QTL mapping

Paul Gepts
PLB152: Plant Genetics

Sources

- **** Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196**
- *Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio E, Vale G (2005) Marker assisted selection in crop plants. Plant Cell Tissue and Organ Culture 82:317-342 (for this lecture: especially pp. 317-321)
- Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14:214-219
- Xu YB, McCouch SR, Zhang QF (2005) How can we use genomics to improve cereals with rice as a reference genome? Plant Molec Biol 59:7-26
- Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621-630
- Kelly JD, Gepts P, Miklas PN, Coyne DP (2003) Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. Field Crops Res 82:135–154
- Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560

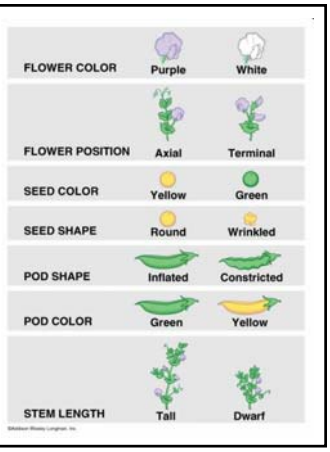
Outline

- What is a qualitative/quantitative trait?
- QTL mapping
- Towards MAS: marker-assisted selection

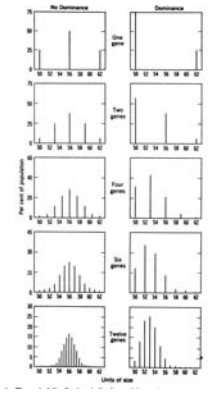
Difference between a qualitative and a quantitative trait

- Qualitative:
  - Presence or absence:
    - Growth habit: tall vs. dwarf
    - Pigmentation: pigmented vs. non-pigmented
    - Disease reaction: resistant vs. susceptible
  - Generally single gene trait
  - Markers are qualitative traits
- Quantitative:
  - Quantity:
    - Tallness
    - Yield
  - Generally more than one gene and/or environmental effects

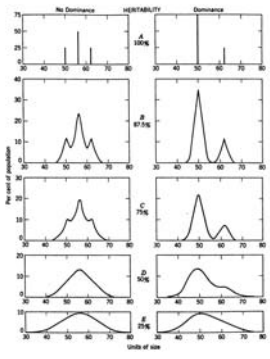
Examples of qualitative traits: see Mendel!



Effect of the number of genes on segregation



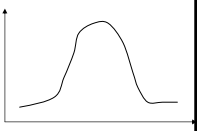
Effect of the magnitude of the environmental effect on segregation



From Allard 1992

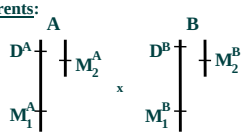
Consequence of the nature of segregation of quantitative traits

- No one-on-one relationship between phenotype and genotype:
  - One cannot infer the genotype underlying a phenotype
  - Uncertainty as to the number of genes and/or magnitude of environmental effects
- Solution: proposed by Sax (1923)
  - "Mendelize" the segregation
  - Correlate segregation of the quantitative trait with that of qualitative trait, i.e., markers
  - QTL = quantitative trait locus = gene!



Methodology of Sax (1923)

Parents:



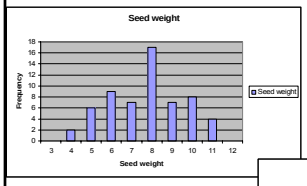
D: gene for domestication
M: molecular markers 1 & 2

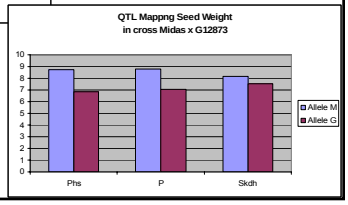
Segregating progeny: subpopulations based on segregation for molecular markers

M_1^A (mostly D^A; some D^B) vs. M_1^B (mostly D^B; some D^A)

M_2^A (D^A + D^B) vs. M_1^B (D^A + D^B)

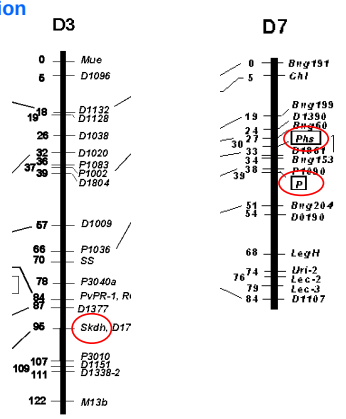
Example: seed weight





Mapping information

Based on results:
- No seed weight gene near *Skdh* locus
- One or more seed weight genes near *P* and *Phs*; same or different?



Methodology of Sax (1923) (contd.)

- Repeat previous analysis for set of markers distributed at regular intervals on each chromosome in the genome
- Results: i.e. data obtained
  - minimum number of genes distinguishing the two parents
  - magnitude of the phenotypic effect of individual genes
  - total proportion of phenotypic variation in the segregating population for a given trait that can be accounted for in genetic terms
  - linkage relationships among genes
- Same basic approach in human genetics but some modifications: e.g., affected siblings method

Methods to detect QTLs

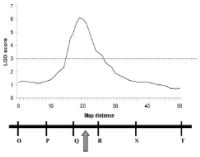
- Single-marker analysis:
  - Divide population into subpopulations based on allelic segregation of individual loci
  - Measure phenotypic trait and average in each subpopulation
  - Determine if statistically significant differences by:
    - t-test, ANOVA, linear regression (R^2)
  - Disadvantage:
    - Underestimation of the effect of a linked QTL due to recombination between marker and QTL
    - Solution: use markers at spacing < 10-15 cM

Methods to detect QTLs

- Simple interval mapping:
  - Analyzes intervals between adjacent markers instead of single markers
  - Eliminates problems of recombination between marker and QTL
  - Statistically more powerful
- Composite interval mapping
  - Analyzes intervals between adjacent markers + additional markers unlinked to the interval markers to focus on the interval and eliminate confounding effects from other QTLs:
    - Elsewhere in the genome: "background noise"
    - Linked to QTL: biases location of QTL

QTL mapping statistics

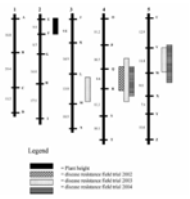
- LOD = logarithm of the odds ratio
  - (P of observing the data under linkage)/ P of observing the same data under no linkage)
  - Key value: ≥ 3; this means 1000 more likely under linkage
- Permutation test:
  - Shuffle phenotypic data but keep marker data constant
  - Repeat 1000 times
  - Determine significance level such that false positives is below a certain level.



Reporting QTL data

- Generally on linkage maps:
  - Rectangles:
    - Encompassing most closely linked markers
    - Exceeding significance level
- "Major" and "minor" QTLs:
  - R^2 > 30%: major; <10%: minor
- Confidence interval:
  - Around the max LOD location:

  Region that corresponds to a decline of 1 LOD from the peak

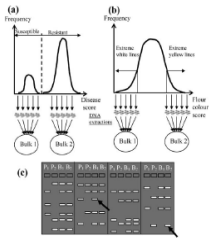


Methodological concerns

- Number of markers:
  - Approx. every 10 cM → total number depends on genetic size of the genome; usually around 100-200
- Shared markers:
  - Try using > 2-3 shared markers per chromosome to allow comparisons among different maps
- Size of the mapping population:
  - Larger populations: >150-200
    - More accurate mapping of QTLs
    - Detection of smaller QTLs
    - Less overestimation of magnitude of QTLs
- Confirmation of QTL mapping
  - Different population from same parents
  - Near-isogenic lines

Short-cuts for gene tagging

- Bulked segregant analysis:
  - Two bulks of DNA samples from ~ 10 individuals each, differing for the trait of interest
  - For traits controlled by major genes or QTLs
- Trait-based marker analysis:
  - Select individuals with extreme phenotypes
  - Only for one trait at a time



Towards marker-assisted selection

- MAS = select for linked marker instead of the phenotypic trait
  - Time savings compared to complex field trials
  - Elimination of unreliable or unfeasible (e.g., quarantine) phenotypic evaluations
  - Selection at a different development or growth stage
  - Gene pyramiding
  - Limit transfer of linked, undesirable genes ("linkage drag")
  - Selection for traits with low heritability

From Francia et al. 2005

Development of markers for MAS

- High-resolution mapping
  - Identify tightly linked markers: < 1 cM
  - May require populations > n = 1000
- Validation of markers
  - Testing markers in other genetic backgrounds
    - Presence of the QTL in these backgrounds?
    - Polymorphism in other genotypes?
- Marker conversion:
  - Technically more reliable and simpler markers:
    - RAPD, RFLP, AFLP → STS, SCAR

Marker-assisted selection

- When do you apply MAS?
  - Magnitude and number of genes or QTLs
    - Easier with fewer genes of major effect
  - Heritability of trait:
    - The lower the heritability, the higher the relative efficiency of MAS
  - Cost relative to phenotypic selection
    - Inheritance of the trait
    - Method of phenotypic evaluation
    - Field/greenhouse and labor costs
    - Cost of development of markers
  - Maximum selection efficiency for quantitative traits:
    - Combination of molecular and phenotypic selection.

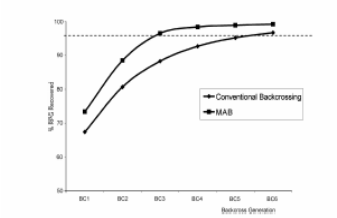
Limitations

- Numerous QTL studies but few have resulted in application of MAS
  - Major QTLs often not found
  - Uncertainty of the QTL position (population size, magnitude of the QTL)
  - Deficiencies in QTL analysis leading to an inaccurate estimation of number and magnitude of QTLs
  - Limited usefulness of markers across backgrounds and environments
  - Epistatic effects
  - QTL x environment effects

Marker-assisted backcrossing

- Traditional BC:
  - After 1 BC: recover 75% of recurrent parent genome
  - But: wide variation in genome composition among BC lines
- MAB:
  - Accelerate BC by selecting for:
    - Donor gene (tightly linked markers)
    - Recurrent parent background (evenly spaced markers)

Earlier recovery of desirable BC line



Conclusion

- Mapping, QTLs, and MAS
  - Tools (among many others) for breeding
  - Do not eliminate need for phenotypic selection:
    - Across locations
    - Across years