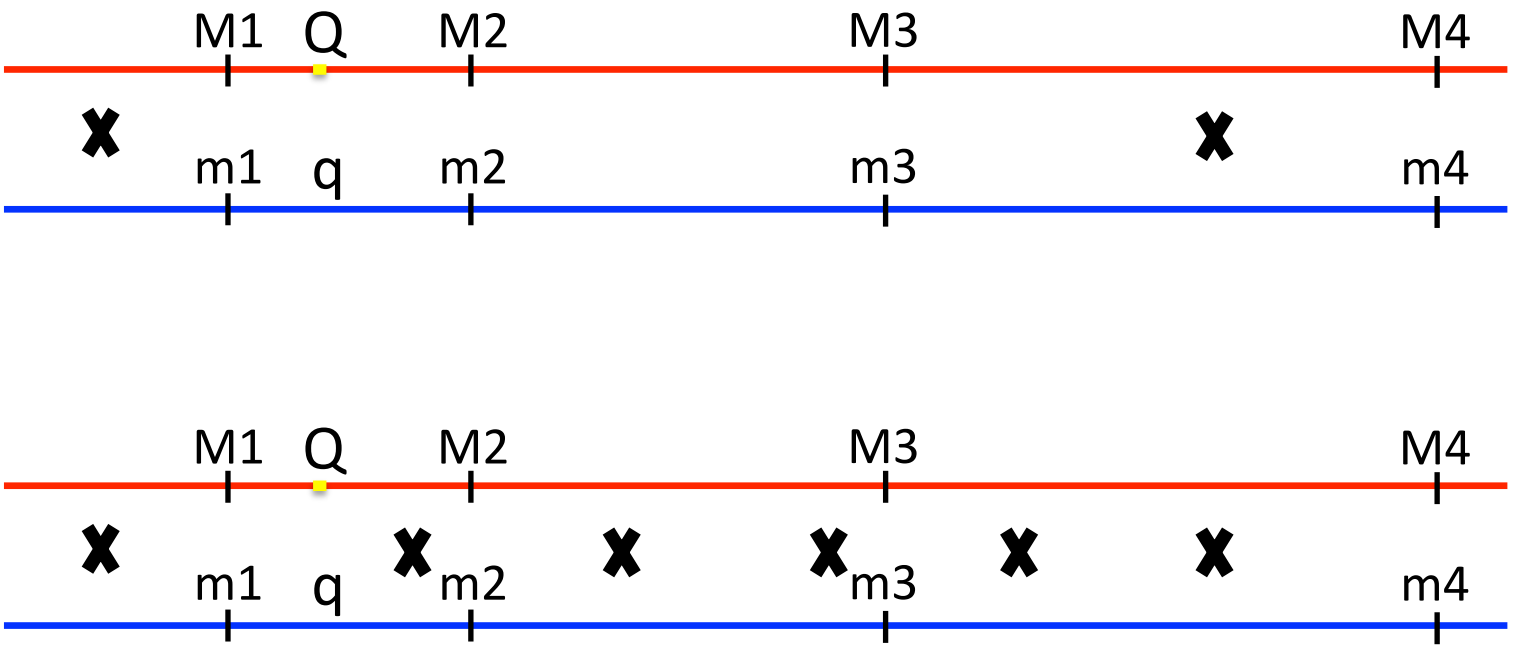

| SNPs | SNP1 | | SNP2 | |
|---|---|---|---|---|
| Alleles | M1 (100) (95Q and 5q) | m1 (100) (95q and 5Q) | M2 (100) (50 Q and 50 q) | m2 (100) (50 Q and 50 q) |
| Mean resistance | 25% | 65% | 45% | 48% |
| *p*-value | 0.0001 | | 0.2 | |

# Linkage mapping

- Linkage
  - Two loci or markers that are physically near to each other are unlikely to be separated during chromosomal crossover
- Linkage mapping of QTL
  - Estimate the mean and variance of a specific marker locus
  - Relies on differences among the trait means of genotype at a marker locus

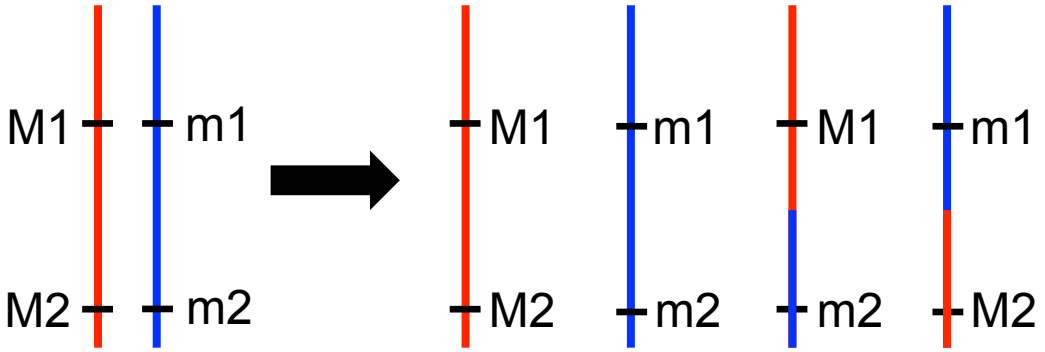
# Recombination and linkage: the basis of linkage mapping

- Linkage and genetic linkage map
  - Markers on the same chromosome tend to be inherited together
  - The closer, the more likely to be inherited together



# Recombination and linkage: the basis of gene mapping

- Meiosis and recombination: two copies of the same chromosome break and rejoin at the same point

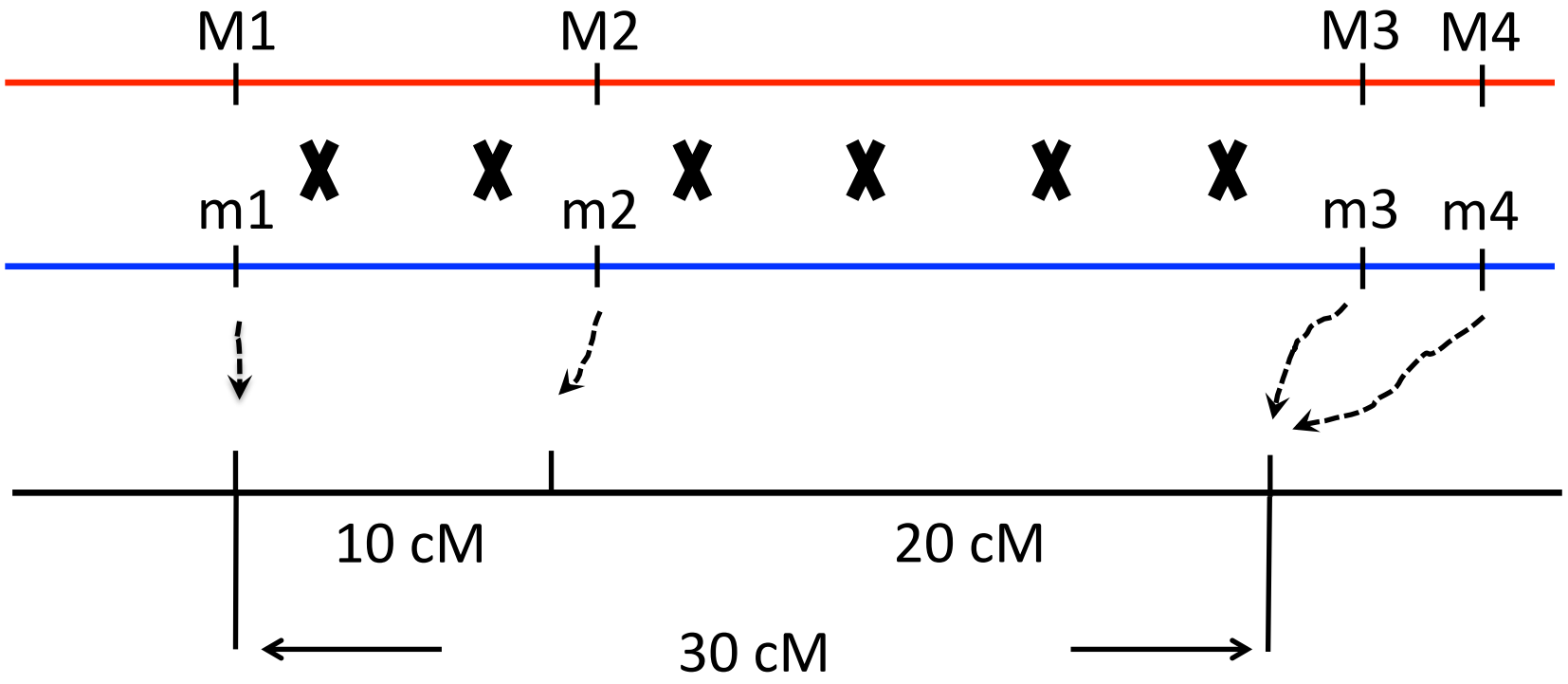



| Gamete | Counts |
|---|---|
| M1M2 | 92 |
| m1m2 | 88 |
| M1m2 | 10 |
| m1M2 | 10 |
| Rec. rate (r) | 0.10 (10 cM) |

**r= (M1m2 + m1M2)%**

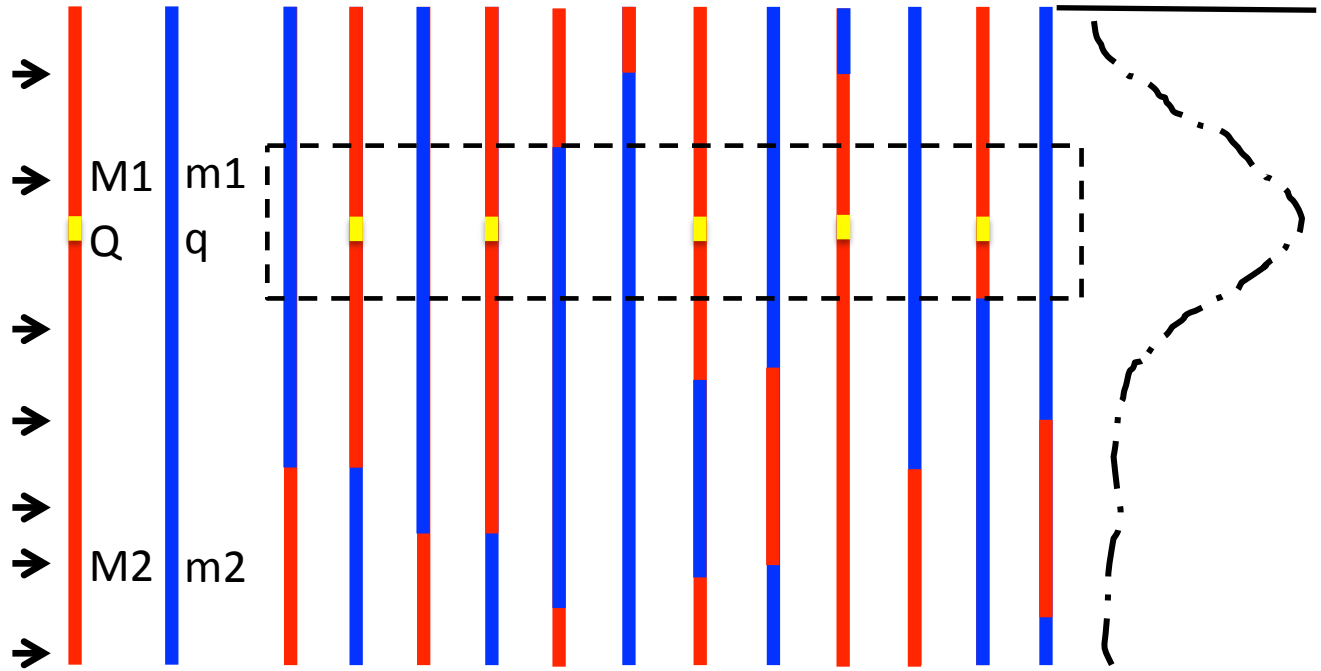
# Recombination and linkage: the basis of gene mapping

- Linkage and genetic linkage map
  - Markers on the same chromosome tend to be inherited together
  - The closer, the more likely to be inherited together



# Linkage mapping

## --Linked markers to locate the candidate gene





| | Q | q |
|---|---|---|
| M1 | 100 | 0 |
| m1 | 0 | 100 |


| | Q | q |
|---|---|---|
| M2 | 50 | 50 |
| m2 | 50 | 50 |

# Single marker analysis of QTL mapping

- Single marker regression analysis

$$y_j = b_0 + bx_j + e_j$$

  - Considering one marker at one time
  - Differences among the means of MM, Mm, and mm individuals can be tested for significance with t-test, F-test, ***linear regression with then number of M alleles*** (two in MM, one in Mm, and zero in mm)
- Limitations
  - Location of the QTL relative to the marker can not be determined
  - Two or more adjacent markers could detect either the same or different QTL

(Zeng, 1994)

# Interval QTL mapping

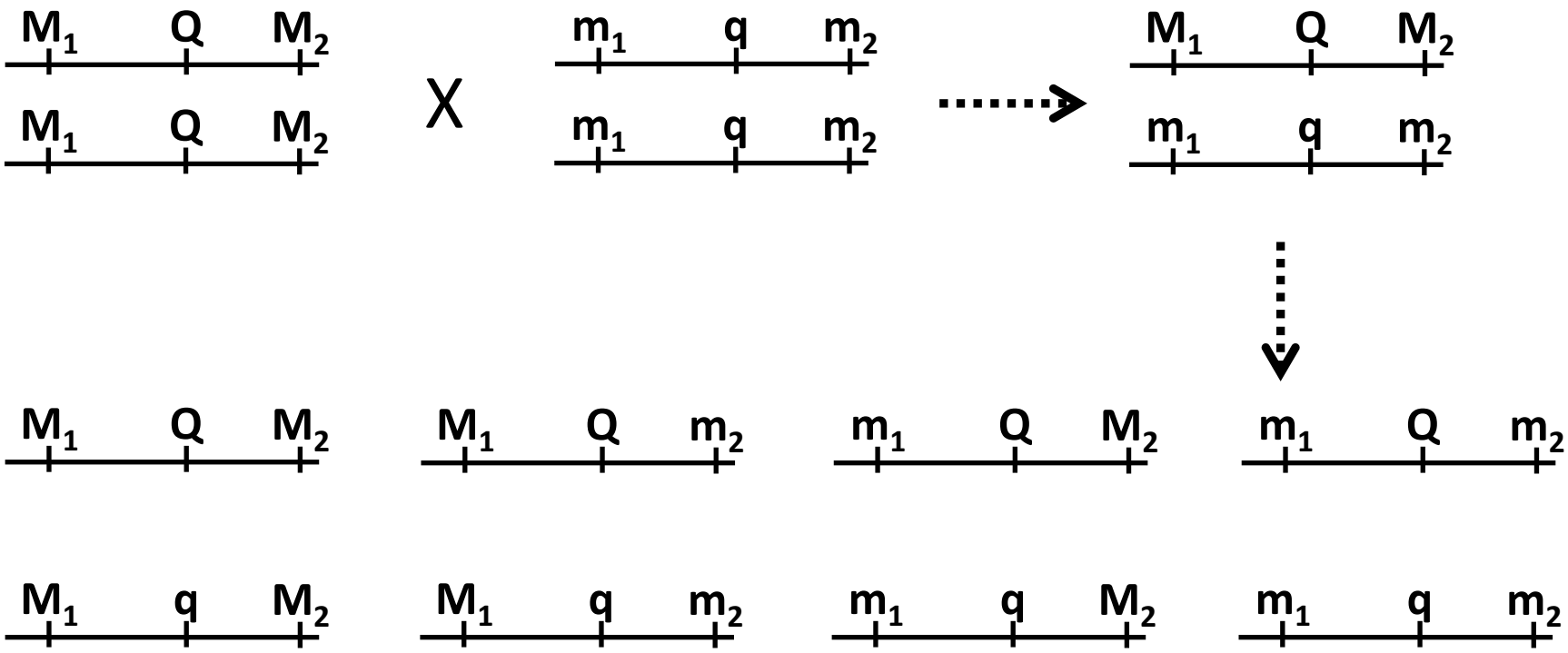
- Use of a pair of markers to disentangle distance (*r*) and QTL effect (*a*) from the test statistics
- Multiple linear regression to locate QTL and estimate QTL effects

$$y_j = b_0 + b^* x_j^* + e_j$$

- Advantages
  - Position of the QTL can be inferred
  - Effects of the QTL can be estimated

# Conditional probability given a pair of markers

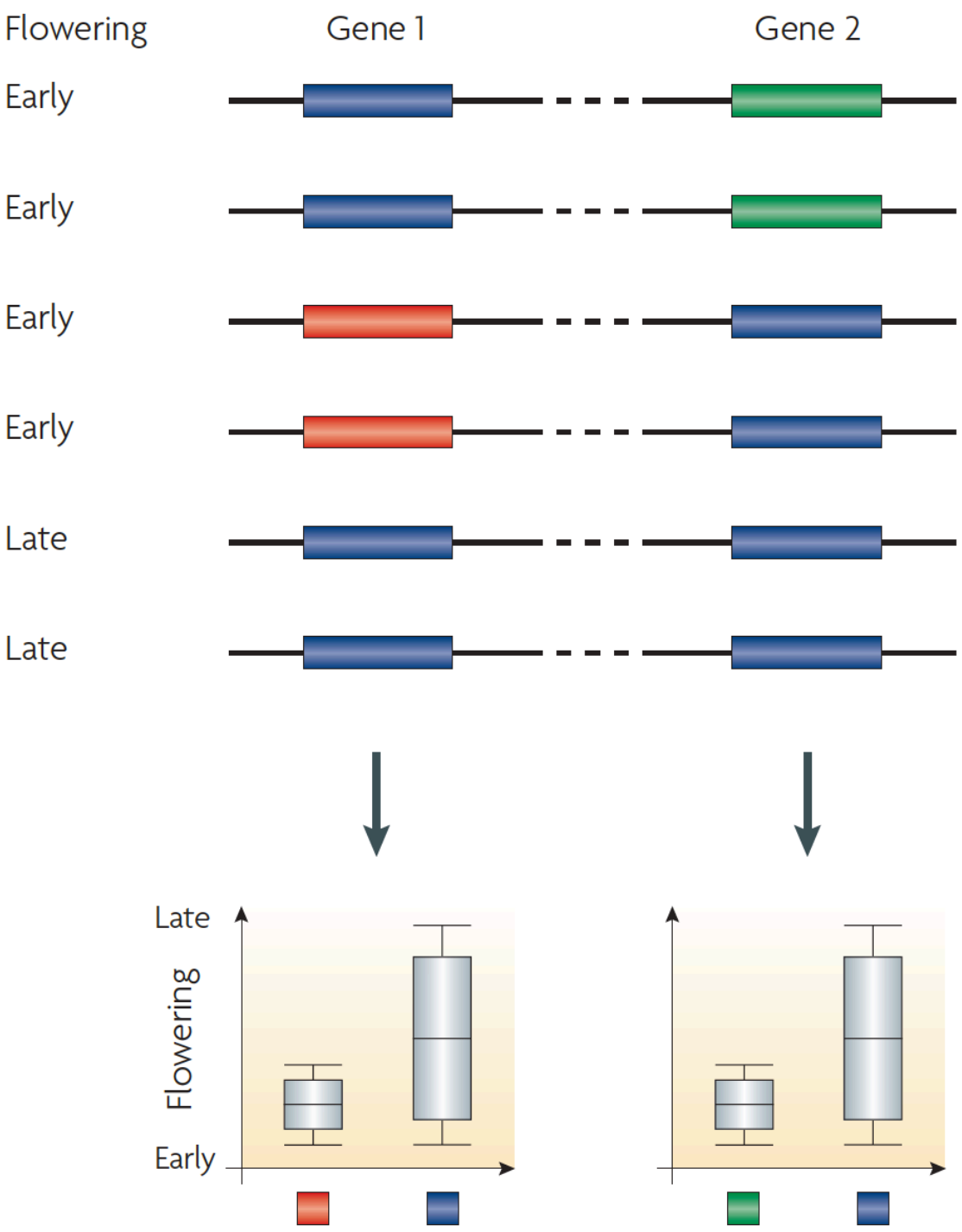
- $P(QQ_l M_1 M_1 M_2 M_2)$=? $P(qq_l M_1 M_1 M_2 M_2)$=?
- $P(QQ_l M_1 M_1 m_2 m_2)$=? $P(qq_l M_1 M_1 m_2 m_2)$=?



# Interval QTL mapping

- Limitations
  - One QTL at a time, overestimate effect of the QTL detected (Xu 2003)
  - One QTL at a time, genetic heterogeneity impede detection of QTL with relative small effects
  - If there is more than one QTL on a chromosome, the test statistic at the position being tested will be affected by all those QTLs, ghost QTL from interval mapping

- Genetic heterogeneity can impede detection of the genes/QTL




(Bergelson and Roux, 2010)

# Composite interval mapping of QTL

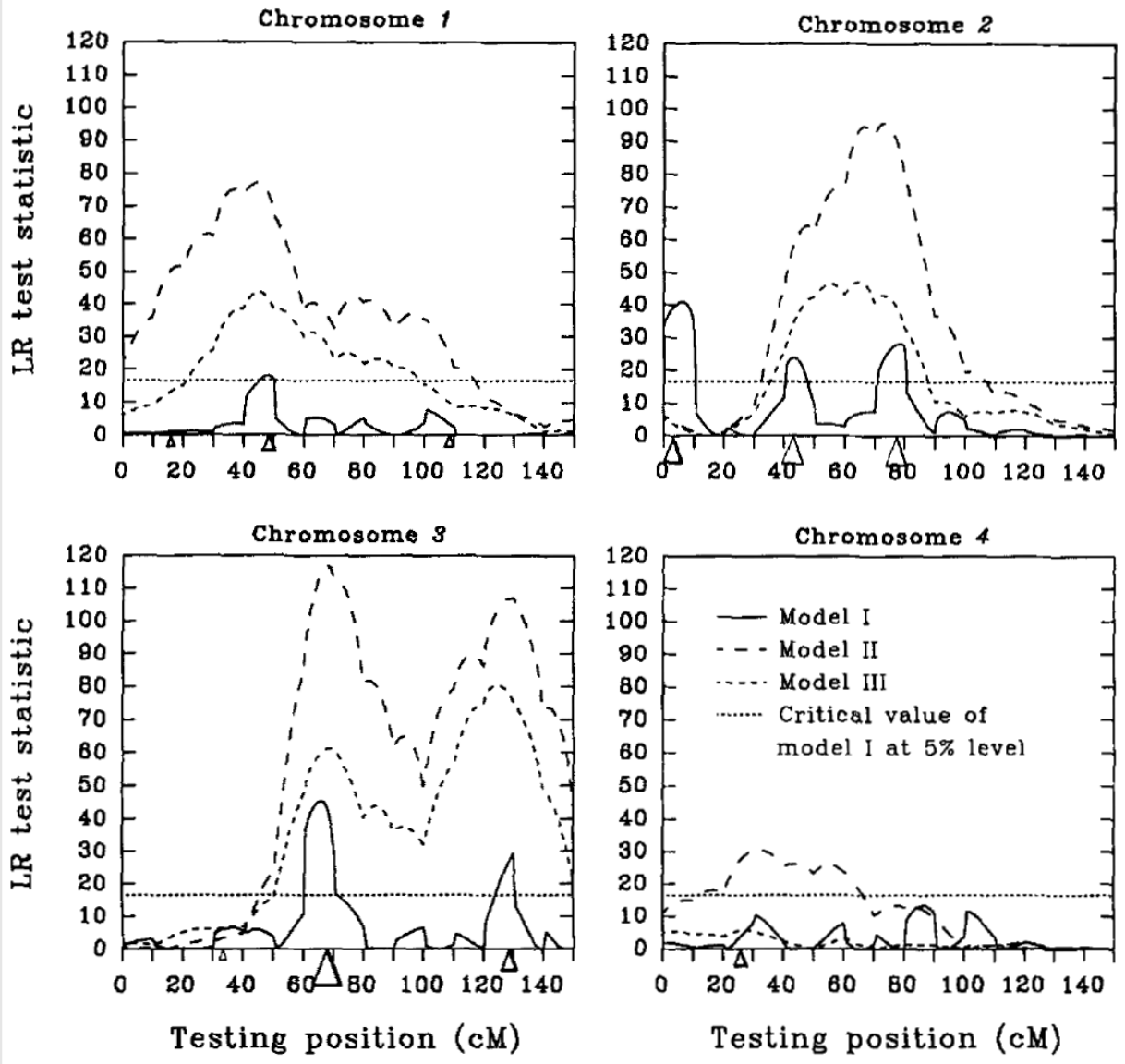
- Joint use of interval mapping and multiple regression
- Location of QTL is estimated by interval mapping
- Effects of other QTL are accounted by multiple regression, with a selected subset of markers as co-factors in the regression analysis

$$y_j = b_0 + b^* x^*_j + \sum_{k \neq i, i+1} b_k x_{jk} + e_j$$

(Zeng, 1994)

# Figure 2. A simulation example of QTL mapping on an hypothetical backcross population

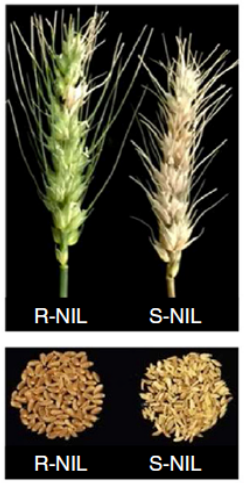
- Model I
  - composite interval mapping
- Model III
  - simple interval mapping




(Zeng, 1994)

# FUSARIUM HEAD BLIGHT in wheat

- FUSARIUM HEAD BLIGHT, caused by Fusarium graminearum is an important disease in wheat producing areas of the USA and Canada
- Epidemics of FHB from 1993 to 1997 resulted in devastating economic losses to the wheat industry of the region, with 1993 estimates alone surpassing one billion dollars (McMullen et al., 1997)
- Fusarium head blight causes both severe yield reduction and decreases grain quality
- In addition, infected grain may contain harmful levels of mycotoxins that prevent its use for human consumption or feed
- Control of FHB has been difficult because of the ubiquitous nature and wide host range of the pathogen and dependence of the disease on unpredictable climatic conditions
- FHB resistance showed low heritability and was highly affected by environments, which limited traditional phenotypic selection




Waldron et al., 1999

# RFLP Mapping of QTL for Fusarium Head Blight Resistance in Wheat

- A population of 112 $F_5$-derived recombinant inbred (RI) lines developed by single seed descent from the spring wheat cross ‘Sumai 3’ (resistant)/’Stoa’ (moderately susceptible)
  - Sumai 3 is a Chinese cultivar known for its Type II resistance to FHB
  - Stoa is a hard red spring cultivar released by North Dakota State University in 1984
- Fusarium head blight phenotypic response of the RI lines and checks was evaluated in two experiments, each with three replications, 1994 and 1995
- As plants reached anthesis, approximately 10 spikes per RI line at the same stage of development in each replicate were inoculated by placing a 10+L droplet of conidial suspension within the glumes of a single spikelet near the center of the head
- Three weeks after inoculation, spikes were scored individually for visual symptoms on a scale of 0 to 100% FHB

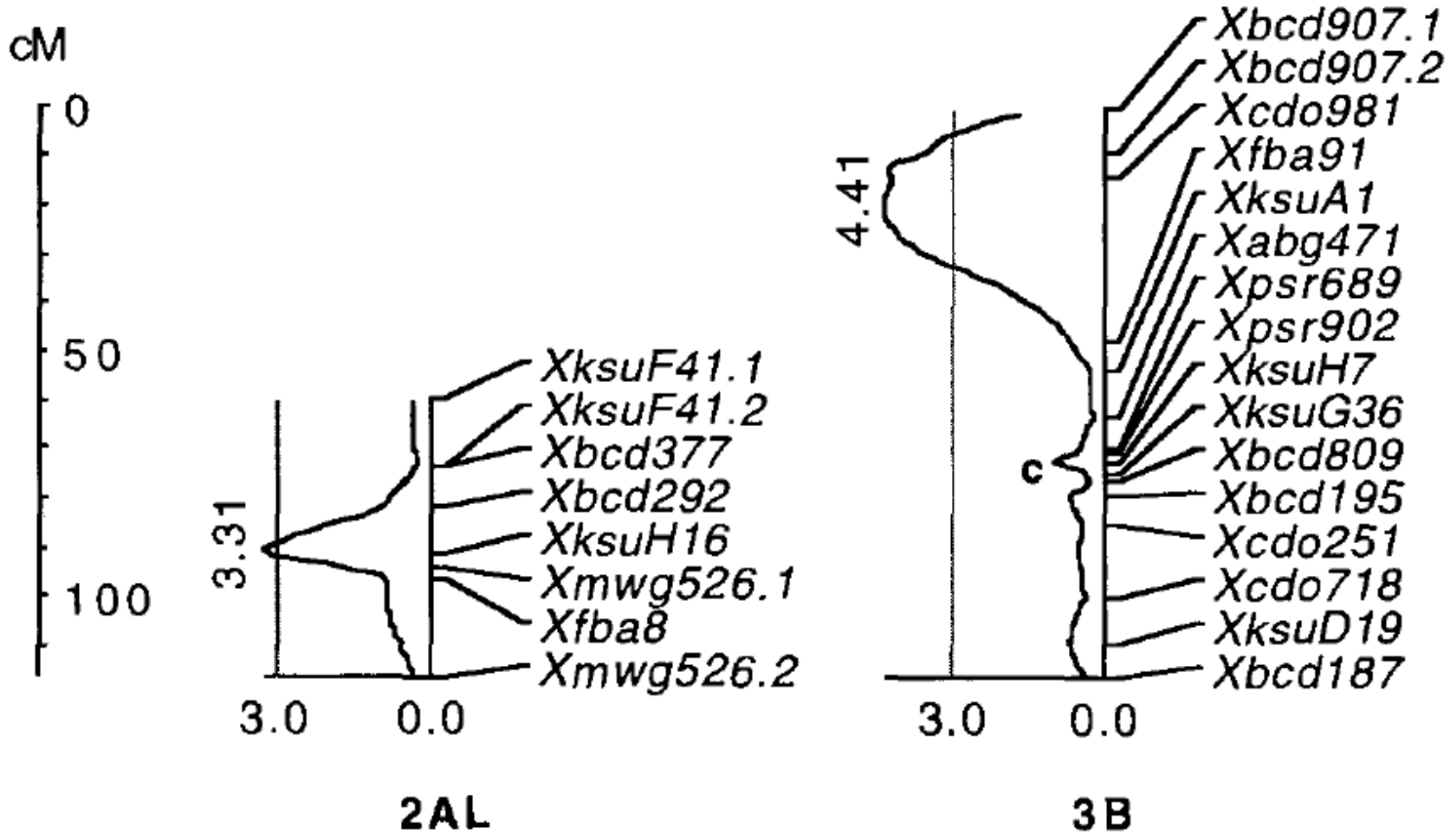
Waldron et al., 1999

# RFLP Mapping of QTL for Fusarium Head Blight Resistance in Wheat

- RFLP linkage maps were constructed using MAPMAKER
- A  total of 292 clones were mapped, yielding 360 loci that formed 38 linkage groups with 46 of the markers unlinked.
- Five genomic regions were significantly ($P < 0.01$ in either Exp. 1,2, or their combined mean) associated with FHB resistance
- The best single marker, Xcdo982, was linked to the QTL on chromosome 3BS. This marker explained 15.4% of the phenotypic variation

Waldron et al., 1999

# RFLP Mapping of QTL for Fusarium Head Blight Resistance in Wheat




Waldron et al., 1999

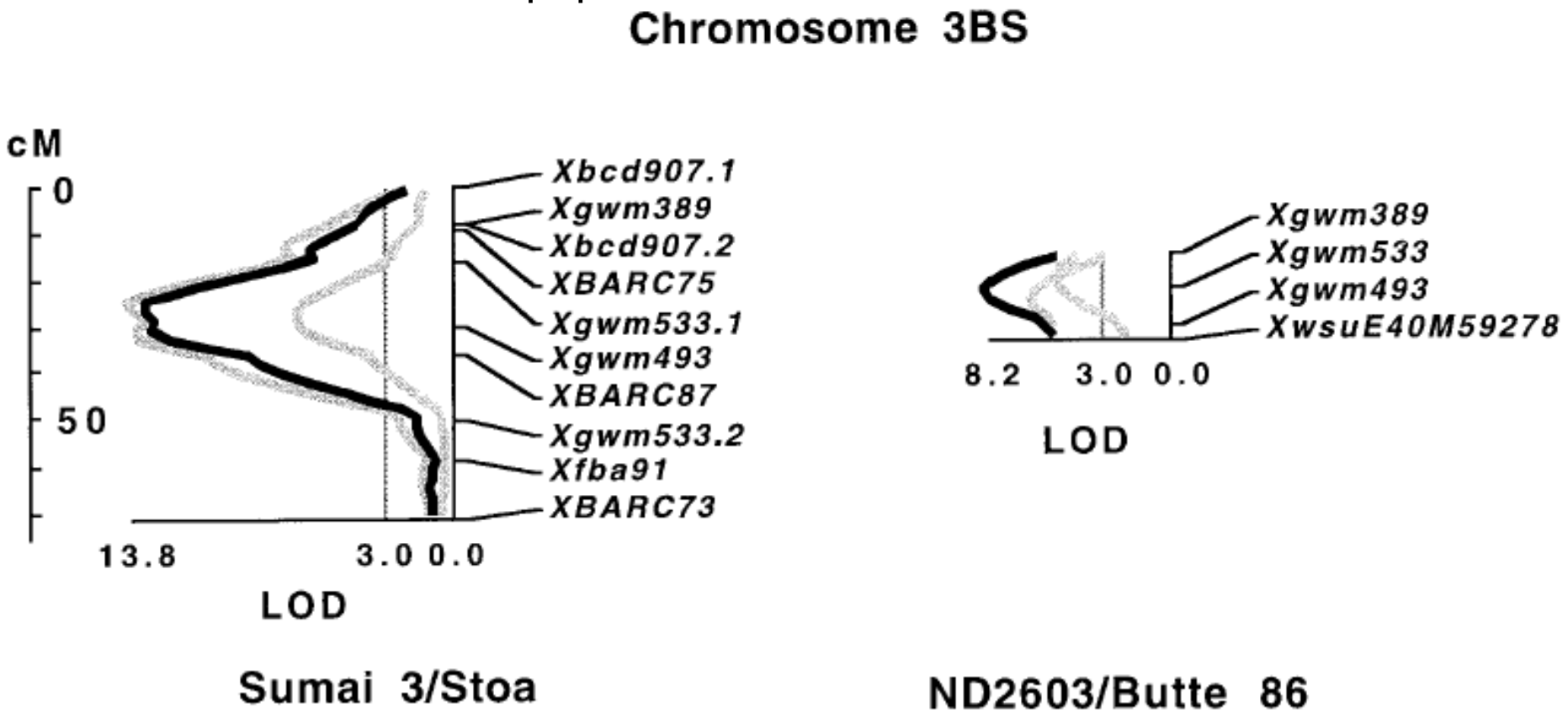
# DNA markers for Fusarium head blight resistance QTLs in two wheat populations

- The objectives of the present research were to verify the FHB QTLs identified in the Sumai 3/Stoa population with another population and obtain more closely linked markers to Qfhs.ndsu-3BS
- A population of 139 F5-derived recombinant inbred lines (RIL) from the cross ND2603 (Sumai 3/Wheaton) (resistant)/ Butte 86 (moderately susceptible) was evaluated
- RFLP mapping in the Sumai 3/Stoa population was described by Waldron et al. (1999). Only those RFLP markers significantly associated ($P<0.05$) with FHB on the Sumai 3/Stoa population were screened for polymorphism and mapped in the ND2603/Butte 86 population.
- Primers for all microsatellites (SSRs) published by Röder et al. (1998) were synthesized and screened for polymorphism among the four parents of these populations

Anderson et al., 2001

# DNA markers for Fusarium head blight resistance QTLs in two wheat populations

- Interval analysis of data for chromosome 3B for Fusarium head blight resistance in the Sumai 3/Stoa and ND2603/Butte 86 recombinant inbred populations




Anderson et al., 2001

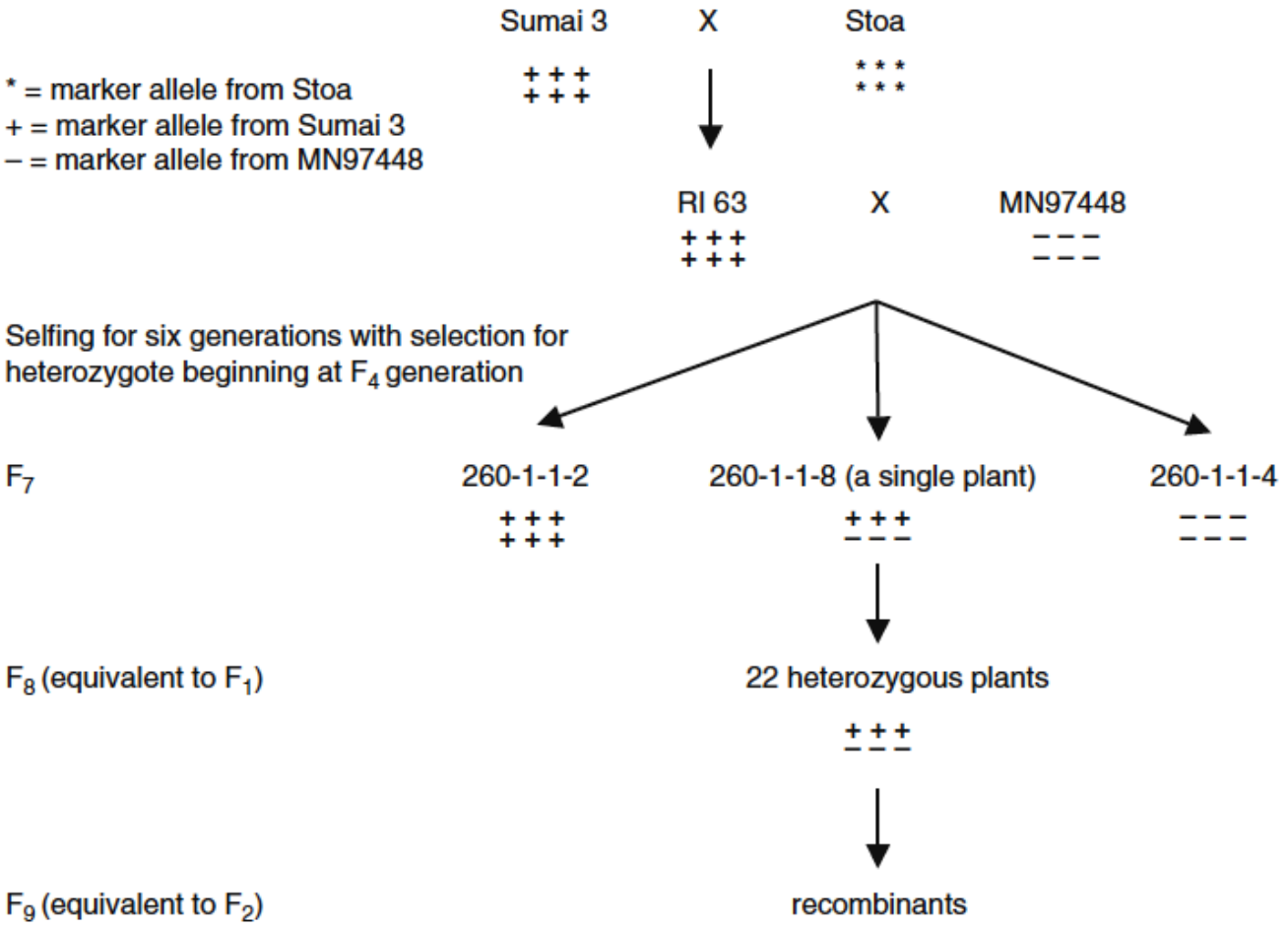
# Fine mapping of the genomic region harboring a major QTL for resistance to FHB in wheat

- This QTL is flanked by two simple sequence repeat (SSR) marker loci, Xgwm533 and Xgwm493, and has been verified in several mapping populations
- The objectives of this study were to construct a fine wheat genetic map of the Qfhs.ndsu-3BS region
- An FHB-resistant recombinant inbred line, RI 63, derived from the cross Sumai 3 (resistant)/Stoa (susceptible) was hybridized with an FHB-susceptible line, MN97448

Liu et al., 2006

# Fine mapping of the genomic region harboring a major QTL for resistance to FHB in wheat

- Three SSR markers, gwm533, BARC133, and gwm493, were used to genotype the parental lines and to select plants heterozygous for all three markers




Liu et al., 2006

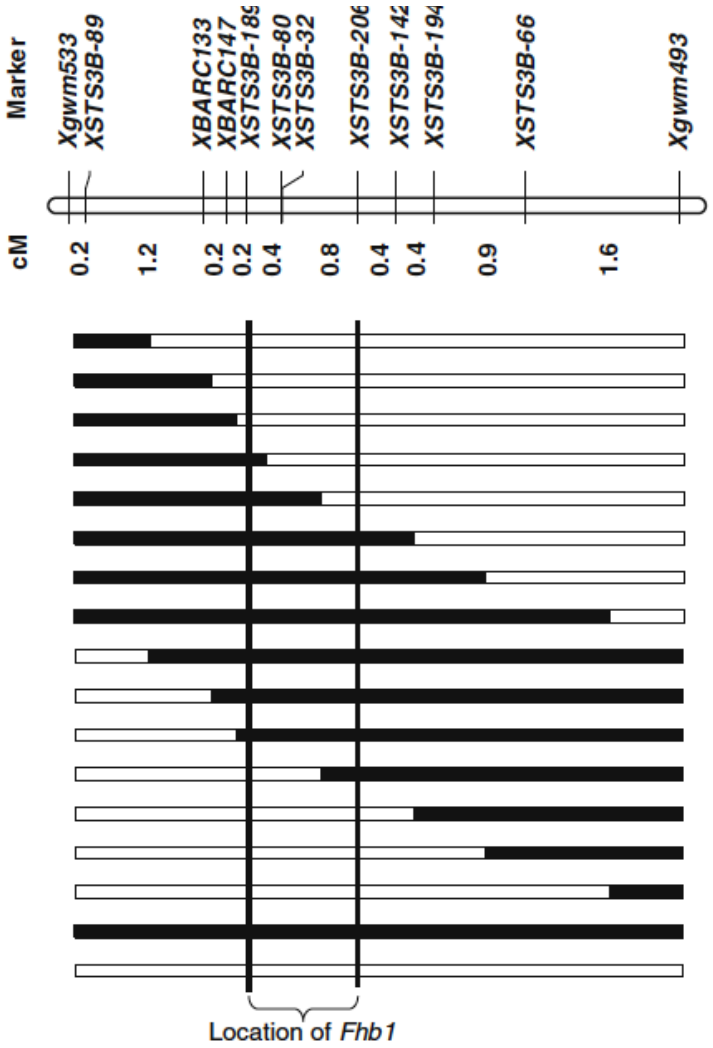
# Fine mapping of the genomic region harboring a major QTL for resistance to FHB in wheat

- The self-pollinated seeds (equivalent to F2) from these 22 heterozygous plants were grown to identify recombinants.
- Two SSR markers flanking Qfhs.ndsu-3BS, gwm533 and gwm493, were used to identify the recombinants
- Among the 3,156 plants (equivalent to F2 plants for the region of interest) screened for recombinants with the two SSR marker loci, Xgwm533 and Xgwm493, 382 recombinants were identified. Nine recombinants were homozygous for both of the two SSR markers
- All 382 recombinants were genotyped with two more SSR markers, BARC133 and BARC147, and eight STS markers

Liu et al., 2006

# Fine mapping of the genomic region harboring a major QTL for resistance to FHB in wheat

- All 382 recombinants were genotyped with two more SSR markers, BARC133 and BARC147, and eight STS markers
- Based on the FHB phenotypes of three HR lines, HR37, HR45, and HR56, Qfhs.ndsu-3BS was placed into a 1.2-cM marker interval flanked by STS3B-189 and STS3B-206

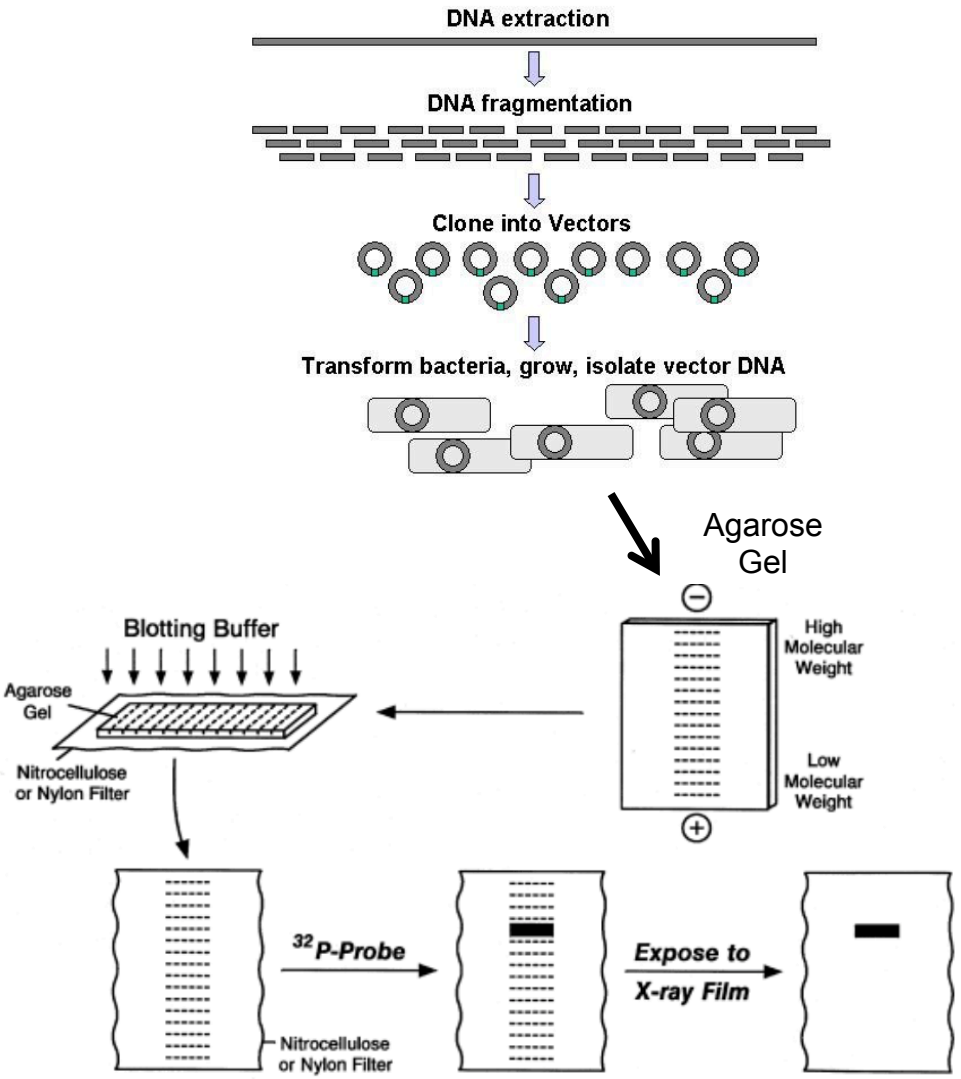



| Line | %FHB UMN | %FHB KSU |
|---|---|---|
| HR4 | 55 (S) | 73 (S) |
| HR41 | 40 (S) | |
| HR46 | 57 (S) | |
| HR45 | 41 (S) | 66 (S) |
| HR56 | 3 (R) | 9 (R) |
| HR68 | 6 (R) | |
| HR63 | 4 (R) | |
| HR74 | 7 (R) | |
| HR11 | 4 (R) | |
| HR20 | 5 (R) | |
| HR26 | 7 (R) | |
| HR37 | 37 (S) | 70 (S) |
| HR77 | 49 (S) | |
| HR78 | 53 (S) | |
| HR85 | 61 (S) | 76 (S) |
| 260-1-1-2 | 5 (R) | 10 (R) |
| 260-1-1-4 | 47 (S) | 64 (S) |
| $LSD_{0.05}$ | 13 | 9 |

Liu et al., 2006

# Toward positional cloning of Fhb1, a major QTL for FHB resistance in wheat

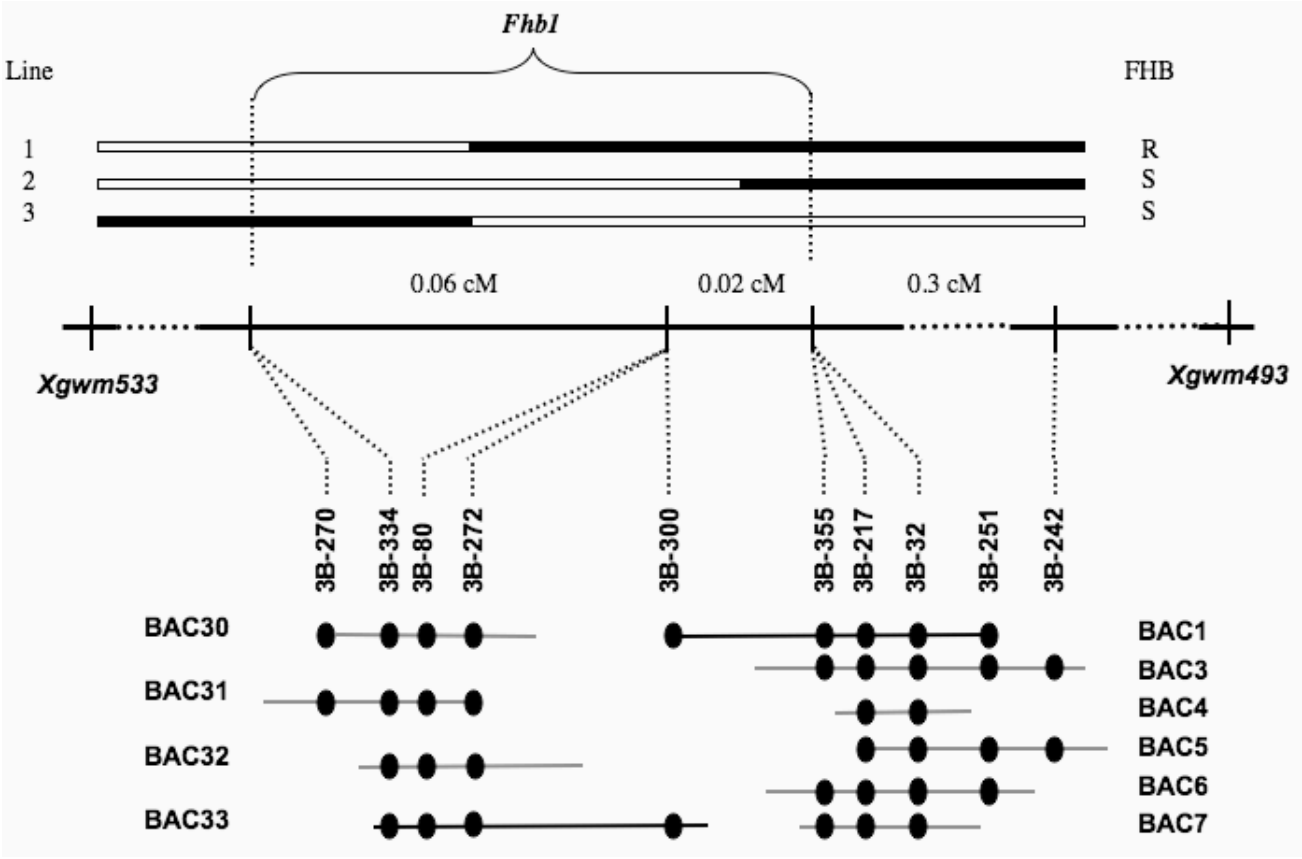
- The PCR products of two STS markers, STS3B-32 and STS3B-80, near Fhb1 (Liu et al. 2006) were used as probes to screen the BAC filters of the chromosome 3B library of 'Chinese Spring'
- New DNA markers were developed from the BAC sequences to further narrow the region spanning *Fhb1* locus
- Polymorphic markers were mapped in the fine mapping population




Liu et al., 2008

# Toward positional cloning of Fhb1, a major QTL for FHB resistance in wheat

- *Fhb1* was placed into the interval flanked by DNA markers 3B-334 and 3B-355
- *Fhb1* was narrowed down to a 261-kb region and seven candidate genes were identified



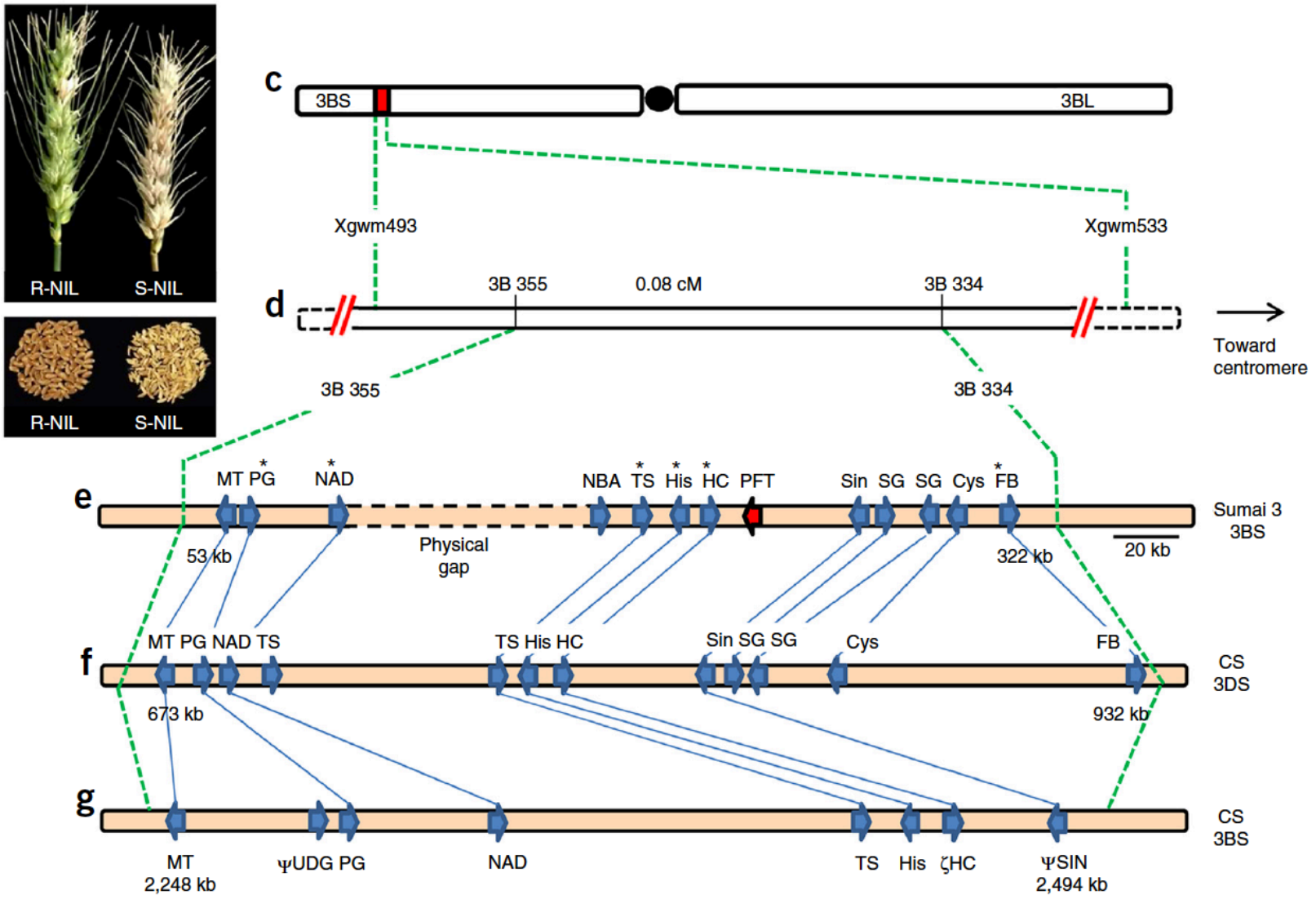

Liu et al., 2008

# Map-based clone of *Fhb1* in wheat

- Used these two flanking markers, other markers derived from CS 3BS, and BAC end-based markers to screen a Sumai 3 BAC library
- Assembled eight overlapping BACs by fingerprinting
- Sequenced four BACs (476D8, 71I24, 383G12 and 572D13) forming two contigs (~350 kb in total) with a physical gap in the *Fhb1* region, and assembled and annotated them
- Thirteen genes were annotated on the Sumai 3 sequence
- Moreover, six genes of the thirteen had been ruled out previously by gene complementation
- Of the seven remaining genes, *PFT* and *NBA* are known to have probable roles in plant defense

Rawat et al., 2016

# Map-based clone of *Fhb1* in wheat




Rawat et al., 2016

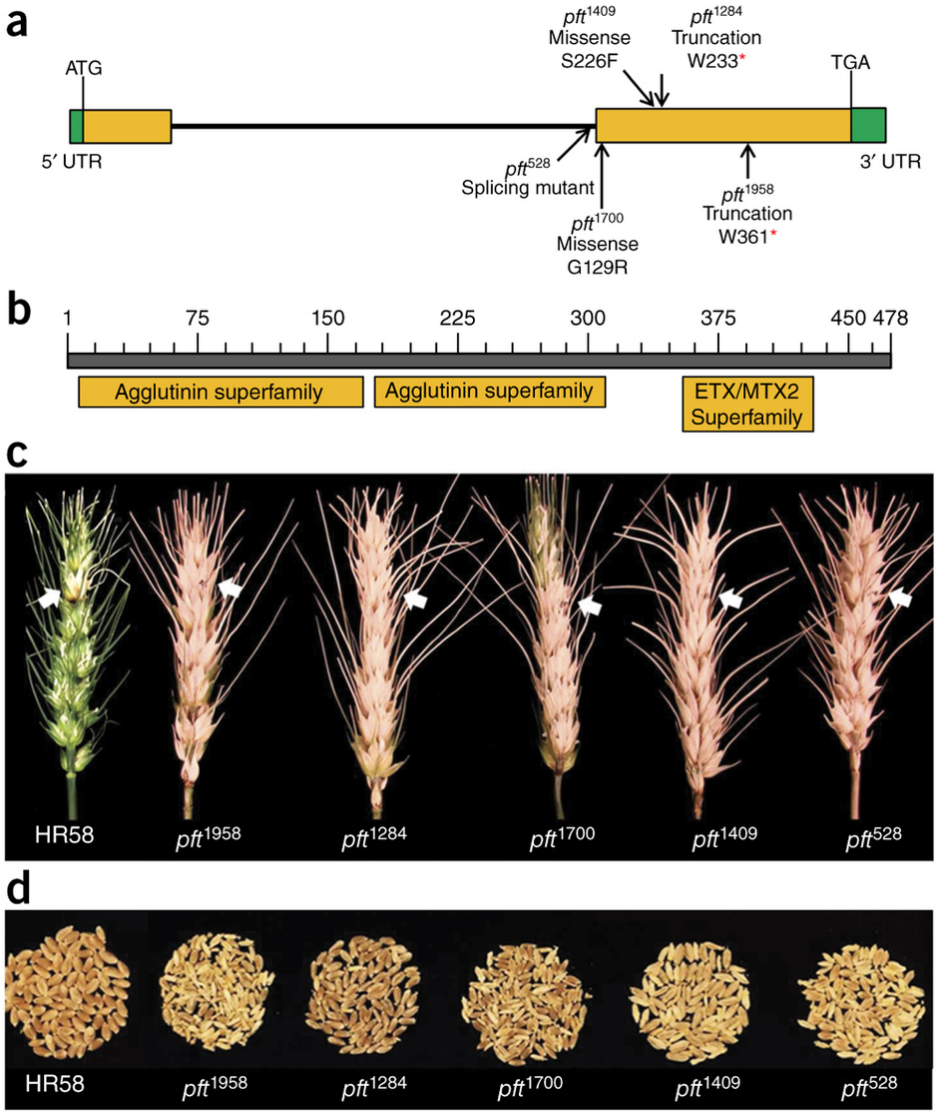
# Map-based clone of *Fhb1* in wheat

- To identify the major genetic determinant of FHB resistance, expression analyses for the annotated genes were performed using quantitative RT-PCR in spikes, inoculated with *Fusarium* macroconidia vs. water, of resistant near-isogenic line (R-NIL) with *Fhb1* and susceptible NIL (S-NIL) lacking the *Fhb1* locus13.
  - The *PFT* and Nb-ARC domain-containing (*NBA*) genes were expressed only in R-NIL and not in S-NIL, whereas the other genes had similar expression patterns in both the NILs
- Furthermore, in an association panel comprising landraces and cultivars known to vary for the presence of *Fhb1*, *NBA* was present in a susceptible haplotype containing cultivars Nanda 2419, Jingzhou 1 and Emai 6.
- Therefore, we excluded *NBA* and considered *PFT* as the putative candidate for *Fhb1*.
- *PFT* is a 3,472-bp gene with two exons generating a 1,437-bp mRNA
- Assessed the candidacy of *PFT* for *Fhb1* using targeting induced local lesions in genome (TILLING) approach, RNA interference (RNAi)-induced gene silencing, association mapping and gene complementation by transformation.

Rawat et al., 2016

# Map-based clone of *Fhb1* in wheat
## --TILLING mutants indicate *PFT* is the causal gene

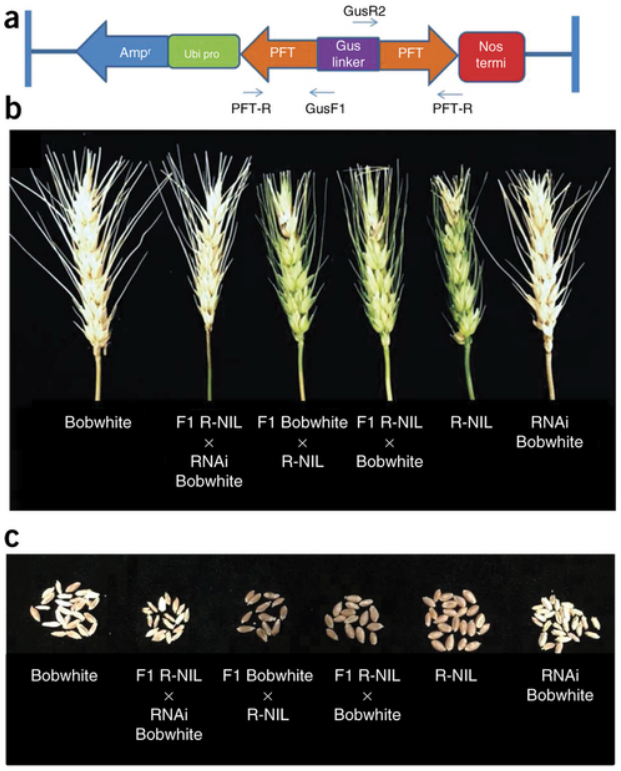
- Five mutations of PFT in homozygous state caused the plants to be susceptible to FHB
- Resequencing of exons of all the other genes in the *Fhb1* region in the susceptible mutants revealed no mutations, including in *NBA*, confirming that the unique mutations in *PFT* were responsible for the loss of resistance to FHB in these plants

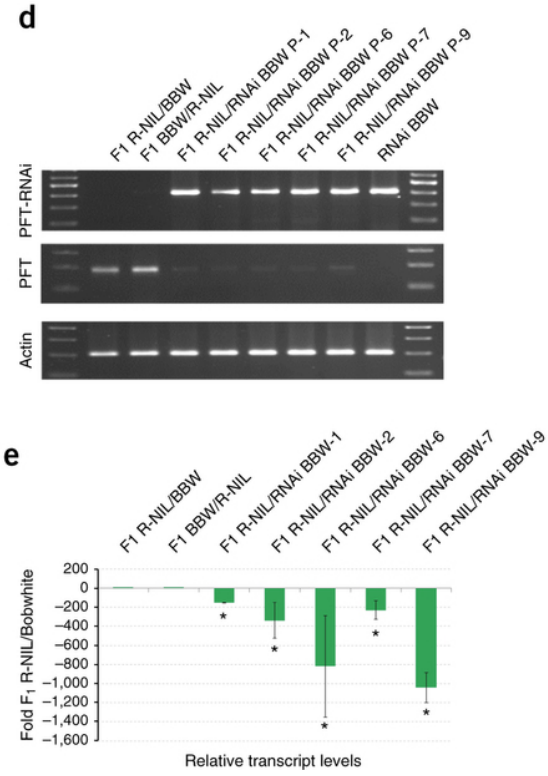



Rawat et al., 2016

# Map-based clone of *Fhb1* in wheat
## --*PFT validation* with RNAi gene silencing

- Introduced an RNAi construct of *PFT* into the wheat cultivar Bobwhite, which is amenable to transformation but does not have *Fhb1*
- Sumai 3 and the R-NIL with *Fhb1* were not amenable to tissue culture and, thus, were not responsive to transformation. F1 plants from reciprocal crosses of the R-NIL and Bobwhite




Rawat et al., 2016

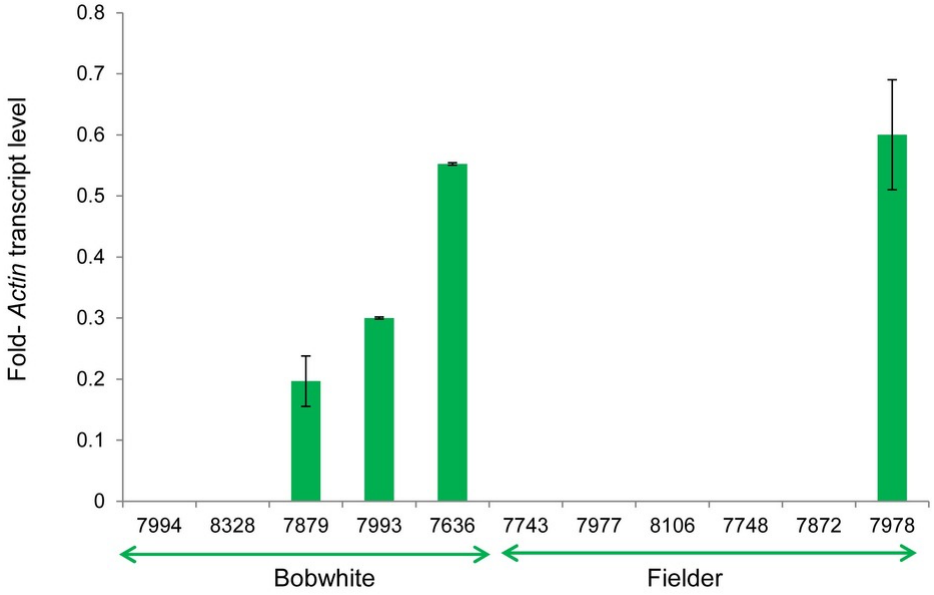
# Map-based clone of *Fhb1* in wheat
## --*Association mapping*

- Sequenced the 13 genes from the *Fhb1* region in an association mapping panel of 40 wheat landraces and cultivars known to differ for their FHB phenotype
- One resistant and four distinct susceptible haplotypes were visible in the region
- Among the 40 wheat genotypes in the panel, all *Fhb1* resistant lines carried the *PFT* gene
- Among the susceptible genotypes, 15 genotypes belonging to haplotypes S1, S2 and S4 were null for *PFT*, whereas three genotypes comprising haplotype S3 carried the same two SNPs each in the *PFT* gene. One SNP was a silent change and the second SNP caused aberrant splicing of mRNA as found in the susceptible TILLING mutant *pft*528, confirming that *PFT* is necessary for FHB resistance

Rawat et al., 2016

# Map-based clone of *Fhb1* in wheat
## --Functional validation of *PFT* with transgenic lines

- Generated transgenic plants expressing the *PFT* gene in hexaploid wheat cultivars Bobwhite and Fielder



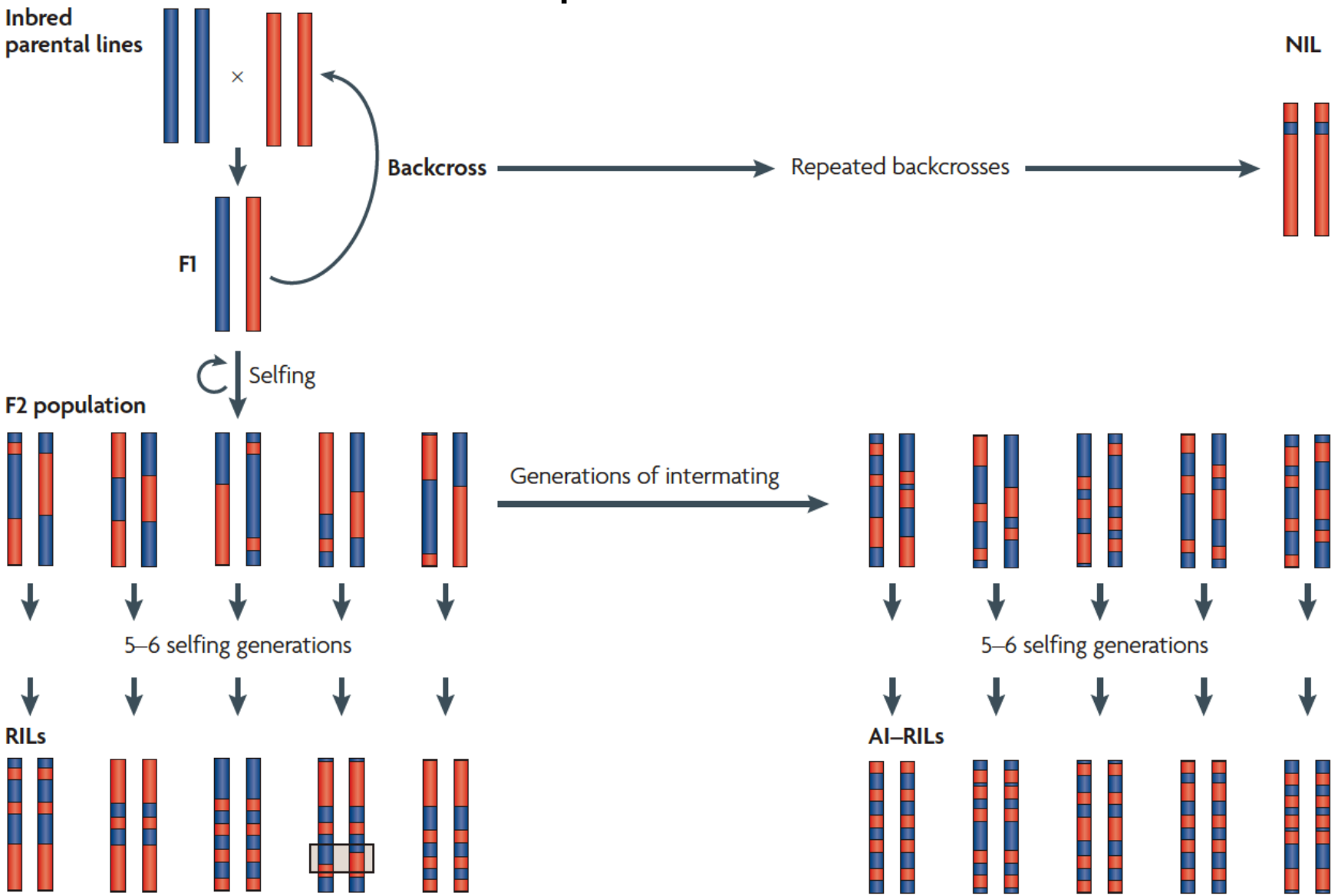


Rawat et al., 2016

# Methods to validate a gene's function

- Association mapping in a population with broad genetic background
- Mutations of the candidate genes
- RNA silencing
- Transgenic line

# Mapping populations derived from two inbred parents



# Population types for QTL mapping in plant

- **$F_2$, $F_3$, and $BC_1$**
  - No enough seeds for replications
  - Fast and early generation QTL detection
- **RILs**
  - Enough seeds for replications and repeated experiment
  - Late generations and time consuming
- **DH**
  - Enough seeds for replication and repeated experiment
  - Faster to get the population, but costly population development and may not work for some parents
- **$F_1$ derived from non-inbred parents**
  - Clones for phenotyping

# Drawbacks of QTL mapping

- QTL mapping has a number of drawbacks; for instance, genetic variation in the mapping population is usually quite restricted with only two parents used to initiate the QTL mapping population
- Because a QTL mapping population usually consists of early-generation crosses (usually F1 or F2 ), the number of recombination events per chromosome is small, which in turn limits the resolution of the genetic map
- In many organisms the generation of mapping populations through controlled crosses is either time-consuming or not even possible, further restricting the utility of QTL mapping
- When a QTL of large effect is identified, tracking down the causal gene is a tedious and time-consuming task
- A single large-effect QTL often breaks down into multiple, closely linked QTLs of smaller, and sometimes opposite, effects on the phenotype

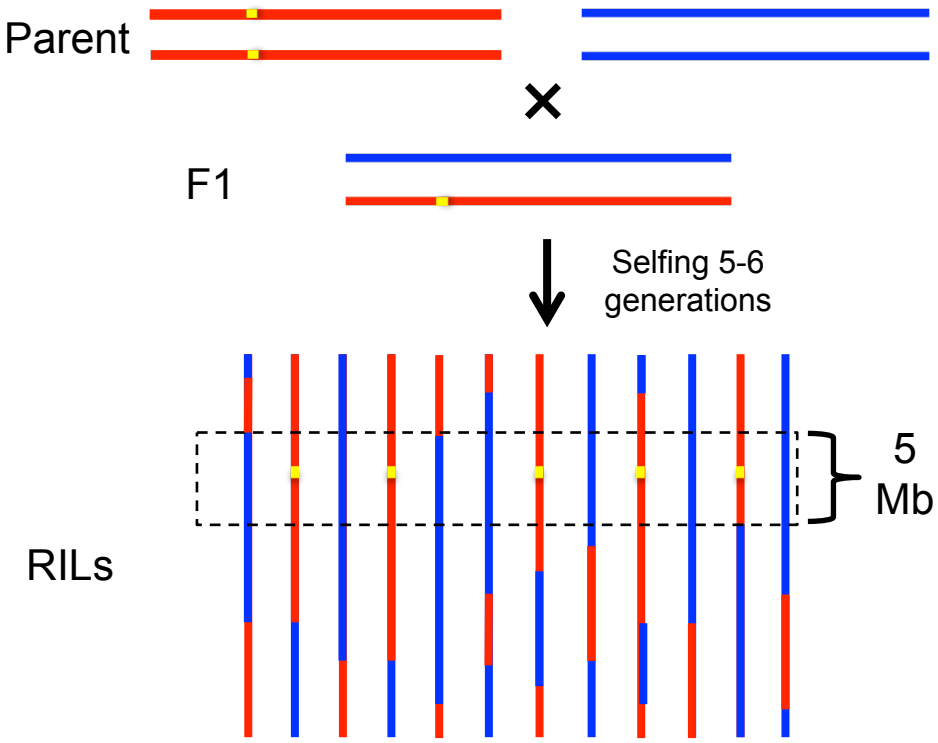
# Genome-wide association mapping (or LD mapping)

- The wealth of molecular markers developed over the last decade has opened up the possibility to directly study statistical associations (linkage disequilibrium, LD) between genetic markers and adaptive traits in natural populations, so-called association genetics (Nordborg & Weigel, 2008).
- Linkage disequilibrium (LD) mapping in plants detects and locates quantitative trait loci (QTL) by the strength of the correlation between a trait and a marker. (Ian mackay and Powell, 2007)
- It offers greater precision in QTL location than family-based linkage analysis (Ian mackay and Powell, 2007)

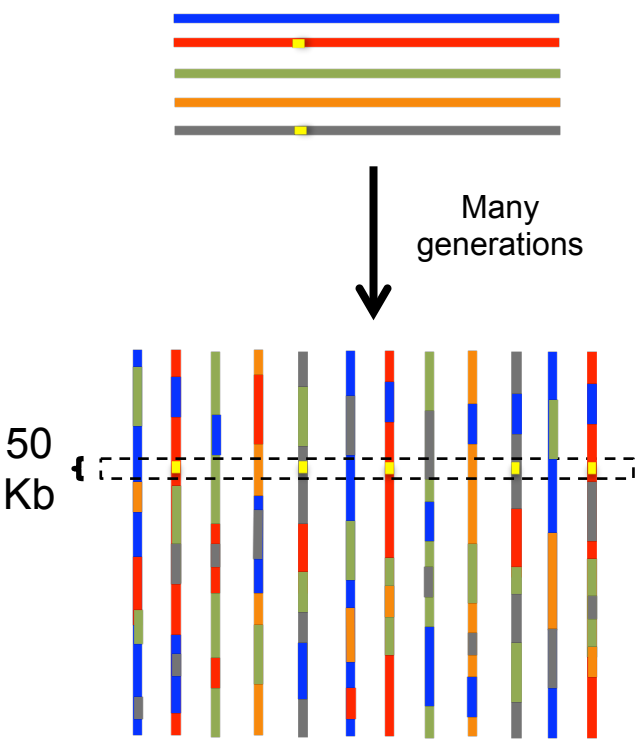
# QTL vs association mapping

- Both QTL and association mapping rely on co-inheritance of functional polymorphism and neighboring DNA markers
- Difference is the mapping population

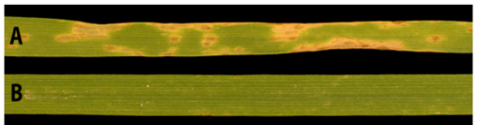
**QTL mapping:** recombination from a few generations and low resolution




**Association mapping:** recombination from many generations and high resolution



# Marker-trait association




R
S

| | | |
|---|---|---|
| Sample 1 | R | ……ACGGT…CGGCA……..TGAT……AAGGG…… |
| Sample 2 | R | ……ACGGT…CGGCA……..TGAT……AAGGG…… |
| Sample 3 | R | ……ACGGT…CGGCA……..TGAA……AAGGC…… |
| Sample 4 | R | ……ACCGT…CGGCA……TGAA……AAGGC…… |
| Sample 5 | S | ……ACCGT…CGGCT……..TGAA……AAGGG…… |
| Sample 6 | S | ……ACCGT…CGGCT……..TGAA……AAGGC…… |
| Sample 7 | S | ……ACCGT…CGGCT……..TGAT……AAGGC…… |
| Sample 8 | S | ……ACCGT…CGGCT……..TGAT……AAGGG…… |
| …… | | …… |

# *Example 1:* GWAS of tan spot resistance in durum wheat

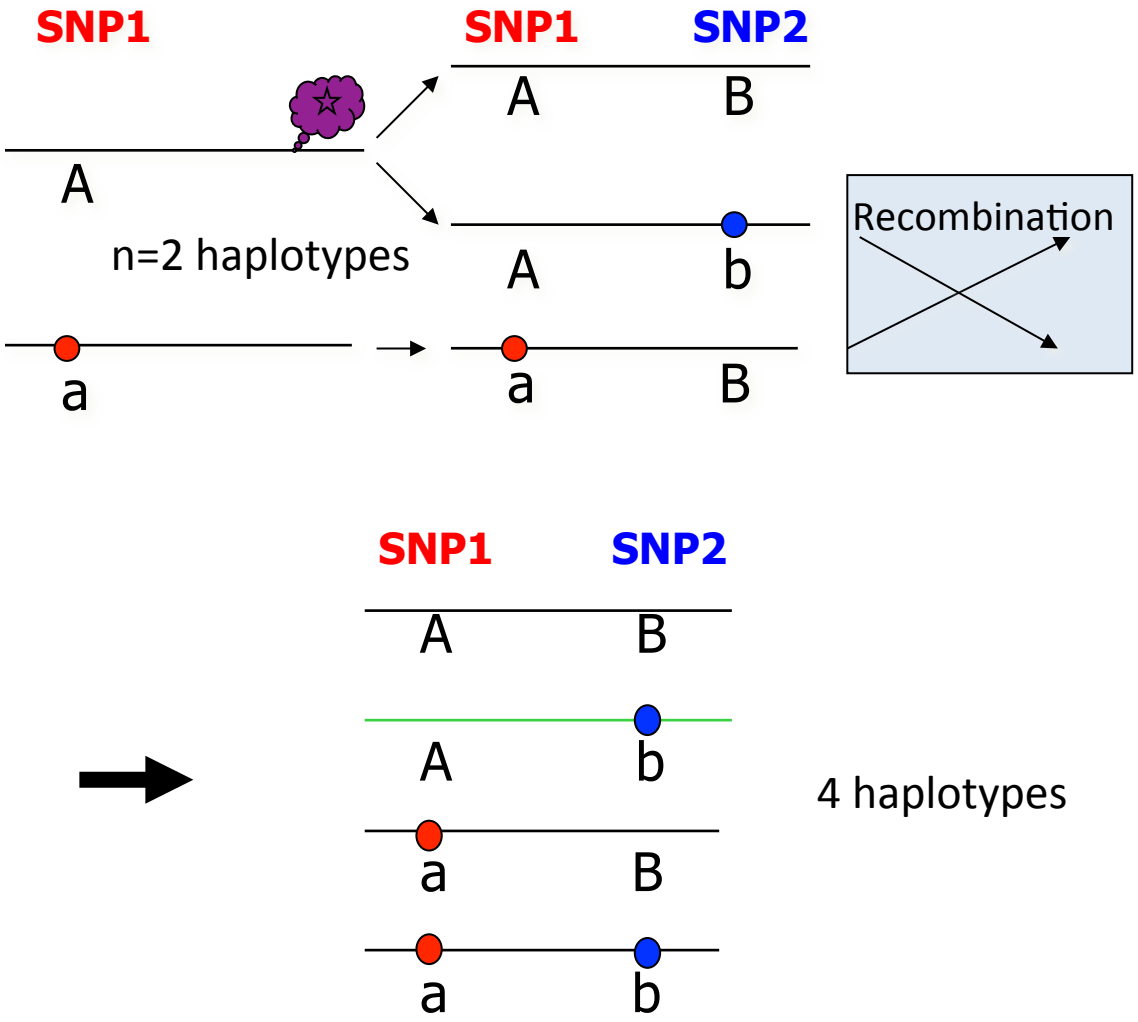
| Ref | Alt | Chr | Position | Est-Ref | Est-Alt | *p-value* |
|---|---|---|---|---|---|---|
| C (296) | T (29) | chr1A | 1207866 | 1.98 | 2.02 | 0.90 |
| A (157) | G (170) | chr1A | 1238074 | 1.95 | 2.05 | 0.49 |
| A (111) | G (208) | chr1A | 1336691 | 2.06 | 1.94 | 0.47 |
| … | … | … | … | … | … | … |
| A (133) | T (198) | chr5B | 545940331 | 2.70 | 1.30 | 1.02E-14 |
| C (119) | G (211) | chr5B | 545943215 | 2.85 | 1.15 | 7.55E-18 |
| G (126) | C (205) | chr5B | 545943463 | 2.75 | 1.25 | 1.19E-15 |
| A (162) | T (169) | chr5B | 546537120 | 1.40 | 2.60 | 8.90E-12 |

# Linkage disequilibrium (LD)

- What is LD?
  - LD is the nonrandom association of alleles at different loci in a given population
- Why estimate LD level?
  - Average LD level determine number of markers needed for genome wide association mapping and power to detect a QTL
- How to estimate LD level?
- What factors affect estimated LD level?
- **Reference paper**
  - Flint-Garcia, S.A., Thornsberry, J.M. and Buckler IV, E.S., 2003. Structure of linkage disequilibrium in plants. *Annual review of plant biology*, *54*(1), pp.357-374.

# Allele, haplotype, and genotype

- Allele: a variant form of a given locus
- Haplotype: specific combination (phasing) of alleles occurring on the same chromosomal segment, AB, Ab, aB, and ab
- Genotype: AB/AB, AB/Ab, ……



# Allele and genotype frequencies at two loci

| | | SNP2 Allele: B | SNP2 Allele: b | |
|---|---|---|---|---|
| SNP1 Allele | A | p(AB/AB) | p(Ab/Ab) | p(A) |
| | a | p(aB/aB) | p(ab/ab) | p(a) |
| | | p(B) | p(b) | |

Genotypes of inbreds: AB/AB, Ab/Ab, aB/aB, and ab/ab

p(AB/AB)+p(aB/aB)= p(B)

p(Ab/Ab)+p(ab/ab)= p(b)

# Linkage equilibrium vs Linkage disequilibrium (LD)

- Linkage equilibrium
  - Observed haplotype frequency p(ab) = p(a) x p(b)
- Linkage disequilibrium (LD)
  - Observed haplotype frequency p(ab)≠ p(a) x p(b)
  - is the nonrandom association of alleles at different loci in a given population
- D, a basic LD statistic, is the difference between observed haplotype and expected haplotype frequencies
  - P(ab) = p(a) p(b) + D
  - *or* P(AB) = p(A) p(B) + D

# LD measure D

- D=p(AB) – p(A) p(B) = 0.86 – 0.9*0.9
- D = 0.05

| | B | b | |
|---|---|---|---|
| A | p(AB/AB)=0.86 | p(Ab/Ab)=0.04 | p(A)=0.9 |
| a | p(aB/aB)=0.04 | p(ab/ab) =0.06 | p(a)=0.1 |
| | p(B)=0.9 | p(b)=0.1 | |

# LD measure D

- D=p(AB) – p(A) p(B) = 0.30 – 0.5*0.5
- D = 0.05
- **D estimation biased by allele frequencies**

| | B | b | |
|---|---|---|---|
| A | p(AB/AB)=0.30 | p(Ab/Ab)=0.20 | p(A)=0.5 |
| a | p(aB/aB)=0.20 | p(ab/ab) =0.30 | p(a)=0.5 |
| | p(B)=0.5 | p(b)=0.5 | |

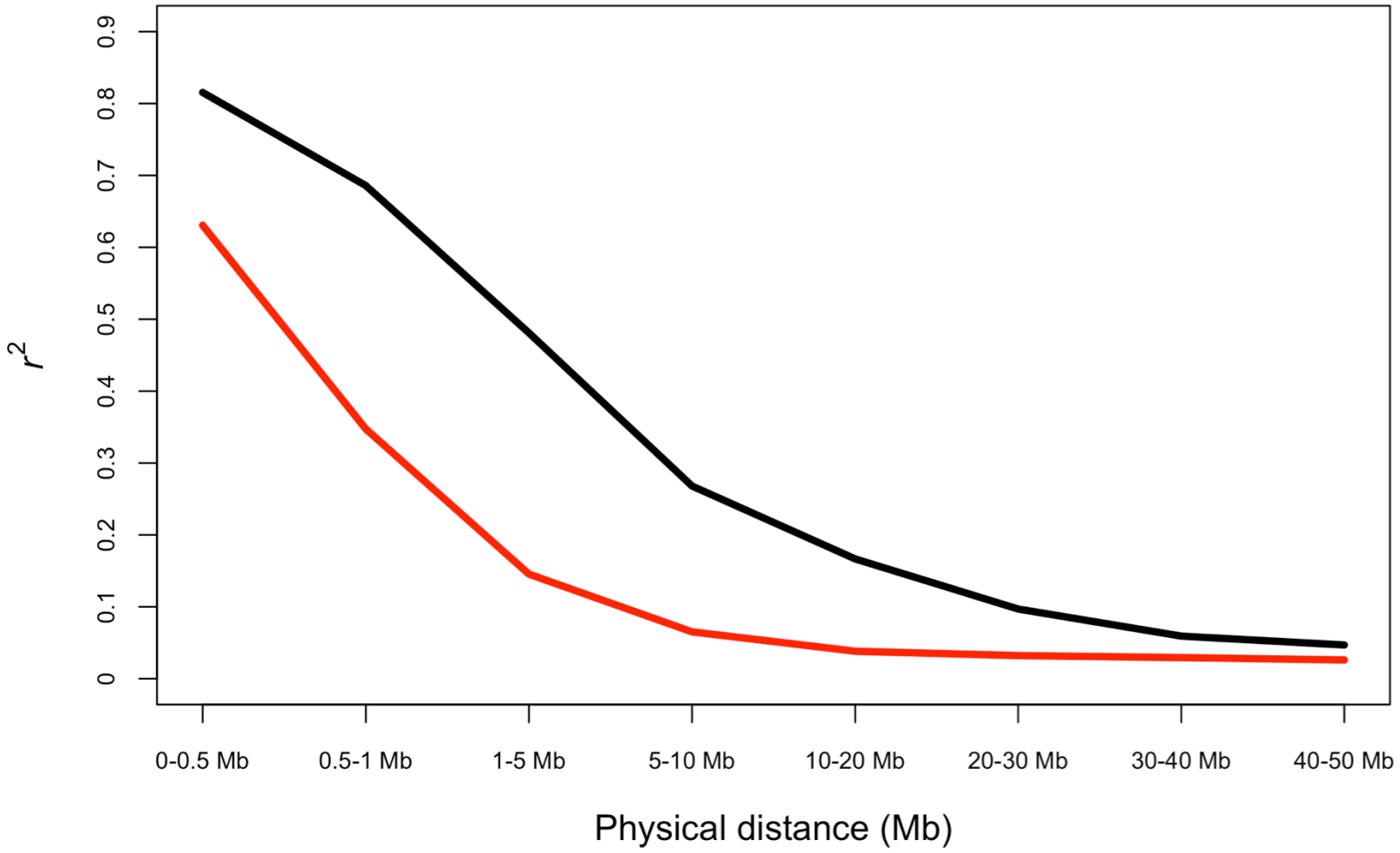
# LD measure D' and $r^2$

- D', Standardize D by rescaling to a proportion of its maximal value for the given allele frequencies
  - $D' = D / D_{max}$
  - $D_{max} = \min (p(A)p(b), p(a)p(B))$ $D > 0$
  - $D_{max} = \max (-p(A)p(B), -p(a)p(b))$ $D < 0$

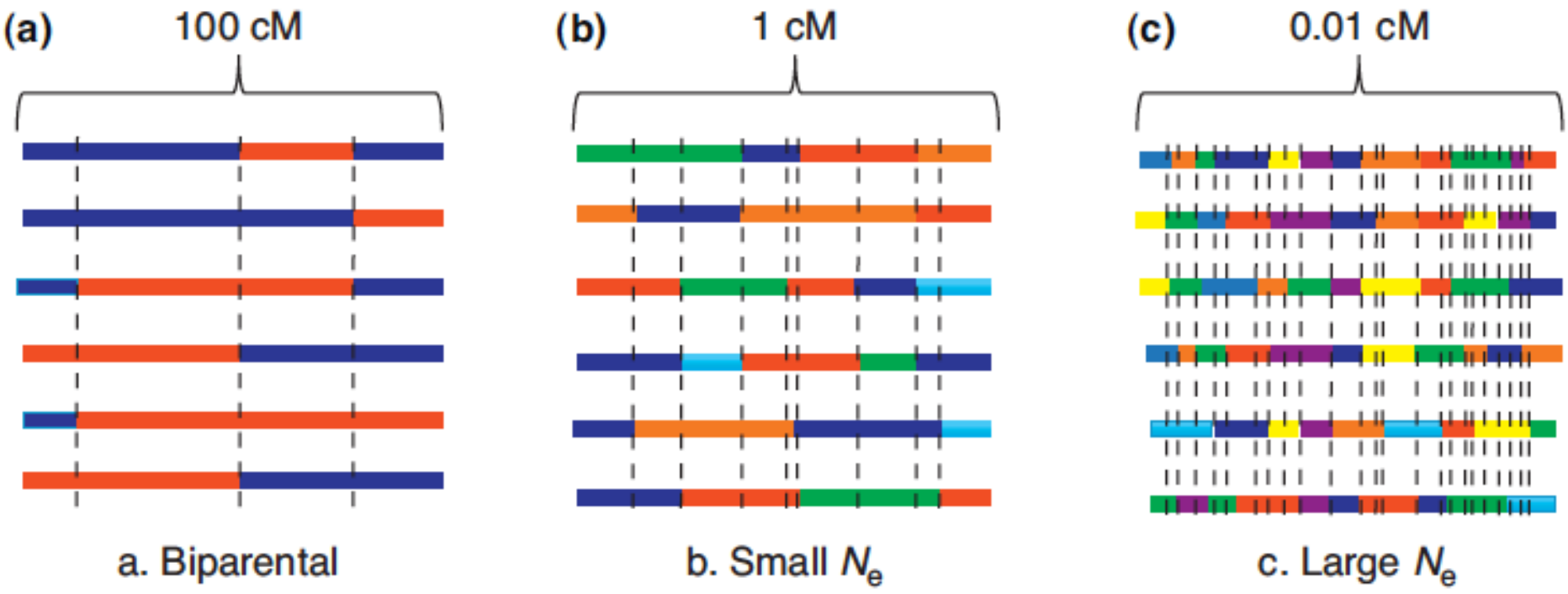
- $r^2$, correlation coefficient between pairs of loci

$$r^2 = \frac{D^2}{p_A p_a p_B p_b}$$

# Distribution of the pairwise linkage disequilibrium measure $r^2$ depending on the physical distance between SNPs

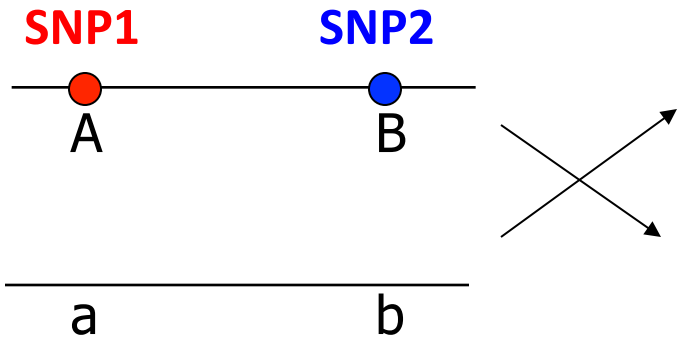


# LD level among different populations



# Linkage and LD

- Linkage: the tendency of DNA sequence that are close together on a chromosome to be inherited together
  - Two genetic markers that are physically near to each other are unlikely to be separated during chromosomal crossover
  - Markers on different chromosomes are perfectly unlinked
- Linkage disequilibrium (LD)
  - is the nonrandom association of alleles at different loci in a given population
  - not necessary on the same chromosome
- What factors affect LD?




Probability Recombination occurs = θ

Probability Recombination *<u>does not</u>* occur = 1-θ

Initial LD between SNP1 - SNP2: $D_0$

After 1 generation

Preservation of LD:

$D_1 = D_0(1-\theta)$

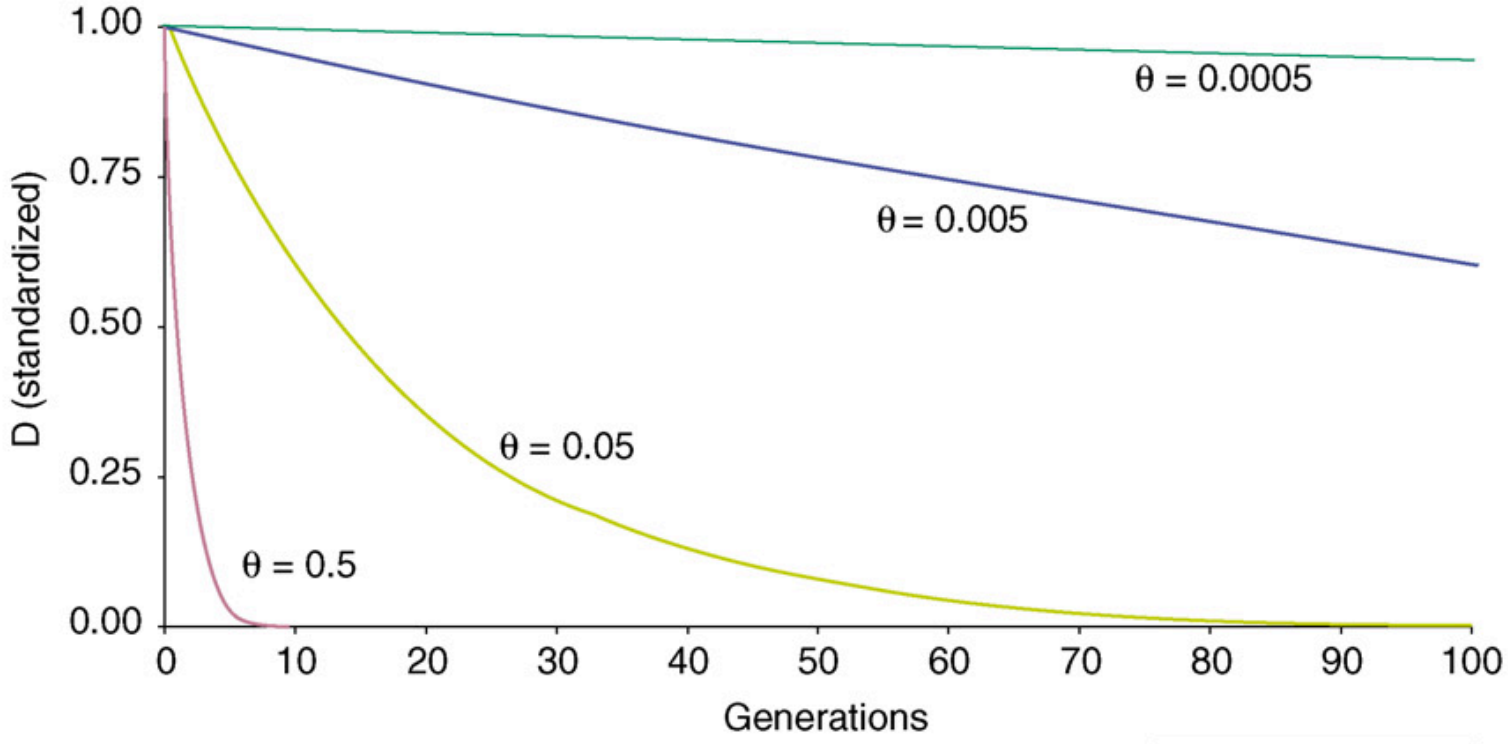
After t generations:

$D_t = D_0 (1- \theta)^t$

# Figure 1. Decay of linkage disequilibrium with time for four different recombination fractions

- For unlinked loci, u = 0.5 and LD decays rapidly within a small number of generations
- For closely linked loci, the decay in LD is extremely slow.

Abbreviation: D = coefficient of linkage disequilibrium




Mackay and Powell, 2007

TRENDS in Plant Science

# What factors affect LD level?

- Linkage
- Recombination
- **Migration (admixture)**
- Mutation
- Population size (genetic drift)
- Selection
- Epistatic interaction

# Admixture causes LD between unlinked markers

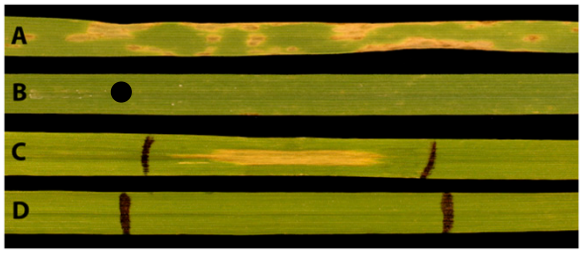
- Assume the frequencies of alleles A and B are 0.8 in pop_1 and 0.2 in pop_2,
  - Frequency of the AB/AB is $0.8^2$=0.64 in population 1
  - Frequency of the AB/AB is $0.2^2$=0.04 in population 2
- The two population are mixed in equal proportions
  - Expected frequency of AB/AB is $0.5^2$=0.25 if A and B are unlinked
  - Observed frequency of AB/AB is 0.34
- Spurious associations due to population structure
  - If A is a QTL, B will be falsely identified as a significant marker, called spurious association

# Population structure

- **Population stratification** (or **population structure**) is the presence of a systematic difference in allele frequencies between subpopulations in a population, possibly due to different ancestry, especially in the context of association studies

- Cause
  - The basic cause of population stratification is nonrandom mating between groups, often due to their physical separation

# Marker-trait association

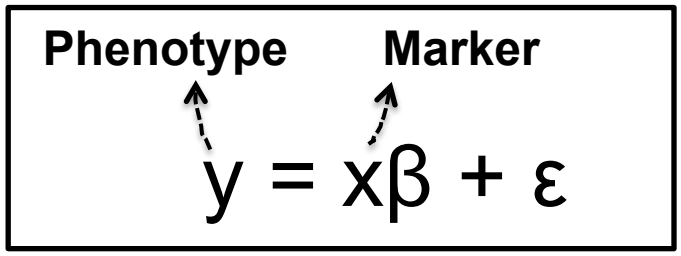



R
S

| | | |
|---|---|---|
| Sample 1 | **R** | ……ACGGT…CGGCA……..TGAT……AAGGG…… |
| Sample 2 | **R** | ……ACGGT…CGGCA……..TGAT……AAGGG…… |
| Sample 3 | **R** | ……ACGGT…CGGCA……..TGAA……AAGGC…… |
| Sample 4 | **R** | ……ACCGT…CGGCA……TGAA……AAGGC…… |
| Sample 5 | **S** | ……ACCGT…CGGCT……..TGAA……AAGGG…… |
| Sample 6 | **S** | ……ACCGT…CGGCT……..TGAA……AAGGC…… |
| Sample 7 | **S** | ……ACCGT…CGGCT……..TGAT……AAGGC…… |
| Sample 8 | **S** | ……ACCGT…CGGCT……..TGAT……AAGGG…… |
| …… | | ……..…… |

# *Example 1:* GWAS of tan spot resistance in durum wheat

- 371 durum wheat landraces
- Tan spot resistance
  - ToxA, scored as 0 (R) and 3 (S)
  - Pti2, scored as 1, 2, 3, 4, and 5 from R to S
- ~60,000 SNP markers
- Association analysis



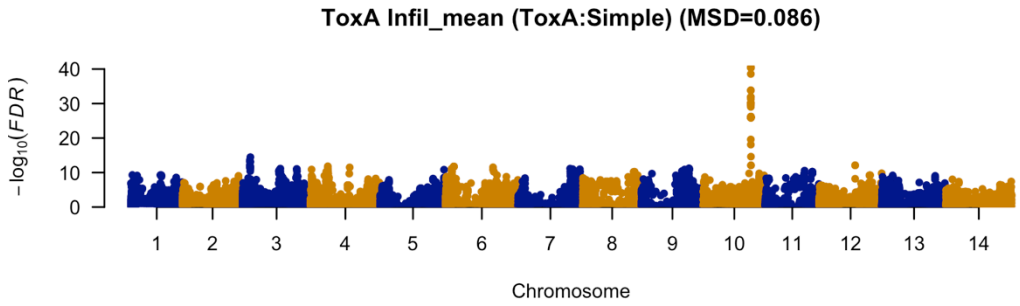
# Statistical Models for single marker and trait association

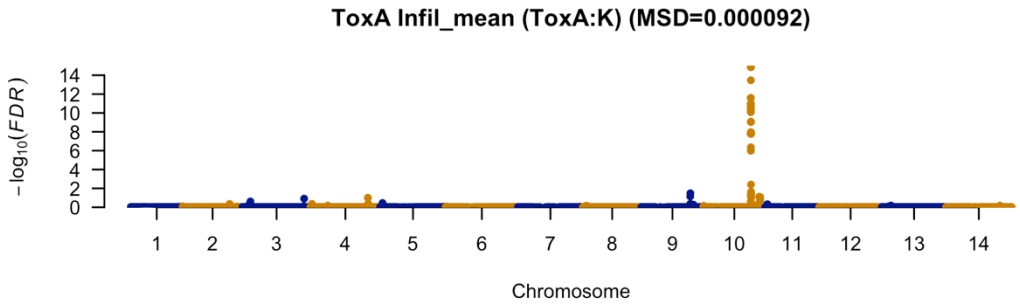
- Generalized linear model
  1. Simple model $y = x\beta + \varepsilon$ $\varepsilon \propto N(0, \sigma_\varepsilon^2 I)$
  2. P or Q structure $y = x\beta + P + \varepsilon$ $\varepsilon \propto N(0, \sigma_\varepsilon^2 I)$
- Linear mixed model
  3. Kinship (K) $y = x\beta + (g) + \varepsilon$ $\varepsilon \propto N(0, \sigma_\varepsilon^2 I)$, $g \propto N(0, \sigma_g^2 K)$
  4. PK $y = x\beta + P + (g) + \varepsilon$ $\varepsilon \propto N(0, \sigma_\varepsilon^2 I)$, $g \propto N(0, \sigma_g^2 K)$
  5. CMLM (compression of kinship)

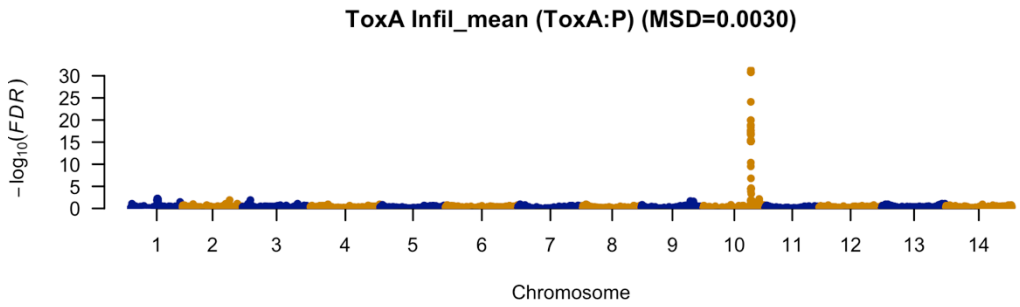
# *Example 1:* GWAS of tan spot resistance in durum wheat

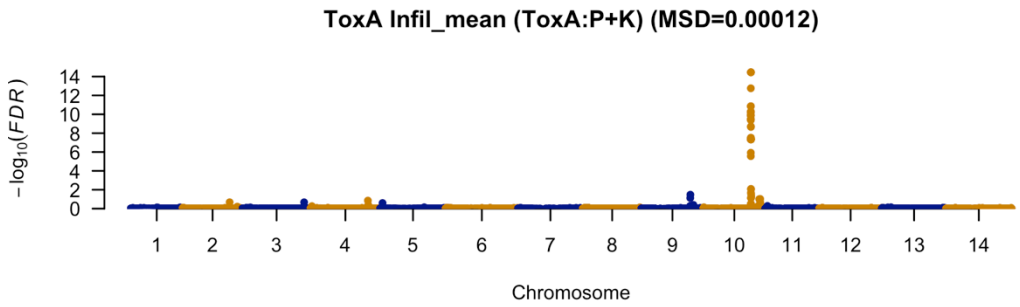
| Ref | Alt | Chr | Position | Est-Ref | Est-Alt | *p-value* |
|---|---|---|---|---|---|---|
| C (296) | T (29) | chr1A | 1207866 | 1.98 | 2.02 | 0.90 |
| A (157) | G (170) | chr1A | 1238074 | 1.95 | 2.05 | 0.49 |
| A (111) | G (208) | chr1A | 1336691 | 2.06 | 1.94 | 0.47 |
| … | … | … | … | … | … | … |
| A (133) | T (198) | chr5B | 545940331 | 2.70 | 1.30 | 1.02E-14 |
| C (119) | G (211) | chr5B | 545943215 | 2.85 | 1.15 | 7.55E-18 |
| G (126) | C (205) | chr5B | 545943463 | 2.75 | 1.25 | 1.19E-15 |
| A (162) | T (169) | chr5B | 546537120 | 1.40 | 2.60 | 8.90E-12 |

- Manhattan plots of ToxA resistance in durum wheat



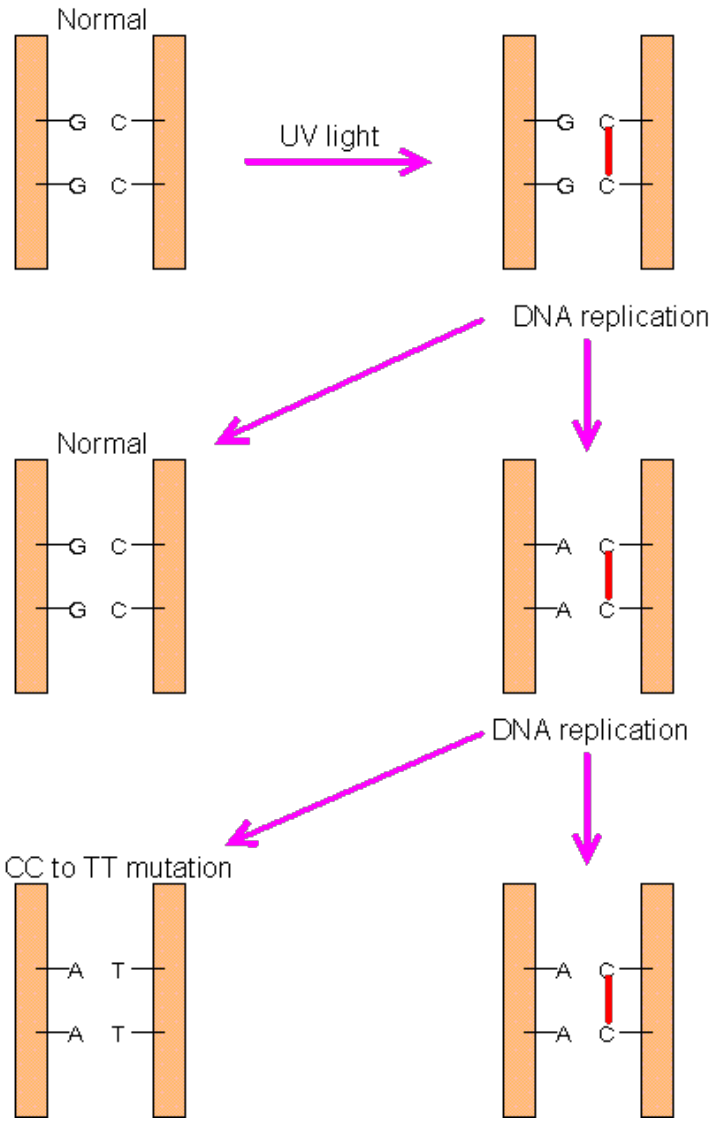



# Nest-generation sequencing to isolate mutant genes from forward genetic screens

- Whole genome sequence yeast radiation-induced mutants
- MutSeq, rice EMS-induced mutants
- MutRenSeq, wheat EMS-induced mutants

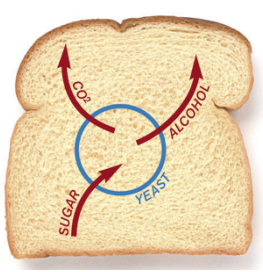
# Ultraviolet light (UV) radiation induced mutations

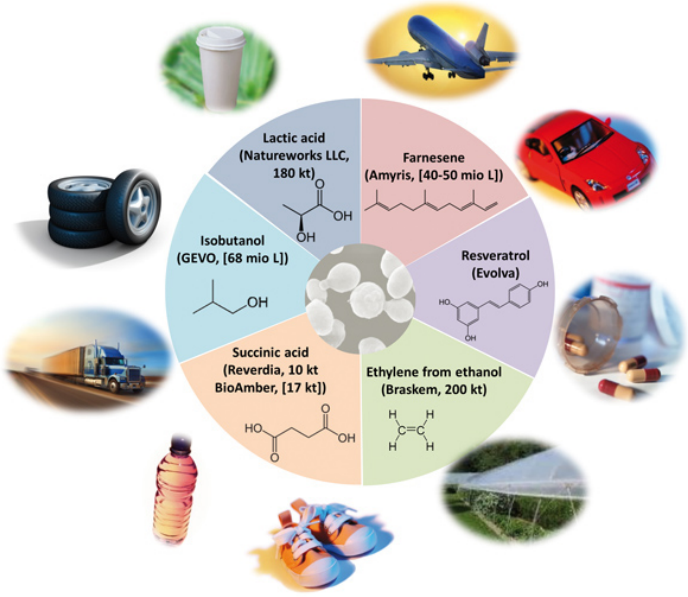
- UV, non-ionizing radiation, induces pyrimidine dimers mutation
  - Causes two consecutive pyrimidine bases on one strand to bind together
  - E.g., leads to a CC to TT mutation
- Ultraviolet radiation from the sun induced pyrimidine dimers is a primary cause of skin cancer in human



# *Example:* UV induced mutations in Yeast -- Modified cell factories

- Yeast is a widely used cell factory for production of
  - Food and beverages
    - Fermentation, converts sugar to carbon dioxide ($CO_2$) and alcohol
    - Beer, bread, yogurt, etc.
  - Pharmaceuticals
    - Antibiotics, hormones, and anti-cancer drugs
    - 20% biopharmaceuticals produced in yeast including insulin, vaccines, etc.
    - 300 biopharmaceuticals have sales over $100 billion
  - Fuel and chemicals
    - Bioethanol, citric acid, etc.
    - Production is $3000 billion in industry

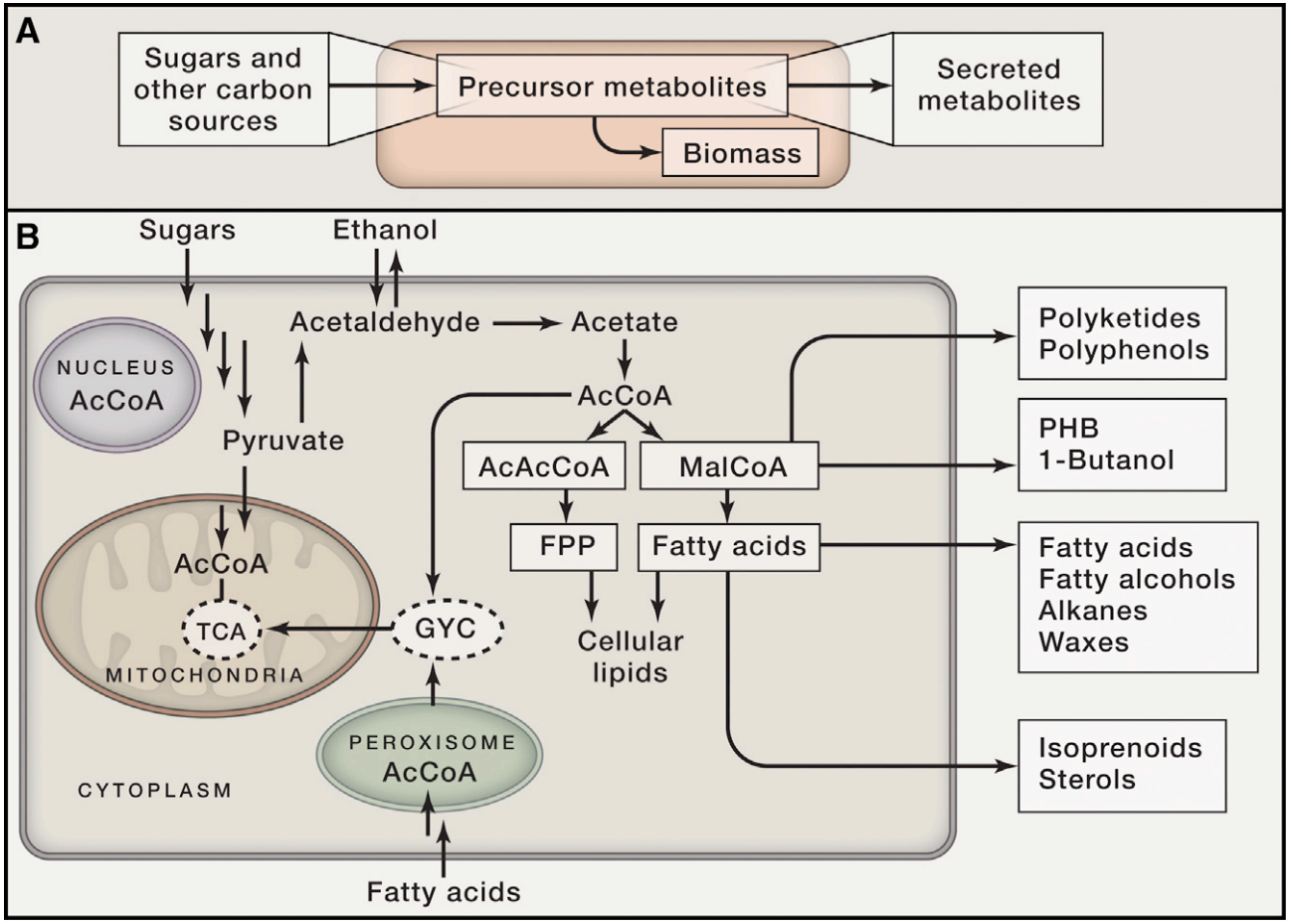



**Jens Nielsen 2014; Nielsen and Keasling 2016**

# *Example:* UV induced mutations in Yeast -- Modified cell factories

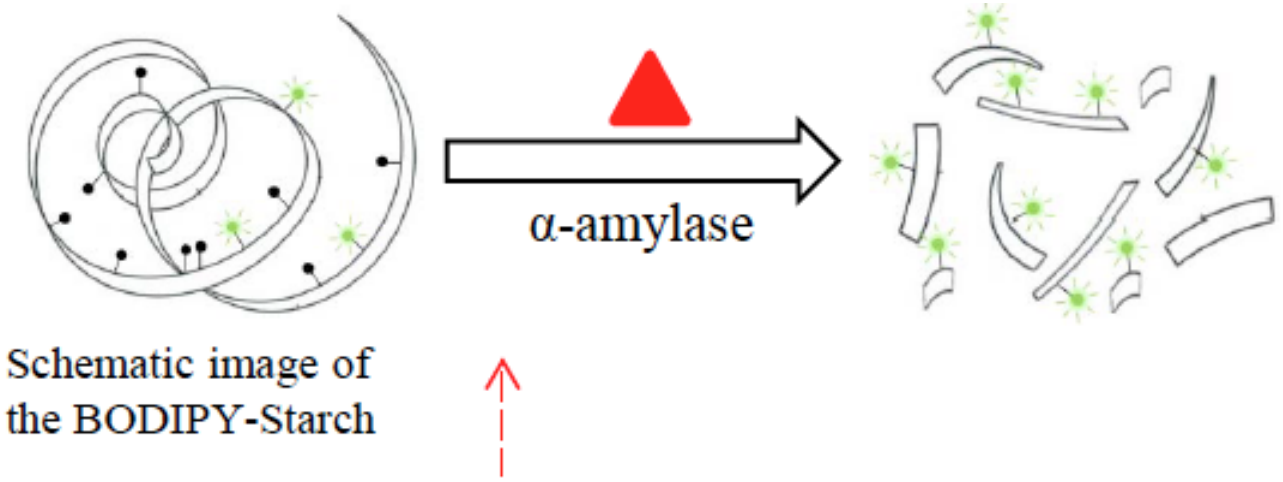
- All carbon sources are converted to 12 precursor metabolites that are used for biosynthesis of all secreted metabolites
- Understand pathways and genes to make yeast cells into efficient factories




Nielsen and Keasling 2016

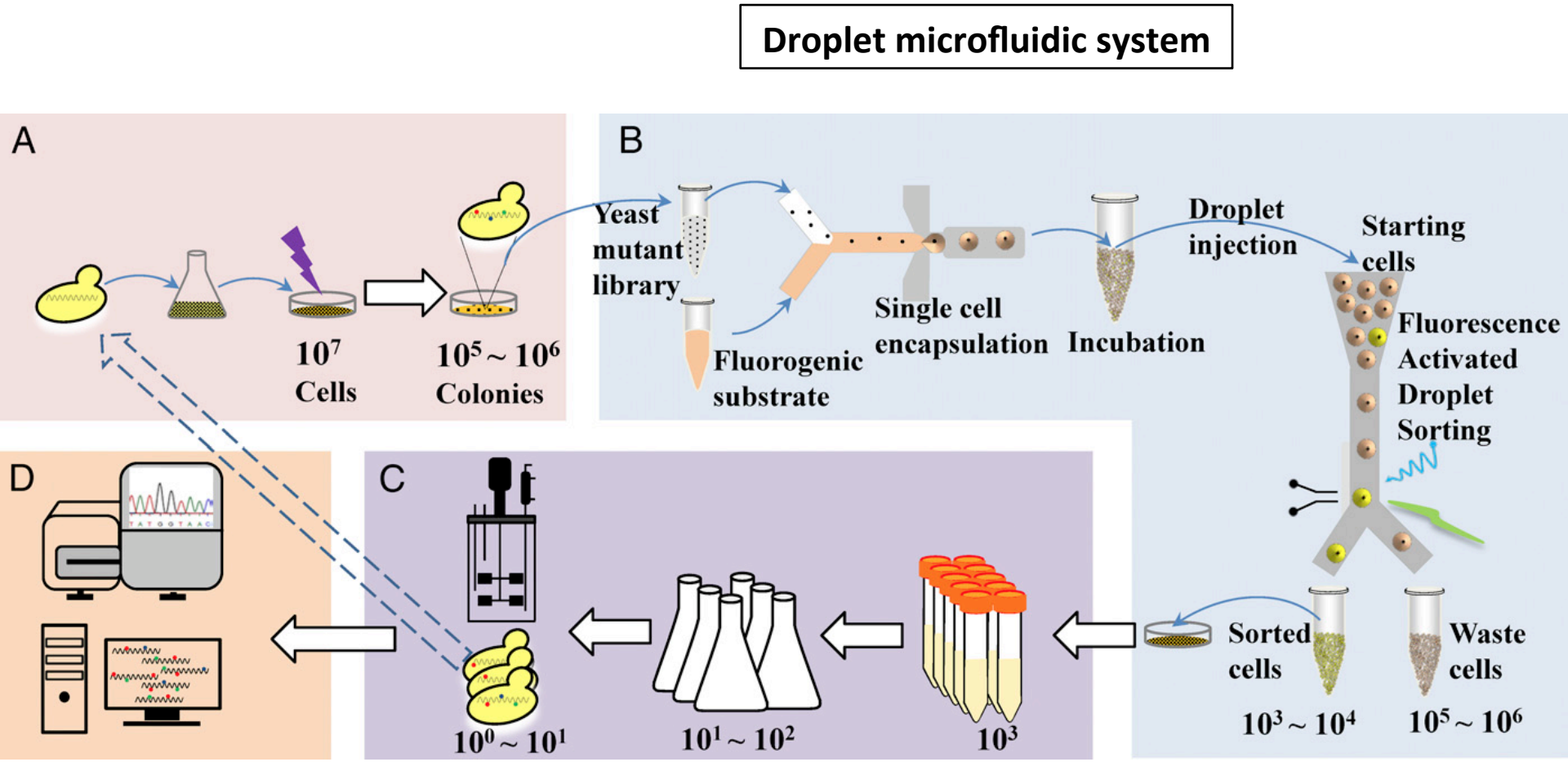
# *Example:* UV induced mutations in Yeast
## -- Modified cell factories increasing α-amylase

- Ultraviolet light (UV) irradiation induce mutations
- Select mutations having increased α-amylase production
- Identify genes involved for genetic engineering




**Huang et al., 2015**

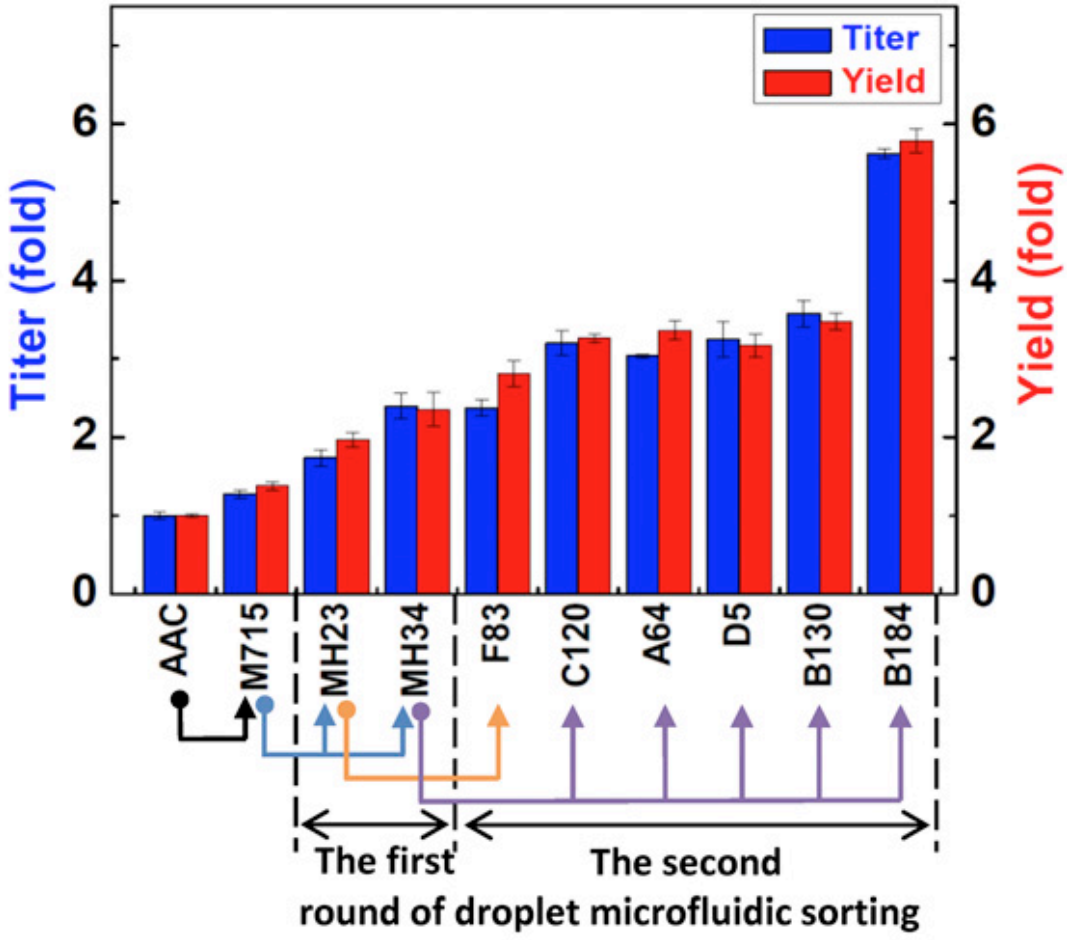
# *Example:* UV induced mutations in Yeast
## -- modified cell factories to produce α-amylase




Huang et al., 2015

# *Example:* UV induced mutations in Yeast
## -- modified cell factories to produce α-amylase

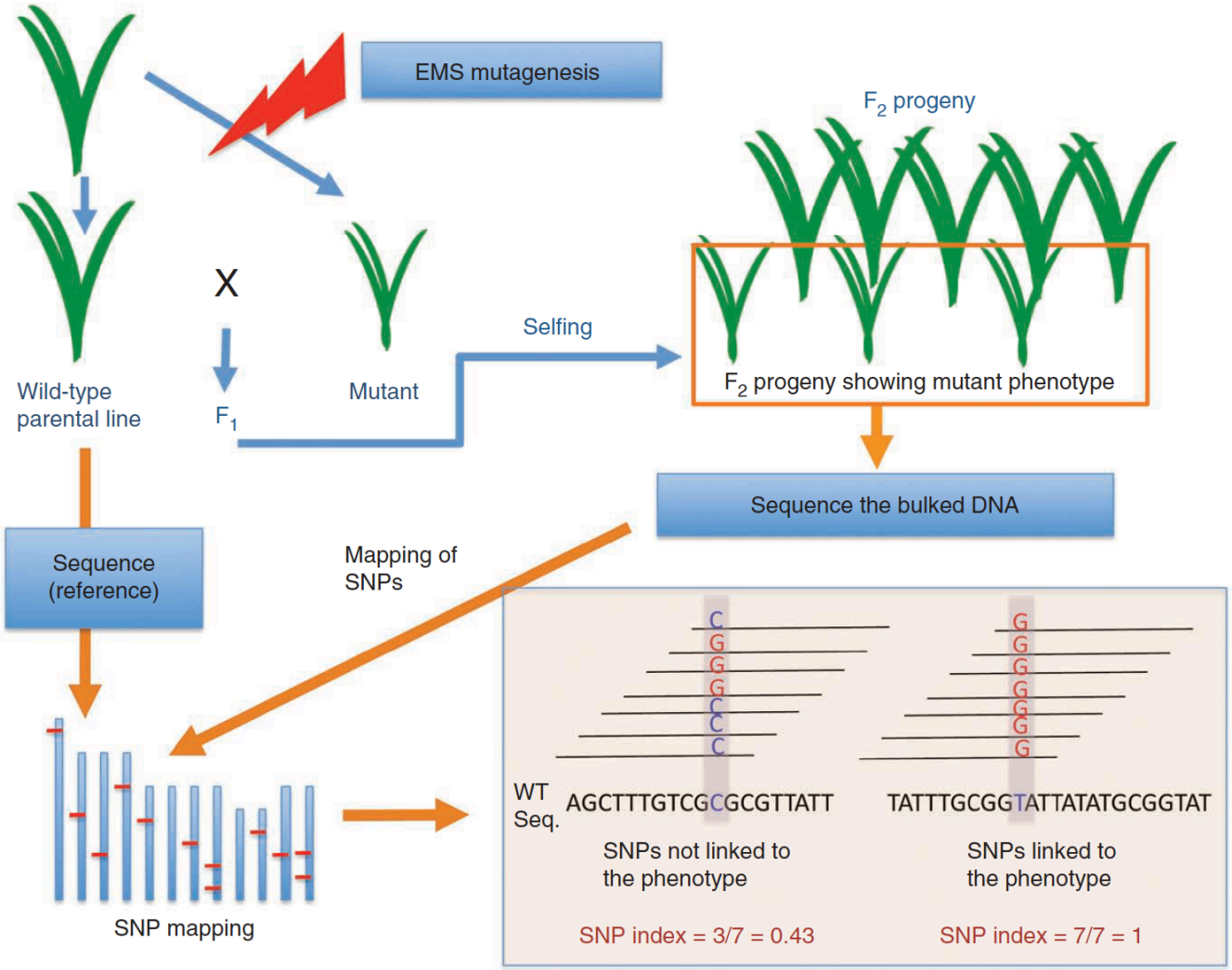
- Two cycles of selection of UV induced mutation leads to some yeast strains with improved α-amylase production




Huang et al., 2015

# MutMap

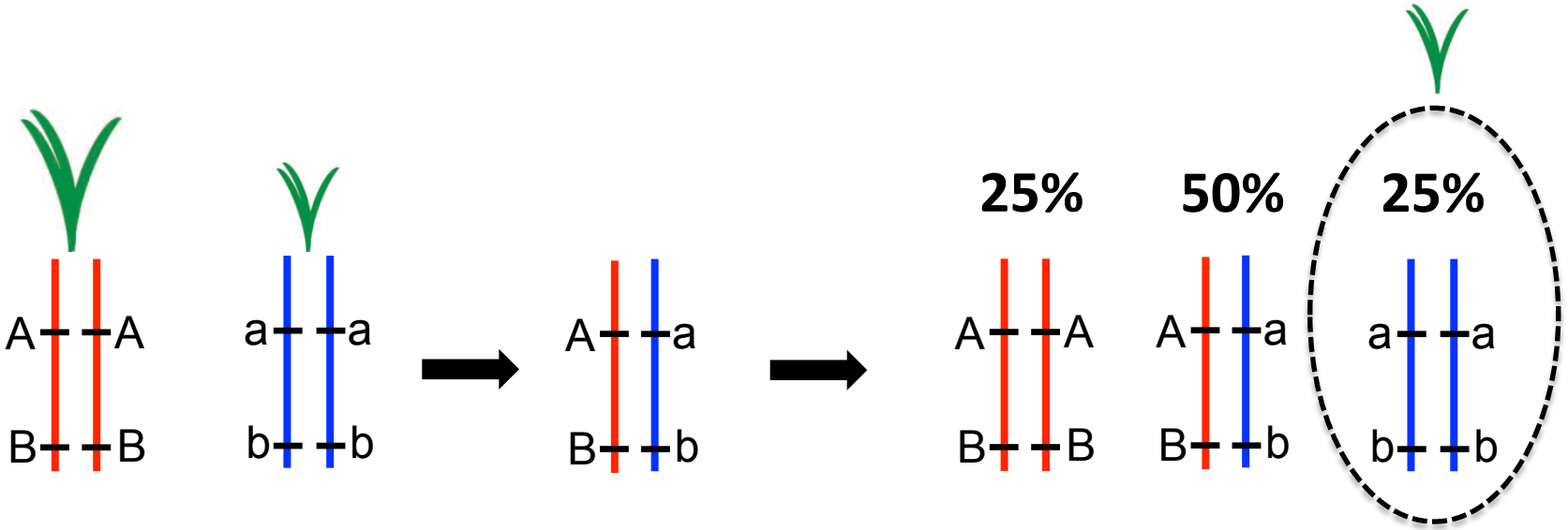
- EMS induce 2 to 10 mutations/ Mb of diploid DNA
- Rice: 400 Mbp, about 2,000 mutations per line




Abe et al., 2012

# MutMap

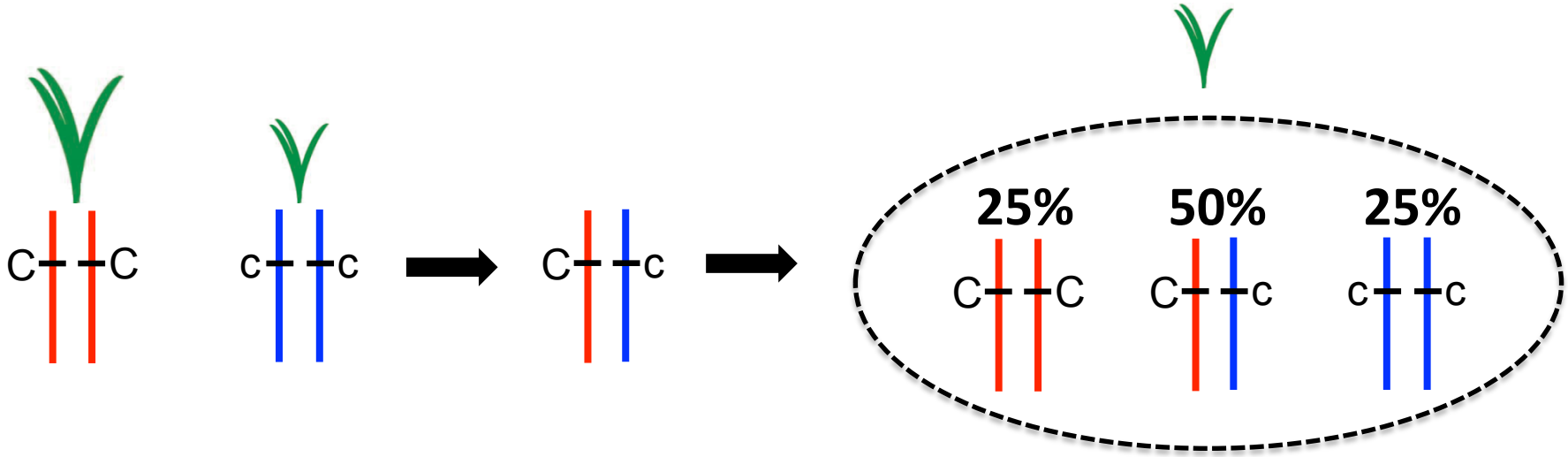
- SNP index is frequency of mutant allele in a group of individuals with mutant phenotype
- The causal SNP and very closely linked SNPs should show 100% mutant and 0% wild-type reads (SNP index=1)




Abe et al., 2012

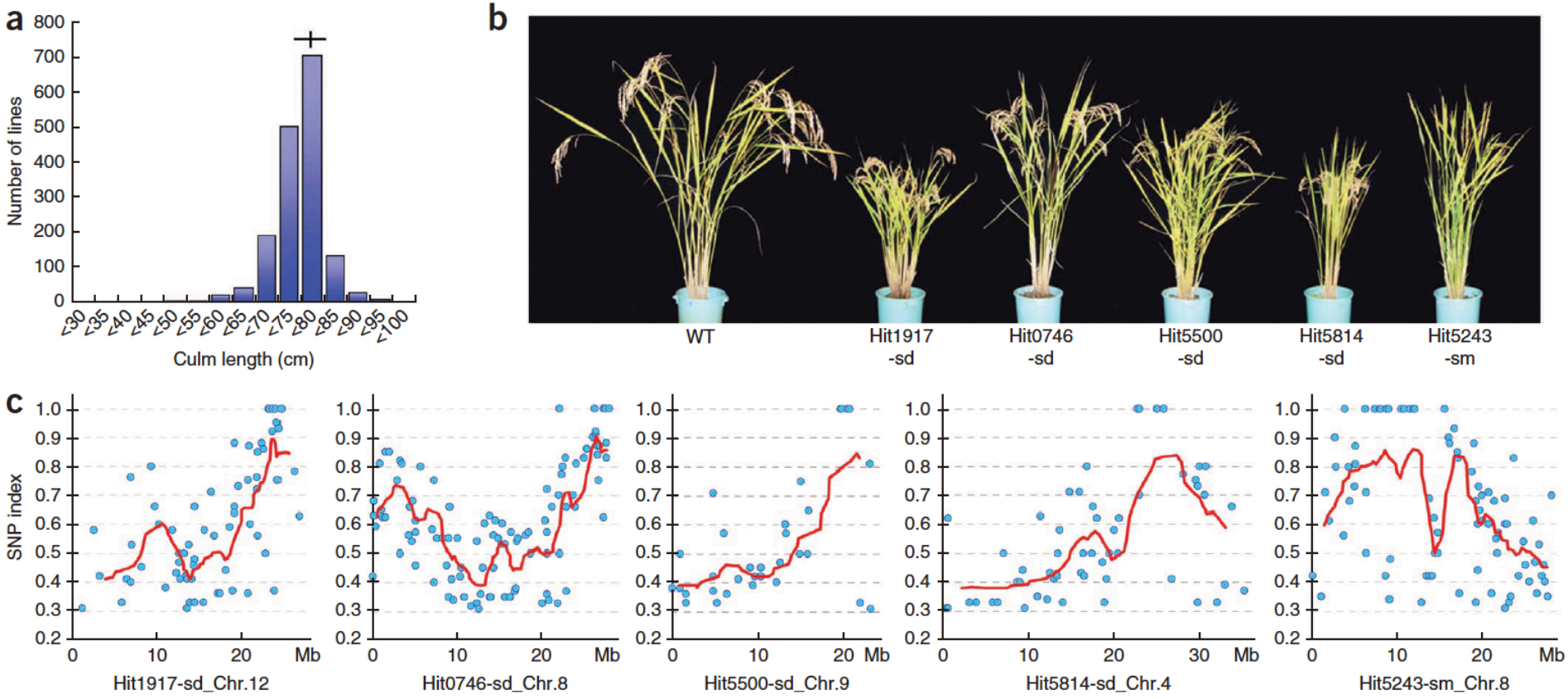
# MutMap

- SNP index is frequency of mutant allele in a group of individuals with mutant phenotype
- SNPs that are unlinked to the SNP responsible for the mutant phenotype have SNP index of 0.5




Abe et al., 2012

# Identification of genomic regions harboring causal mutations for five rice mutants using MutMap

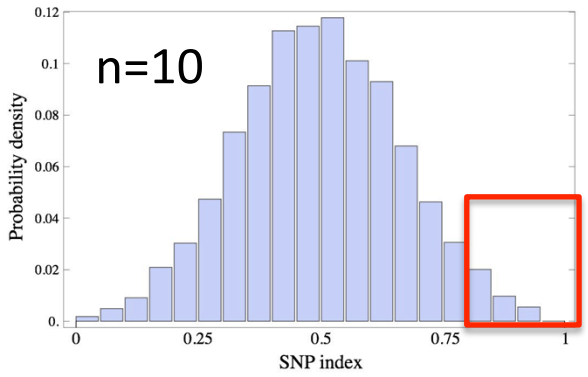


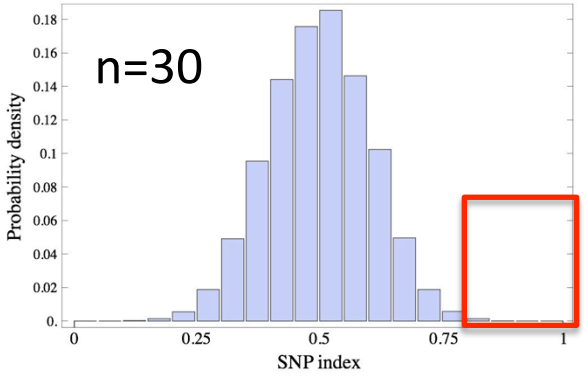

Abe et al., 2012

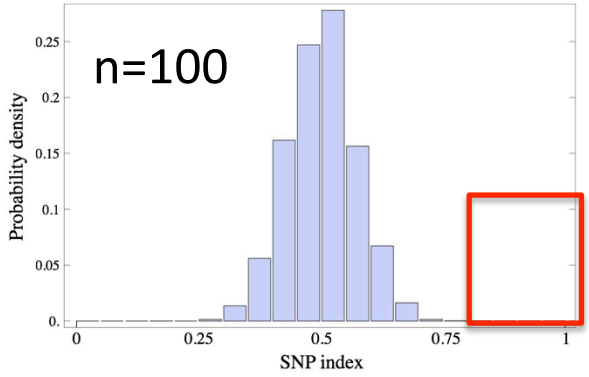
# Effect of bulk size (n) on the levels of false positive SNPs

- False positive error, a SNP (Cc) is not causal SNP, but was wrongly considered as a causal SNP
  - Theoretically, SNP index = 0.5
  - Estimation of SNP index is close to 1, when bulk size is small

Effect of n (G= 100 is fixed)



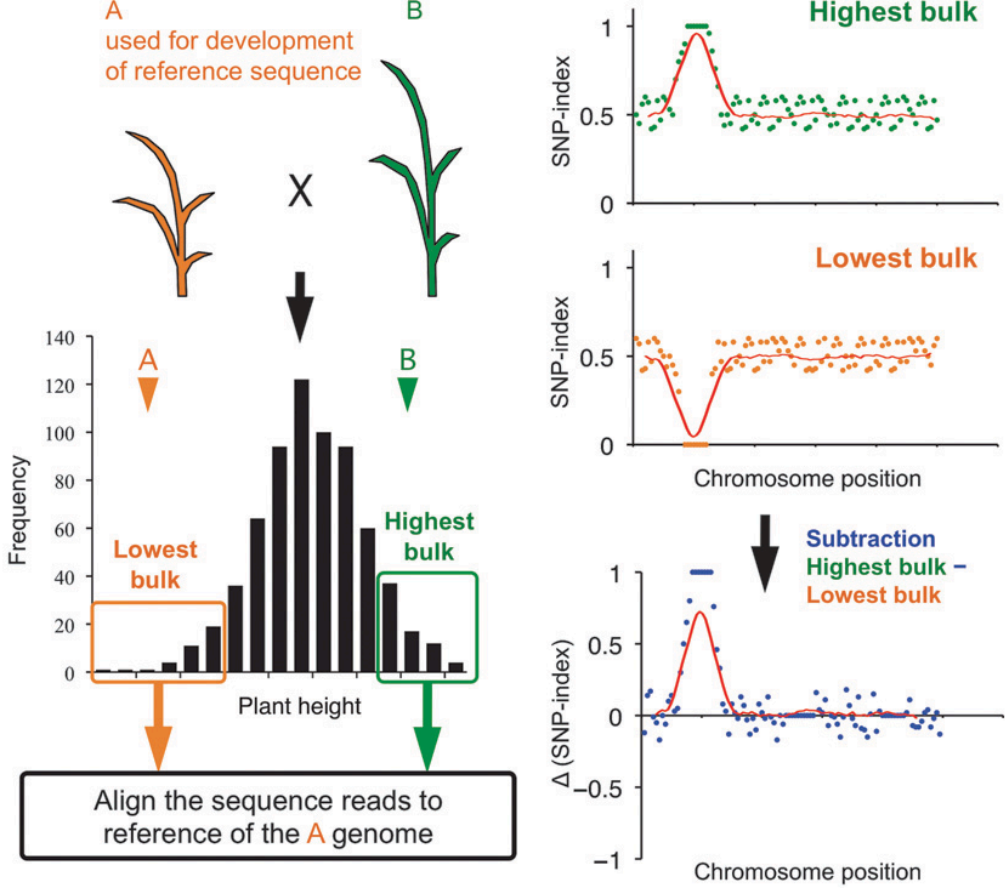



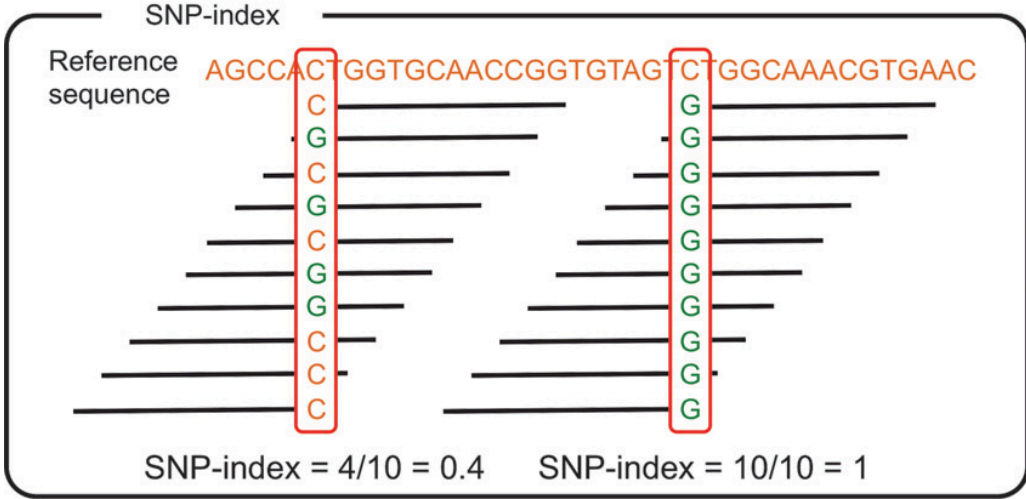




Abe et al., 2012

# QTLseq

- Two wild type of parents
- Two bulked DNA samples are generated from the progeny showing contrasting phenotypes



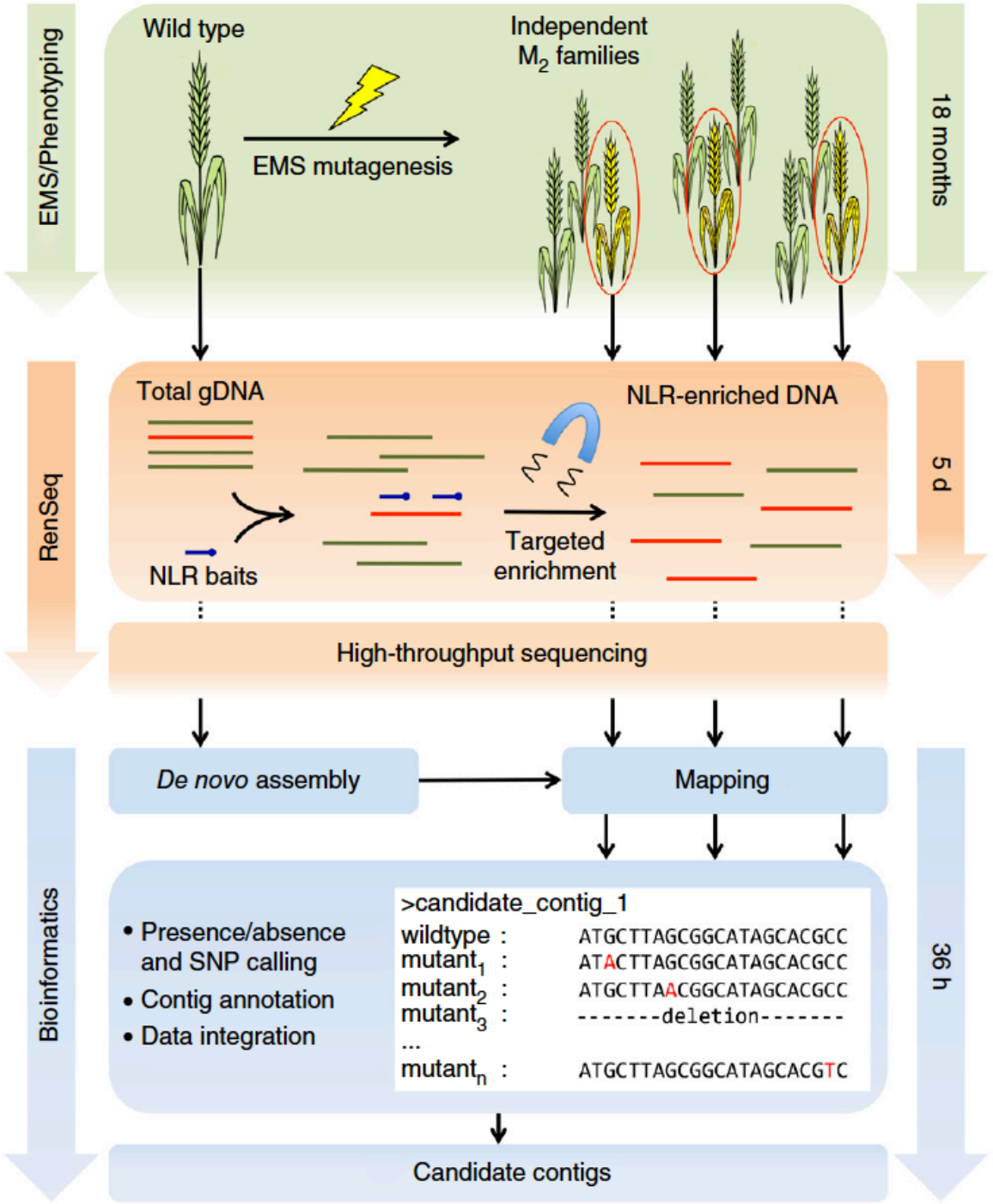

Takagi et al., 2013

# Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture

- Plant diseases can devastate crop yields and pose a threat to global food security
- Many R genes are present in gene families, with members in close physical proximity, such that dissection of the locus by recombination is not practical
- Most R genes encode proteins with nucleotide binding and leucine-rich repeats (NLRs)
- Sequencing of R genes of ethyl methane sulfonate (EMS)-derived, loss-of-resistance mutants with wild-type progenitors, called “MutRenSeq”. It enables the rapid identification of genes responsible for resistance without any positional fine mapping
- MutRenSeq combines chemical mutagenesis with exome capture and sequencing for rapid R gene cloning

**Steuernagel et al., 2016**

- Step 1 (green): EMS mutagenesis of resistant plant, creation of independent M2 families and screening for susceptible mutants (highlighted in yellow) fine mapping
- Step 2 (orange): target enrichment using a *Triticeae* NLR–specific bait library and sequencing of the wild-type and susceptible mutants (indicated by arrows)
- Step 3 (blue): data analysis and candidate calling




**Steuernagel et al., 2016**

# Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture

- MutRenSeq to clone the stem rust resistance gene *Sr22*, which was introgressed into wheat chromosome 7A from the diploid A-genome relatives (*T. boeoticum* and *T. monococcum*)
- In cultivar Schomburgk, *Sr22* confers resistance to commercially important races of the stem rust pathogen, including the Ug99 race
- Deployment of *Sr22* has been hampered owing to poor agronomic performance associated with the *Sr22*-introgression conferred by linked gene alleles (linkage drag)
- Efforts to clone *Sr22* in wheat with standard map-based approaches were unsuccessful owing to suppressed recombination in the *Sr22* region

**Steuernagel et al., 2016**

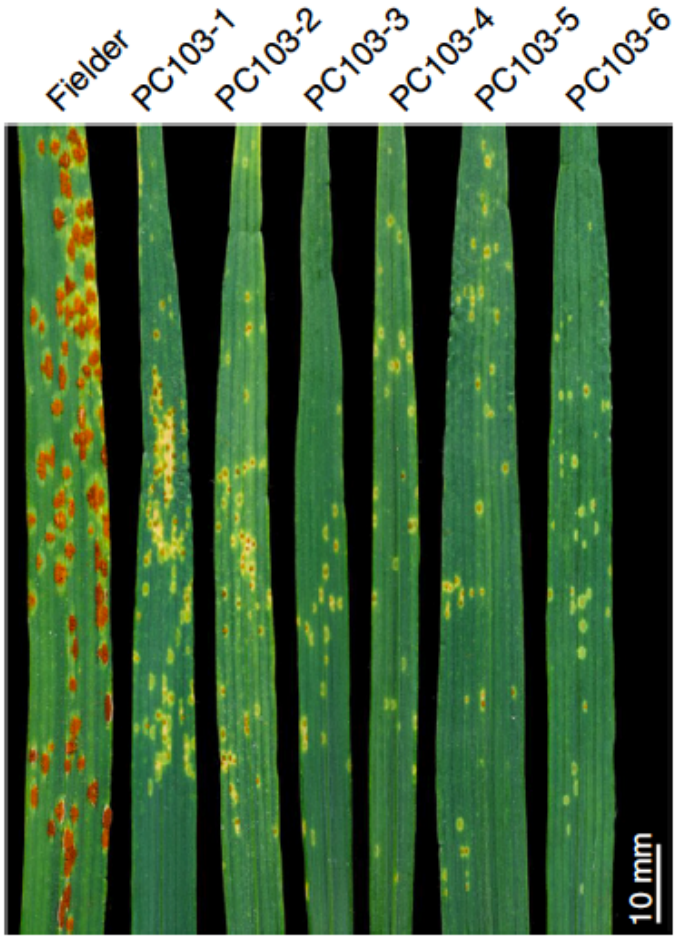
# Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture

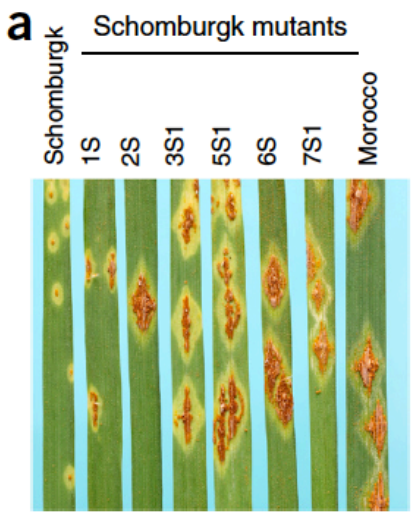
- Carried out an *Sr22* EMS suppressor screen using Schomburgk seeds and identified six independent susceptible mutants from 1,300 M2 families
- Sequenced the genomic NLR complement of Schomburgk (wild-type *Sr22*) and the six mutants using Illumina short-read sequencing and compared the mutant NLR complements to wild type
- The number of mutations ranged from 44 to 84, and identified 23 contigs that were mutated in two mutants, three contigs that were mutated in three mutants, and **a single 3,408-bp contig, that contained independent mutations in five of the six mutants**

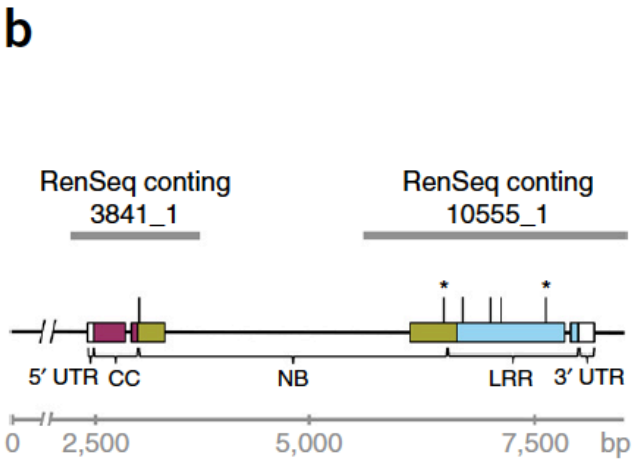
**Steuernagel et al., 2016**

# Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture

- All six mutations are GC to AT transitions that cause nonsense (two) or missense (four) mutations
- "To further verify *Sr22* cloning, we used the sequence to generate a PCR molecular marker, which co-segregated with *Sr22* in 2,300 gametes"
- All the transgenic lines were resistant to wheat stem rust with an infection phenotype similar to that of Schomburgk *Sr22*




**Steuernagel et al., 2016**

# Potential questions for final exam

- Briefly describe QTL mapping and map-based clone
- Methods to validate a candidate gene
- What is genome wide association mapping?
- What are the differences between QTL mapping and genome wide association mapping?
- What are linkage and linkage disequilibrium (LD)?
- What factors affect LD level?
- What factors affect genome wide association mapping accuracy?
- Sequencing methods to detect mutant genes from forward genetic screens, whole genome sequence, MutMap, and MutRenSeq

# References


- Anderson, J.A., Stack, R.W., Liu, S., Waldron, B.L., Fjeld, A.D., Coyne, C., Moreno-Sevilla, B., Fetch, J.M., Song, Q.J., Cregan, P.B. and Frohberg, R.C., 2001. DNA markers for Fusarium head blight resistance QTLs in two wheat populations. *Theoretical and Applied Genetics*, *102*(8), pp.1164-1168.
- Bergelson, J. and Roux, F., 2010. Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature Reviews Genetics*, *11*(12), p.867.
- **Borevitz, J.O. and Chory, J., 2004. Genomics tools for QTL analysis and gene discovery. *Current Opinion in Plant Biology*, *7*(2), pp.132-136.**
- Hamblin, M.T., Buckler, E.S. and Jannink, J.L., 2011. Population genetics of genomics-based crop improvement methods. *Trends in Genetics*, *27*(3), pp.98-106.
- Huang, M., Bai, Y., Sjostrom, S.L., Hallström, B.M., Liu, Z., Petranovic, D., Uhlén, M., Joensson, H.N., Andersson-Svahn, H. and Nielsen, J., 2015. Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proceedings of the National Academy of Sciences*, *112*(34), pp.E4689-E4696.
- Huang, B. Emma, et al. "MAGIC populations in crops: current status and future prospects." *Theoretical and Applied Genetics* 128.6 (2015): 999-1017.
- Ingvarsson, Pär K., and Nathaniel R. Street. "Association genetics of complex traits in plants." *New Phytologist* 189.4 (2011): 909-922.
- Jander, Georg, et al. "Arabidopsis map-based cloning in the post-genome era." *Plant Physiology* 129.2 (2002): 440-450.

# References


- Liu, S., Pumphrey, M., Gill, B., Trick, H., Zhang, J., Dolezel, J., Chalhoub, B. and Anderson, J., 2008. Toward positional cloning of Fhb1, a major QTL for Fusarium head blight resistance in wheat. *Cereal Research Communications*, *36*(Supplement 6), pp.195-201.
- Liu, S., Zhang, X., Pumphrey, M.O., Stack, R.W., Gill, B.S. and Anderson, J.A., 2006. Complex microcolinearity among wheat, rice, and barley revealed by fine mapping of the genomic region harboring a major QTL for resistance to Fusarium head blight in wheat. *Functional & integrative genomics*, *6*(2), pp.83-89.
- **Mackay, I. and Powell, W., 2007. Methods for linkage disequilibrium mapping in crops. *Trends in plant science*, *12*(2), pp.57-63.**
- **Peters, J.L., Cnudde, F. and Gerats, T., 2003. Forward genetics and map-based cloning approaches. *Trends in plant science*, *8*(10), pp.484-491.**
- Rawat, N., Pumphrey, M.O., Liu, S., Zhang, X., Tiwari, V.K., Ando, K., Trick, H.N., Bockus, W.W., Akhunov, E., Anderson, J.A. and Gill, B.S., 2016. Wheat Fhb1 encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. *Nature genetics*, *48*(12), pp.1576-1580.
- **Schneeberger, K., 2014. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature Reviews Genetics*, *15*(10), p.662.**
- Steuernagel, B., Periyannan, S.K., Hernández-Pinzón, I., Witek, K., Rouse, M.N., Yu, G., Hatta, A., Ayliffe, M., Bariana, H., Jones, J.D. and Lagudah, E.S., 2016. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nature biotechnology*, *34*(6), p.652.
- Waldron, B.L., Moreno-Sevilla, B., Anderson, J.A., Stack, R.W. and Frohberg, R.C., 1999. RFLP mapping of QTL for Fusarium head blight resistance in wheat. *Crop Science*, *39*(3), pp.805-811.
- Zhu, Chengsong, et al. "Status and prospects of association mapping in plants." *The plant genome* 1.1 (2008): 5-20.