

Quantitative Genetics, Genetical Genomics, and Plant Improvement

Bruce Walsh. jbwalsh@u.arizona.edu. University of Arizona.

Notes from a short course taught June 2008 at the Summer Institute in Plant Sciences at Wuhan

Overview of This Material

Plant improvement is one of man's oldest, and most important, technologies. Indeed, the amazing economic growth in China owes much of its success to the green revolution, which allowed a smaller fraction of the population to feed the country, providing the rich labor pool behind much of China's recent growth to an economic superpower.

Quantitative Genetics is the study and analysis of traits whose variation is influenced by both genetic and environmental variation. This includes essentially all agriculturally-important traits, such as yield, lodging, flowering times, etc. The machinery of quantitative genetics, which classically has relied upon **variance components** estimated by using *sets of relatives* underpins most of plant and animal improvement, and well as much of human genetics and evolutionary genetics. The machinery of quantitative genetics, while largely statistical in nature (indeed, much of modern statistics comes from the roots of quantitative genetics) easily incorporates information from molecular markers. Indeed, a major growth area in quantitative genetics has been the search for **QTLs** – **quantitative trait loci** – whose variation influences the variation in a trait of interest. Equally important, although less flashy, has been the development of powerful statistical tools (typically based on **mixed models**) such as BLUP estimation of breeding and combining values, and mixed-model-based approaches for the analysis of complex genotype \times environment interaction.

Genetical genomics is a term coined by Risert Jansen, and is simply the application of quantitative-genetic ideas and techniques to genomic traits, such genome-wide transcription data, the genomic patterns of methylation, and metabolomic/ proteomic data. For all of these traits, we are interested in *variation* – how do these features change over tissues/individuals/populations/species. Such variation is influenced by both genetic and environmental factors. The machinery of quantitative genetics can be used to find QTL influencing these traits, as well as exploring how much of the observed variation is heritable, and how we can exploit this variation for plant improvement.

The propose of this series of lectures in the broader context of this workshop is to provide some background in these tools and ideas. Future lectures will explore the exciting fields of functional genomics in much greater detail. Keep in mind these genomic features (no matter how apparently complex) are simply traits, really no different from classically-measured traits such as yield. The goal is to understand how best to use this information for (i) a deeper understanding of plant biology and (ii) enhancement of plant improvement.

Table of Contents for Lecture Notes

I. Classical Quantitative Genetics	1.1
Nature of Quantitative Variation	1.1
Causes of Genetic Variation: The Multiple-factor Hypothesis	1.2
Variances, Covariances and Regression	1.3
The Variance	1.4
The Covariance	1.4
Covariance and Regressions	1.5
Useful Properties of Variances and Covariances	1.5
Contribution of a Locus to the Phenotypic Value of a Trait	1.5
Fisher's Decomposition of the Genotypic Value	1.6
Average Effects, Additive Genetic and Breeding Values	1.8
Genetic Variances	1.8
Epistasis	1.10
Resemblance Among Relatives	1.11
The Genetic Covariance Between Relatives	1.11
Offspring and one parent	1.11
Half-sibs	1.12
Full-sibs	1.12
General relationships	1.13
Components of Environmental Variances	1.14
Narrow -vs. Broad-sense Heritability	1.14
Defining H^2 for Plant Populations	1.14
General and Specific Combining Abilities: GCA, SCA	1.15
2. QTL Mapping: Overview and Using Inbred Lines	2.1
Experimental Designs	2.1
Conditional Probabilities of QTL Genotypes	2.2
Example: Conditional Probabilities for an F_2	2.2
Expected Marker Means	2.3
Linear Models for QTL Detection	2.4
Maximum Likelihood Methods for QTL Mapping and Detection	2.5
Likelihood Maps	2.6
Permutation Tests: Finding the Significance Threshold	2.7
Precision of ML Estimates of QTL Position	2.7
Interval Mapping with Marker Cofactors	2.8
Power and Repeatability: The Beavis Effect	2.9
Dealing with Supersaturated Models: Model Selection	2.10
Bayesian Approaches	2.11
3. QTL and Association Mapping in Outbred Populations	3.1
QTL Mapping Using Sets of Relatives	3.1
Informative Matings	3.1
QTL Mapping Using Sib Families	3.2

General Pedigree Methods	3.3
Association Mapping	3.4
LD: Linkage disequilibrium	3.4
<i>D'</i> and <i>r</i> ² : Measures of LD	3.5
Fine-mapping Major Genes Using LD	3.6
The Candidate-Gene Approach	3.7
Complications from Population Structure	3.7
TDT, the Transmission/Disequilibrium Test	3.8
Linkage vs. Association	3.9
Dense SNP Association Mapping	3.9
Genomic Control	3.10
Structured Association Analysis	3.11
Regression Approaches	3.12
Structure plus Kinship Methods	3.12
4. Using Molecular Markers to Detect Selection	4.1
Searching for Genes that have Experienced Recent Selection	4.1
Expected Patterns of Variation Under Genetic Drift	4.1
Drift and the Coalescence Process	4.2
Hitch-hiking, Linkage Drag, and Selective Sweeps	4.3
Parameters Associated with Selective Sweeps	4.5
Signatures of Selective Sweeps in Crops	4.6
<i>tb1</i> in Maize	4.6
<i>Waxy</i> in Rice	4.7
The Hill-Roberston Effect: Accumulation of Deleterious mutations in domesticated rice genomes	4.7
Basic Logic of Sequence-Based Selection Tests	4.7
Logic Behind Polymorphism-Based Tests	4.7
Logic Behind Divergence Tests	4.8
Logic Behind Joint Polymorphism and Divergence Tests	4.8
Tests Based Strictly on Within-Population Variation	4.9
The Infinite Alleles Model: Ewen's Sampling Formula	4.10
The Infinite Sites Model	4.10
Tajima's <i>D</i> Test	4.11
Genome-Wide Polymorphism Tests	4.12
The Ghost of Lewontin-Krakauer: Genome Wide <i>F</i> _{ST} -based Scans	4.13
Tests Based on Long Haplotypes	4.13
Ascertainment Bias	4.13
Joint Polymorphism and Divergence Tests	4.14
McDonald-Kreitman Test	4.14
Hudson-Kreitman-Aguade (HKA) Test	4.14
Closing Comments on the Search for Domestication Genes	4.16
5. Classical and Molecular Approaches to Plant Improvement	5.1
Accumulating Favorable Alleles and Genotypes	5.1

The Breeder's Equation	5.1
The Selection Intensity, \bar{i}	5.2
Reducing Environmental Noise: Stratified Mass Selection	5.3
Expanding the Basic Breeder's Equation: Accuracy	5.4
Selection on Multiple Traits	5.6
Background: Matrix Algebra	5.6
The Multivariate Breeders' Equation	5.7
Index selection	5.8
The Smith-Hazel Index	5.10
Marker-Based Selection Methods	5.10
Marker-Assisted Introgression	5.11
MAS: Marker-Assisted Selection	5.11
The Lande-Thompson Index	5.11
Genomic Selection	5.14
Heterosis and Hybrid Breeding	5.14
Heterosis in F_2	5.15
Synthetics	5.15
Agricultural importance of heterosis	5.16
Heteroic Groups and Molecular Markers	5.16
6. Quantitative Analysis of Regulatory Variation	6.1
Gene Regulation is a Complex Trait	6.1
QTLs Involved in Protein Regulation	6.1
Microarrays	6.2
A Brief Overview Of The Technology	6.2
Analysis of Microarray Data	6.3
Microarray Analysis Is Best Regarded As An EDA Approach	6.4
Problems (and Pitfalls) of Gene Discovery via Microarray Analysis	6.5
General Patterns of Transcriptional Variation	6.6
Gene Expression Levels are Typically Highly Heritable	6.6
Correlations Between Rates of Regulatory (Transcriptional) and Sequence Divergence	6.7
Correlations Between Regulatory Divergence and Expression Level/Pattern	6.8
Does Divergence in Expression Follow a Neutral Model?	6.8
Analysis of Pathways	6.9
Two-hybrid screen: Construction Protein-protein interaction maps	6.10
Flux and Pathways	6.10
Kascer-Burns Sensitivity Analysis	6.11
Regulatory Networks and Graph Theory	6.12
Erdos-Renyi Random Graphs and Random Boolean Networks	6.13
Graphs: Small Worlds, Scale-Free, and Power Laws	6.15

Additional Textbooks on Quantitative Genetics

General

Falconer, D. S. and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*, 4th Edition

Lynch, M. and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer.

Roff, D. A. 1997. *Evolutionary Quantitative Genetics*. Chapman and Hall.

Mather, K., and J. L. Jinks. 1982. *Biometrical Genetics*. (3rd Ed.) Chapman and Hall.

Plant Breeding

Acquaah, G. 2007. *Principles of Plant Genetics and Breeding*. Blackwell.

Bernardo, R. 2002. *Breeding for Quantitative Traits in Plants*. Stemma Press.

Hallauer, A. R., and J. B. Miranda. 1986. *Quantitative Genetics in Maize Breeding*. Iowa State Press.

Mayo, O. 1987. *The Theory of Plant Breeding*. Oxford.

Sleper, D. A., and J. M. Poehlman. 2006. *Breeding Field Crops*. 5th Edition. Blackwell

Wricke, G., and W. E. Weber. 1986. *Quantitative Genetics and Selection in Plant Breeding*. De Gruyter.

Statistical and Technical Issues

Bulmer, M. 1980. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press.

Kempthorne, O. 1969. *An Introduction to Genetic Statistics*. Iowa State University Press.

Saxton, A. M. (Ed). 2004. *Genetic Analysis of Complex Traits Using SAS*. SAS Press.

Sorensen, D., and D. Gianola. 2002. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer.