DNA marker and association mapping

- **Association mapping** is a forward genetics method to identify gene(s) and/or causal mutation(s) associated with a target trait by using association analysis between DNA markers and phenotype of the target trait.

- **DNA marker** (or genetic marker) is a fragment of DNA within the genome revealing variation/mutation.

- **Single Nucleotide Polymorphism (SNP)** is abundant in the genome and is the most popular type of DNA marker used in genetics studies; most SNPs have two marker alleles.
Association mapping to identify gene(s) associated with tan spot resistance in wheat

<table>
<thead>
<tr>
<th>SampleID</th>
<th>Phenotype</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>1</td>
<td>............ACGGTG...CGGCA..........................TGAT......AAGGGG.............</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 2</td>
<td>2</td>
<td>............ACGGTG...CGGCA..........................TGAT......AAGGGG.............</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 3</td>
<td>1</td>
<td>............ACGGTG...CGGCA..........................TGAAA......AAGGC..........</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 4</td>
<td>1</td>
<td>............ACCGT...CGGCA..........................TGAAA......AAGGC..........</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 5</td>
<td>4</td>
<td>............ACCGT...CGGCT..........................TGAAA......AAGGGG.............</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 6</td>
<td>5</td>
<td>............ACCGT...CGGCT..........................TGAAA......AAGGC..........</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 7</td>
<td>5</td>
<td>............ACCGT...CGGCT..........................TGAT......AAGGC..........</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 8</td>
<td>S</td>
<td>............ACCGT...CGGCT..........................TGAT......AAGGGG.............</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Procedure of association mapping

- Collect phenotypic data of a target trait from a population
- Collect DNA marker data from the same population
- Estimate the phenotypic means and variances of different marker alleles at a marker locus
- Statistical test on differences between the phenotypic means of different alleles at a marker locus
Association mapping of tan spot resistance in durum wheat

- Population: 371 durum wheat landraces
- Phenotype data of tan spot resistance
  - Toxin: ToxA and ToxB (scored as 1 and 0)
  - Isolate: 86-124 and DW5 (scored as 1, 2, 3, 4, and 5 from R to S)
- Marker data
  - Over 60,000 SNP markers on 14 chromosomes
## Association mapping of tan spot resistance in durum wheat

<table>
<thead>
<tr>
<th>Ref</th>
<th>Alt</th>
<th>Chr</th>
<th>Position</th>
<th>Est-Ref</th>
<th>Est-Alt</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C (196)</td>
<td>T (129)</td>
<td>chr1A</td>
<td>1207866</td>
<td>1.98</td>
<td>2.02</td>
<td>0.90</td>
</tr>
<tr>
<td>A (187)</td>
<td>G (150)</td>
<td>chr1A</td>
<td>1238074</td>
<td>1.95</td>
<td>2.05</td>
<td>0.49</td>
</tr>
<tr>
<td>A (211)</td>
<td>G (108)</td>
<td>chr1A</td>
<td>1336691</td>
<td>2.06</td>
<td>1.94</td>
<td>0.47</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>A (133)</td>
<td>T (198)</td>
<td>chr3B</td>
<td>345940331</td>
<td>2.70</td>
<td>1.30</td>
<td>1.02E-4</td>
</tr>
<tr>
<td>C (119)</td>
<td>G (211)</td>
<td>chr3B</td>
<td>345943215</td>
<td>2.85</td>
<td>1.15</td>
<td>7.55E-8</td>
</tr>
<tr>
<td>G (126)</td>
<td>C (205)</td>
<td>chr3B</td>
<td>345943463</td>
<td>2.75</td>
<td>1.25</td>
<td>1.19E-5</td>
</tr>
<tr>
<td>A (162)</td>
<td>T (169)</td>
<td>chr3B</td>
<td>346537120</td>
<td>1.40</td>
<td>2.60</td>
<td>8.90E-4</td>
</tr>
</tbody>
</table>

Galagedara et al., 2020
Manhattan plots of association mapping for reaction to ToxA and ToxB

Galagedara et al., 2020
Manhattan plots of association mapping for reaction to isolates

86-124

DW5

Galagedara et al., 2020
Wheat genome browsers

- https://www.wheatgenome.org/News/Latest-news/IWGSC-Reference-Sequence-v1.0-browser-now-available-at-URGI
- https://plants.ensembl.org/Triticum_aestivum/Info/Index
Genome wide association (GWA) mapping

- GWA mapping needs or prefers to have
  - Reference genome available
  - A large number of genome wide DNA markers available so that to have at least one marker linked to the causal gene or mutation
  - Genotyping platform that can easily genotype large number of DNA markers for hundreds even thousands of individuals with low cost
GWA mapping in human

- Human genome project, the first human genome was completely assembled in 2003

- 1,000 human project, created a haplotype map of 38 millions SNPs by using low-coverage whole-genome and exome sequencing of 1094 human genomes

- Human SNP arrays provide low-cost and high-throughput genotyping of thousands or millions of human DNA
GWA mapping of human diseases

• GWAS catalog
  – https://www.ebi.ac.uk/gwas/
  – https://www.ebi.ac.uk/gwas/diagram
Million human projects

• 1-million-person long-term health study started in 2016, cost $4 billion over 10 years (Kaiser et al., 2016); Researchers will use the data to develop “precision medicine,” or personalized treatments for others

• AstraZeneca, one of the world’s largest pharmaceutical companies, launches a project to sequence 2 million genomes for new drug targets (Ledford et al., 2016)
Personalized medicine

• Pharmacogenomics to personalized medicine
  – Study of genetic variation on drug responses
  – Using an individual’s genome to determine whether or not a particular therapy, or dose of therapy, will be effective
Personalized medicine

• More than 100 Food and Drug Administration (FDA)-approved drugs have pharmacogenomics information in their labels, in diverse fields such as analgesics, antivirals, and anti-cancer therapeutics

  ✓ http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm
Linkage mapping using a bi-parental mapping population

- In 1990s, DNA markers were developed
- Before 2000, few species were whole-genome sequenced
- DNA markers’ physical positions on a chromosome are unknown
- Linkage mapping using a bi-parental mapping population was an approach to locate genes involved in a trait
Linkage mapping using a bi-parental mapping population

1. Create a segregating population for the target trait **using two parents**
2. Genotype DNA markers and **construct genetic linkage map**
3. Phenotype the population for the interested trait
4. Perform marker-trait statistical analysis to find markers linked to the causal genes
Linkage mapping using bi-parental mapping population

Parent1 (R: 1) × Parent2 (S: 5)

\[ F_1 \]

Phenotype

RILs (200 F_7 lines)

Marker

\( M_1 \)–\( m_1 \)
\( M_2 \)–\( m_2 \)
\( M_3 \)–\( m_3 \)
Linkage

- Two markers that are physically near to each other are unlikely to be separated during chromosomal crossover
- The closer, the more likely to be inherited together
Meiosis and homologous recombination

- Two copies of the same chromosome break and rejoin at the same point, which can generate a new **haplotype** of two or more markers

- **Haplotype** is specific combination of alleles occurring on the same chromosomal segment (including two or more markers)
Genetic distance between two markers

- Recombination rate \( (r) \) is percentage of the recombinant haplotypes: \( r = \frac{M1m2 + m1M2}{M1M2 + m1m2 + M1m2 + m1M2} \% \)

![Diagram showing F1 and F7s with gamete counts and recombination rate](image)

<table>
<thead>
<tr>
<th>Gamete</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1M2</td>
<td>92</td>
</tr>
<tr>
<td>m1m2</td>
<td>88</td>
</tr>
<tr>
<td>M1m2</td>
<td>10</td>
</tr>
<tr>
<td>m1M2</td>
<td>10</td>
</tr>
</tbody>
</table>

Rec. rate \( (r) \) = 0.10 (10 cM)
### Grouping of markers

<table>
<thead>
<tr>
<th>Marker1</th>
<th>Marker2</th>
<th>r</th>
<th>cM</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>M2</td>
<td>0.1</td>
<td>10</td>
<td>G1</td>
</tr>
<tr>
<td>M1</td>
<td>M3</td>
<td>0.3</td>
<td>30</td>
<td>G1</td>
</tr>
<tr>
<td>M1</td>
<td>M4</td>
<td>0.3</td>
<td>30</td>
<td>G1</td>
</tr>
<tr>
<td>M1</td>
<td>M5</td>
<td>0.5</td>
<td>50</td>
<td>M5 not in G1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>M3</td>
<td>0.2</td>
<td>20</td>
<td>G1</td>
</tr>
<tr>
<td>M2</td>
<td>M4</td>
<td>0.2</td>
<td>20</td>
<td>G1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>M5</td>
<td>0.5</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>
Ordering markers within a linkage group

- The closer markers are, the more likely to be inherited together
### Linkage mapping

<table>
<thead>
<tr>
<th>SNPs</th>
<th>Marker 1</th>
<th>Marker 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Marker Allele (QTL allele)</strong></td>
<td>M1 (100) (95Q and 5q)</td>
<td>M3 (100) (50 Q and 50 q)</td>
</tr>
<tr>
<td></td>
<td>m1 (100) (95q and 5Q)</td>
<td>m3 (100) (50 Q and 50 q)</td>
</tr>
<tr>
<td><strong>Mean resistance</strong></td>
<td>1.2</td>
<td>4.5</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.0005</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Fusarium Head Blight (FHB) in wheat

- FHB, caused by *Fusarium graminearum*, is a devastating disease of wheat worldwide
- FHB causes both severe loss of grain yield and quality
- Epidemics of FHB from 1993 to 1997 in Northern Great Plains; over one billion dollars loss to wheat industry in 1993 (McMullen et al., 1997); all commercial cultivars were susceptible
- Developing and growing resistant cultivars is the most efficient mean to minimize the negative effects of the disease

Waldron et al., 1999
Linkage mapping of FHB resistance in Wheat

• Create a population segregating for FHB resistance
  – A population of 112 $F_5$-derived recombinant inbred lines (RILs) developed by single seed descent from the spring wheat cross ‘Sumai3’/’Stoa’
  – Sumai3 is a Chinese cultivar with high resistance to FHB
  – Stoa (susceptible to FHB) is a hard red spring cultivar released by North Dakota State University in 1984

Waldron et al., 1999
Linkage mapping of FHB resistance in Wheat

• Phenotyping of FHB resistance
  – Evaluated in two experiments, each with three replications, 1994 and 1995
  – As plants reached anthesis, approximately 10 spikes per line in each replicate were inoculated 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%

Waldron et al., 1999
Initial linkage mapping located a major gene in a large interval on Chr3B

- Marker-trait association analysis
  - Five genomic regions were significantly associated with FHB resistance
  - The best single marker, Xcdo981, was mapped on chromosome 3BS
  - This marker explained 15.4% of the phenotypic variation

Waldron et al., 1999
Initial linkage mapping located a major gene in a large interval on Chr3B.
Fine mapping to narrow down the Chr3B candidate region

Parent1 (R)       M1       M2       M3       M4       M5       M6
Parent2 (S)       m1       m2       m3       m4       m5       m6
RILa (R)          
RILb (R)          
RILc (S)          
RILd (S)          
RILe (S)
Fine mapping of the Chr3B genomic region

* = marker allele from Stoa
+ = marker allele from Sumai 3
- = marker allele from MN97448

Selfing for six generations with selection for heterozygote beginning at F₄ generation

F₇

260-1-1-2
+ + +

260-1-1-8 (a single plant)
+ + +

260-1-1-4
= = =

F₈ (equivalent to F₁)

22 heterozygous plants
+ + +

F₉ (equivalent to F₂)

recombinants

Liu et al., 2006
Near isogenic lines (NILs) to fine mapping

F1

↓ Selfing 6 generations

M1  M6

m1  m6

260-1-1-8

↓ Selfing 2 generations

NIL1

NIL2
Fine mapping of the Chr3B genomic region

- Two SSR markers flanking Qfhs.ndsu-3BS, gwm533 and gwm493, were used to identify the recombinants.

- Among the 3,156 plants screened for recombinants with the two SSR marker loci, Xgwm533 and Xgwm493, 382 recombinants were identified.

- All 382 recombinants were genotyped with two more SSR markers, BARC133 and BARC147, and eight STS markers.
Fine mapping of the Chr3B genomic region

Liu et al., 2006
Linkage mapping and map-based clone of *Fhb1* in wheat

Rawat et al., 2016
Selection of *Fhb1* candidate gene

- Expression analyses for the annotated genes were performed in spikes, inoculated with *Fusarium* macroconidia, of resistant near-isogenic line (R-NIL) with *Fhb1* and susceptible NIL (S-NIL) lacking the *Fhb1*
  - The genes *PFT* and *NBA* were expressed only in R-NIL and not in S-NIL, whereas the other genes had similar expression patterns in both the NILs

Rawat et al., 2016
Selection of *Fhb1* candidate gene

- 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, excluded *NBA* and considered *PFT* as the putative candidate for *Fhb1*
Validation of the *Fhb1* candidate gene, *PFT*

- *PFT* is a 3,472-bp gene with two exons generating a 1,437-bp mRNA

1. Assessed the candidacy of *PFT* for *Fhb1* using induced mutants
2. RNA interference induced gene silencing
3. Gene complementation by transformation

Rawat et al., 2016
EMS induced mutants of the gene *PFT*

- Five mutations of *PFT* 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

Rawat et al., 2016
PFT validation with RNAi gene silencing

- Sumai3 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
Validation of *PFT* with transgenic lines

- Generated transgenic plants expressing the *PFT* gene in wheat cultivar Fielder
Linkage mapping using a bi-parental mapping population to clone gene

- Initial linkage mapping
  1) Create a segregating population for a target trait using two parents
  2) Genotype the population for molecular markers and construct linkage map
  3) Phenotype the population for the target trait
  4) Perform marker-trait statistical analysis to find markers linked to the causal gene

- Fine mapping

- Candidate gene validation
Linkage mapping using induced mutant

Wild type A → EMS → Parent1 (R: 1) × Parent2 (S: 5) → F₁ → Phenotype

RILs (200 F₇ lines)

MutMap: rapid gene isolation using a cross between a mutant and wild type

- Rice: about 2,000 point mutations per mutant induced by EMS
MutMap

- SNP index $\frac{a}{(G+a)}$ is frequency of mutant allele in a group of individuals with mutant phenotype
- The causal mutation or SNP is $M1$; SNP index of $M1 = \frac{a}{(G+a)} = 1$

Abe et al., 2012
MutMap

- The mutations or SNPs (like $M3$) that are unlinked to the SNP responsible for the mutant phenotype
- Expected SNP index for $M3 = \frac{a}{G+a} = 0.5$

Abe et al., 2012
The mutation or SNP (\(M2\)) that is very close to the causal SNP (\(M1\))

- SNP index of \(M2 = a/(G+a) = \text{???)}

Abe et al., 2012
Identification of genomic regions harboring causal mutations for five rice mutants using MutMap

Abe et al., 2012
What need to know for final exam

- What is association mapping?
- What is linkage mapping?
- What is fine mapping?
- Basic procedure of association mapping, linkage mapping, and fine mapping
- Methods used in validation of a candidate gene
- Basic procedure of MutMap

