

Quantitative Genetics, Genomics, and the Future of Plant Breeding

Bruce Walsh, University of Arizona

1. Introduction

Quantitative genetics (in its various guises) has been the intellectual cornerstone of plant breeding for close to 100 years. While the roots of Mendelian genetics, and its rediscovery, are firmly in the hands of plant breeders, it was Fisher's (1918) variance decomposition paper that marks the modern foundation for both quantitative genetics and plant breeding. We are now embarking on the age of genomics, and so it is reasonable to speculate on the implications of both partial and whole genome sequences for quantitative genetics. Likewise, the tools of modern quantitative genetics have been developed in four separate fields: plant breeding, animal breeding, human genetics, and evolutionary genetics. Unfortunately, for a variety of reasons, migration of information between these fields has not been what it should be. Thus, it is also an appropriate time to inquire whether useful tools have been developed in these other fields that may be helpful to plant breeders of today and the genomics-based breeders of the near future.

2. Quantitative Genetics in the Age of Genomics

It took just one hundred years to move from the rediscovery of Mendel to the complete sequencing of a higher plant (*Arabidopsis*). Preliminary analysis (The Arabidopsis Genome Initiative, 2000) of the *Arabidopsis* sequence detected 25,500 genes, almost double the number of detected *Drosophila* genes (13,600) and on the same order as the most recent estimates of the number of human genes (25,000-40,000). Roughly 45% of all detected *Arabidopsis* genes are present in four or more copies, so that the 25,500 genes can be classified into roughly 11,600 protein types. While *Arabidopsis* offers a key portal to the genomes of other higher plants, because of their much larger size, whole-genome sequences of

most of the major crops are unlikely to be forthcoming in the near future. Although there is no reason to expect that the basic set of protein types will be any greater than around 12,000, the extensive evolutionary history of polyploidization and segmental duplication in most higher plants (Wendel, 2000) suggests that far more than the 26,000 or so genes of *Arabidopsis* may be present in many of our major crops. While it may seem obvious that genomic information (both current and forthcoming) will have a major impact on quantitative genetics, just how will this information modify quantitative genetics as currently practiced and how profound will the change be?

2.1 Classical vs. Neoclassical Quantitative Genetics

In the classical (Fisherian) quantitative genetics framework, an observed phenotype (y) is regarded as the sum of genetic (g) and environmental (e) effects plus an interaction between genotype and environmental values ($g \times e$),

$$y = u + g + e + g \times e \quad (1)$$

where u is the population mean. In most plant-breeding situations, this basic model usually involves a further decomposition of the environmental effects (for example, into plot, block, temporal and/or location components), but for brevity, we ignore these (easily introduced) extensions in our discussion. Resemblance between relatives, line-cross analysis, and other approaches are then used to estimate the variance components associated with g , e , and $g \times e$ and these are used to predict selection response, the expected values of individuals from defined crosses, and other useful quantities. Classical quantitative genetics has been enormously successful in serving the needs of the plant breeding community, but the use of just a few statistical descriptors (the variance components) to describe the complex underlying genetics tends to leave an uneasy feeling with many of my molecular colleagues. One extreme view put forward by some of these colleagues is that the age of genomics ushers in the death knell for quantitative genetics, as we can move away from a degree of statistical uncertainty to a framework of known genes. While this view is rather naive (and often seems based more on a general aversion to statistics than solid reasoning), it is quite clear that

genomic information does usher in a transformation of quantitative genetics. The classical framework assumes we only know individual phenotypes and the degree of relationship between these individuals. Genomic information allows for an extended (or neoclassical) framework that also incorporates genetic marker information. In particular, under the neoclassical framework, genetic markers provide information on the genotypic value of an individual. If "m" denotes an observed (often multilocus) marker genotype, the basic model now becomes

$$y = u + G_m + g + e + g \times e \quad (2)$$

where G_m is genotypic value associated with this genotype. The classical model (1) is a standard random-effects model, where our interest is in estimating variances associated with g , e , and $g \times e$. By contrast, the neoclassical approach leads to a mixed model, with the G_m regarded as fixed effects. Genotype-environment interactions involving G_m can be incorporated into (2), as can potential epistatic interactions between a particular marker genotype and the background genes ($G_m \times g$). In the neoclassical framework, the marker genotype effect G_m can be directly estimated as a fixed effect. As we detail below, genomic information may be able to suggest particular genes as consideration for candidates, so that G_m is the genotypic value associated with the multilocus genotypes at the candidates. The percentage of the total phenotypic variance accounted for by the candidates provides a direct measure of their importance. The hope expressed by my molecular colleagues is that the variance accounted for by the classical part of this model ($g + e + g \times e$) is small relative to the variance accounted for by the marker information (G_m).

Even in situations where the known genotypes indeed account for a large fraction of trait variance, the importance of particular genotypes may be quite fleeting. Genotypic values for particular loci are potentially functions of the background genotypes and environments, and hence can easily change as crops evolve and as the biotic (pests and pathogens) and abiotic (farming practices) environments change. Further, mutation will generate new quantitative trait loci (QTLs), and a candidate locus that works well in one population may be (at best) a very poor predictor in another. Even if only a modest number of QTLs influence a trait,

then (apart from clones) each individual is essentially unique in terms of its relevant genotypes and the particular environment effects it has experienced. If epistasis and/or genotype-environment interactions are significant, any particular genotype may be a good, but not exceptional, predictor of phenotype. Quantitative genetics provides the machinery necessary for managing all this uncertainty in the face of some knowledge of important genotypes. Indeed, variance components allow one to quantify just how much of the variation is unaccounted for by the known genotypes. A critical feature of quantitative genetics is that it allows for the proper accounting of correlations between relatives in the unmeasured genetic values (g).

We don't mean to paint an overly harsh view of the importance of being able to identify key genotypes. Rather, we simply wish to introduce a little caution to dampen completely unrestrained enthusiasm. It is clear that there are enormous benefits to being able to predict even a fraction of an individual's genotypic value, given a set of genetic markers. For example, even a small increase in the probability of fixation of an advantageous allele during inbreeding to form pure lines can have a dramatic effect. Suppose an F_1 population is segregating favourable alleles at 10 loci, and we first inbred to fixation and then select among lines. In the absence of selection, the probability of fixation of any single favourable allele is 0.5. A quick binomial calculation shows that a sample size of 2,357 is required to have a 90 percent probability that at least one line contains all favourable alleles. If we are able to increase the probability of fixation by only 50 percent (to 0.75), only 40 individuals are required, roughly a 60-fold reduction. If 20 favourable alleles are segregating, the reduction is 3330-fold (from 2,414,434 to 725). Likewise, with known genotypes in hand, searches for $G \times E$ are much more direct, allowing for the possibility of searching for major genes that are highly adaptive to specific environments.

2.2 Genomics and Candidate Loci

Much of the above discussion of a more generalized view of quantitative genetics has assumed that we know the genotypes (and their effects) at a number of QTLs. Given that very few QTLs have been fully isolated, we are still far from achieving this goal. At present, the genotypes (m) scored in Equation (2) usually consist of anonymous markers shown by statistical association to be linked to QTLs. Such marker-assisted selection can result in a significant improvement of the selection response, particularly when the heritability of a character is low (Lande and Thompson, 1990). Given that even a QTL of large effect is typically only initially localized to a region of around 20-50 cM, anonymous markers can easily be 10 to 25 (or more) cM from the actual QTL. Selecting directly on the QTL genotypes (as opposed to linked markers) increases the efficiency of selection, unless the marker is very tightly linked to the QTL. The relative efficiency of a single generation of selection on linked markers (as opposed to directly selecting on the genotypes) scales as $(1-2c)^2$ or roughly $1-4c$ for a tightly linked marker (where c is the recombination frequency between the marker and QTL).

Hence, we would like to be able to at least localize more tightly linked markers, and ideally, screen potential candidate loci directly to see if they are the QTLs. Direct tests of association with a small set of candidates is a more powerful approach than a genome-wide screen using a set of anonymous markers. The difficult issue is selecting the candidates in the first place, and the hope (indeed often the core assumption) of genomics is that the full genome sequence will, in time, greatly facilitate the selection of candidate genes. A variety of genomics tools (reviewed below) has indeed been suggested to help in the search for candidates. At present, the use of these tools is generally restricted by economical, rather than biological, constraints. One of the major trends expected over the next decade will be to make these tools economically feasible for just about any trait or crop of interest.

2.2.1 Basic Genomic Tools

Perhaps the single most useful tool is dense marker maps. It is these maps that allow for QTL mapping, association studies, and marker-assisted introgression, to name just a few uses. The most obvious tool is the whole genome sequence. With the complete sequence in hand (or even a partial sequence highly enriched for coding sequences), one can construct any number of DNA chips. These microarrays containing a large number of defined DNA sequences can be used for screening the expression of a large number of genes (via hybridization) in particular tissues (expression array analysis), for probing a related genome for homologous genes of interest, and many other interesting possibilities we are only beginning to consider. Besides faster and cheaper sequencing, a major factor facilitating future genomic projects is the ability to use sequence homology to bootstrap from a model system to a related species. For example, we can use sequences from *Arabidopsis* to probe for homologous genes in related species, and any recovered sequences can be, in turn, used to design probes for even more distantly related species. Given the tendency for many plant genes to exist in large multigene families, the advantage of a full genome sequence is that all members of a particular family can be used as probes, increasing the chance of identifying at least one homologous gene from a related species. Once again, any recovered homologues can themselves be used as probes for other family members within the target species. Thus, centered around a key model system, we can imagine the search for homologous genes spreading out in phylogenetic space like ripples on a pond, reaching ever more phylogenetically distant species.

2.2.2 Prediction of Candidate Genes

With a genomic sequence in hand, either from the plant of interest or from a sufficiently close relative, how can one use this information to find possible candidate genes? The most straightforward approach is to search the genome for sequences with homologies to known candidate genes from another species. For example, a gene known (say) to create variation in plant architecture in maize (*Zea mays* L.) can be used to probe related grasses. If homologues can be found, association tests between trait values and variation in the potential candidate(s) can be performed. A more brute-force approach is to first limit a QTL to a confidence region

(as small as possible) and then use the genomic sequence from that region to either suggest candidates for further testing (see below) or by simply screening all the genes in this region using expression arrays to search for those whose expression pattern is consistent with the character of interest. Even in such apparently direct expression studies, some caution is in order. For example, a gene turned on in seed tissue is certainly a candidate for yield, but another gene expressed only in root tips (and hence likely excluded from further consideration) may have a more important effect on yield if it increases the plant's ability to gather and store energy.

More generally, the hope is that the DNA sequence itself may provide clues to its potential for a candidate. The growing field of proteomics has generated an extensive catalog of known protein motifs, offering the possibility of making some (albeit crude) deductions about the functions of particular open reading frames, such as whether the resulting protein spans a membrane, is involved in transport, is directed to a particular organelle, etc. Even such partial information may be informative in keeping or excluding potential candidates. The hope for the future is that we will be able to read the regulatory sequences to deduce the expression pattern of a gene directly from its DNA sequence, in essence, during the array expression studies *in silico* (via the computer). Again, the above-mentioned caveat (that expression in a very different, and unexpected, tissues may have a dramatic effect on the character of interest) holds.

2.2.3 Transgenics

The tool from biotechnology that perhaps excites breeders the most is the ability to construct transgenics, importing a novel gene into an organism. While transgenics can be constructed for many crop species, their phenotypes are rather unpredictable. Insertion of a new sequence cannot currently be targeted to specific sites, but rather is largely random, with the location of insertion significantly influencing the level of gene expression. Further, plants often suppress multiple-copy genes, which can further impact a transgenic. Even when these issues are

resolved, it is clear for the foreseeable future that the transgenic technology is restricted to importing genes of major effect. The success in improving a character by importing a modest to large suite of genes of smaller individual effects (but perhaps a great cumulative effect) is less certain, given the above concerns about consistency of expression. However, this could also work in the breeder's favour in that a gene of modest effect may have a more dramatic effect than that expected due to position effect. A further complication is that the introduction of genes of large effect (perhaps generated by using a high expression promoter on a gene of otherwise modest effect) can often have significant pleiotropic consequences on a number of characters besides the target, and, hence, can reduce crop performance in other aspects. Selection for lines possessing modifiers to reduce any associated deleterious effects is, thus, a key step in the improvement of an initial transgenic line. Quantitative-genetic machinery can suggest those lines with the greatest potential for modifiers, for example, by searching for lines with large $G_m \times g$ interactions in favour of less deleterious side effects.

2.3 Fishing for Useful Variation in Natural or Weakly Domesticated Populations

The area where genomics may eventually offer the largest payoff to breeders is in the search of useful genes in natural and/or weakly domesticated populations. The source populations or species from which modern crops descend harbour far more genetic diversity than is present in the limited set of highly domesticated lines currently in use for food production. The ability to localize genes of significant effect, and, subsequently, introgress these into cultivars without generating undesirable side effects on performance, is a key aim of breeders. As we detail in the next section, a variety of useful approaches for searching natural populations for genes of interest have been developed in other fields of quantitative genetics.

3. Useful Tools from Other Fields of Quantitative Genetics

The board arena of quantitative genetics consists of four rather distinct fields --- plant breeding, animal breeding, evolutionary genetics, and human genetics (it could be argued that

tree breeders form a fifth field). Although all four draw upon the basic foundations of quantitative genetics, each has rather distinct, and often non-overlapping, literatures, and the information flow across the fields has often been rather restrictive. One consequence of this restricted flow is that approaches for a specific problem are often independently reinvented. A more interesting consequence is that since practitioners in each of the fields are faced with unique issues and constraints, each has developed a number of useful tools that are (unfortunately) often not widely known to outsiders. We (Lynch and Walsh, 1998; Walsh and Lynch, 2002) have recently tried to bring all these tools and approaches together into a unified general framework for quantitative genetics. Since many of these field-limited tools are both largely unknown, and yet of potential interest, to plant breeders, I conclude by briefly reviewing a few of the more promising approaches (especially those with applications at the quantitative genetics - genomics interface). Extension of some of these approaches, to be of value to plant breeders, may require some nontrivial modifications.

3.1 Plant Breeding

For starters, it is useful to first remind plant breeders of some of the tools they routinely use that are not well known (or at least not widely appreciated) to geneticists outside of the field. As a consequence of having to deal with a diversity of mating systems (most importantly selfing) and sessile individuals, issues that plant breeders tend to focus on more than other quantitative geneticists include the creation and selection among inbred lines and their hybrids, $G \times E$, and competition. Some important tools have already migrated from plant breeders to quantitative genetics as a whole. One example is line-cross based analysis (generation-means, diallels), which has seen an increasing use in evolutionary genetics. Somewhat surprisingly, many quantitative geneticists have been a little slow in drawing upon the wealth of field-plot designs, especially analyses for dealing with $G \times E$, that plant breeders have accrued. For example, while additive main effects multiplicative interactions, or AMMI, models (Gollob, 1968; Mandel, 1971; Gauch 1988, 1992; Zobel et al., 1988; Gauch and Zobel, 1989) and biplots (Gabriel, 1971; Kempton, 1984) have become important tools for plant breeders (as

several papers in this symposium illustrate), they are generally unknown outside the field. The correct formulation for the covariance between relatives under inbreeding (e.g., Cockerham, 1983) is another important tool developed by plant breeders that has remained largely unappreciated (but see Abney et al., 2000).

3.2 Animal Breeding

Animal breeders face designs involving complex pedigrees, large half-sib or (more rarely) full-sib families, long life spans, and overlapping generations (many of these same issues are faced, to an even greater extent, by tree breeders). The machinery of predicting breeding values by best linear unbiased prediction, or BLUP (reviewed by Henderson, 1984; Mrode, 1996; Lynch and Walsh, 1998) and the estimation of variance components by restricted maximum likelihood estimation, or REML (reviewed by Searle et al., 1992; Lynch and Walsh, 1998) have been developed by animal breeders to address these concerns. BLUP/REML easily allow for arbitrary pedigrees (through specification of appropriate relationship matrices) and for the estimation of a large number of fixed factors. This BLUP/REML framework is a very appealing one from a genomics standpoint, as scored genotypes of interest can be treated as fixed effects, and complex (fixed and/or random) models with both background genotypes and structured environmental effects can also be introduced.

A second area that may be of interest to plant breeders is the extensive work of animal breeders on maternal effects designs (e.g., Lynch and Walsh, Chapter 23). Although several of these designs are not easily transferred to plant breeding systems (some are based on cross-fostering offspring), they, nonetheless, are useful reading when thinking about the importance of maternal effects, a topic that often seems overlooked by plant breeders. The widespread availability of cloned individuals can greatly facilitate the estimation of maternal effects, and hence a determination of their importance. Recent theoretical work on the quantitative-genetic implications of endosperm by Shaw and Waser (1994) is a related topic of interest.

Finally, a major push towards the use of Bayesian methods of analysis is coming from the animal breeders (e.g., Gianola and Fernando, 1986). Just as likelihood methods replaced method-of-moments and other estimators when they became computationally feasible in the mid-late 1970s, a variety of Markov Chain Monte Carlo simulation approaches (such as the Gibbs sampler) have allowed Bayesian posteriors to be computed for even very complex models (Geyer, 1992; Tierney, 1994; Tanner, 1996). The very appealing feature of a Bayesian analysis is that a marginal posterior distribution incorporates all the uncertainties introduced by having to estimate other parameters of less interest. For example, a model that estimates the additive genetic variance also must estimate a number of other variance components and fixed effects. The marginal posterior for the additive variance naturally incorporates all the uncertainty introduced by having to estimate these additional nuisance parameters. Bayesian analysis provides a powerful framework for analysis for the expected growing complexity of neoclassical models.

3.3 Evolutionary Genetics

As the search for potentially useful genes moves to natural populations, machinery from evolutionary and population genetics may prove useful. The issues of concern to evolutionary geneticists involve estimating the nature and amount of selection on a defined suite of characters and the population genetics of evolution. Three useful developments from this field may be of interest to plant breeders.

First, methods for estimating the nature of natural selection on any characters of interest have been developed (Lande and Arnold, 1983; Arnold and Wade, 1984a,b; Schluter, 1988; Crespi and Bookstein, 1989; Schluter and Nychka, 1994; Willis, 1996). This machinery allows the breeder to estimate the nature of natural selection on any measurable suite of characters, separating selection into direct and indirect effects (due to selection on correlated characters). A detailed understanding of the nature of natural selection in either wild or

domesticated populations can provide the breeder with valuable insight into characters that can further improve performance.

Second, there is a rich literature from population genetics dealing with detection of selection from a population sample of DNA sequences (reviewed by Kreitman, 2000). An interesting application of these methods was the finding of reduced levels of polymorphism (consistent with directional selection) in the 5' control region of the *teosinte-branched 1* gene involved in major morphological differences between teosinte (*Zea mexicana* L.) and domesticated maize (Wang et al., 1999). With a collection of candidate genes in hand, one can search for signatures of selection in homologues from natural populations. Much of the theory underlying tests of selection follows from the explosive development of coalescent theory (reviewed by Hudson, 1991; Tavaré and Balding, 1995; Fu and Li, 1999), which describes the genealogy (the distribution of the times to a common ancestor) for a random sample of a particular DNA sequence from the population. There are obvious extensions of this theory to deal with issues of concern to quantitative geneticists, such as estimating the degree of relationship based on molecular data and the fine-mapping of QTLs using very tightly linked markers (e.g., Slatkin, 1999; Zollner and von Haeseler, 2000).

Finally, there has been considerable progress in the theoretical analysis of finite locus models (as opposed to the traditional infinitesimal models routinely used by breeders), and these developments are reviewed in Burger (2000). In particular, the response to selection when the underlying distribution of genotypic (or breeding) values is not Gaussian has received significant attention (Barton and Turelli, 1987; Turelli, 1988; Turelli and Barton, 1990, 1994). Such developments in finite-locus models provide a useful framework for predicting selection response when partial genotypic information is available.

3.4 Human Genetics

The final field of quantitative genetics from which plant breeders may wish to draw upon is developments in human genetics, where small family sizes and a lack of controlled mating designs are common occurrences. Despite these obvious limitations, human geneticists have been rather successful at mapping genes, and some of their tools may prove useful to plant geneticists, especially when trying to isolate genes of interest from natural or weakly domesticated populations for which defined inbred lines may not be available.

One powerful approach has been to use sib-pairs to map QTLs (reviewed and extended by Abel and Muller-Myhsok, 1998; Monks et al., 1998; McPeck, 1999; and Elston and Cordell, 2001), and these approaches can be applied to the offspring from single plants in natural populations (although suitable modifications would have to be introduced to account for selfing). One complication that both human geneticists and plant breeders working with natural populations face when attempting association studies (between candidate genotypes and trait values) is that false positives can be created by population substructure (or stratification). For example, if a marker is very common in a particular subpopulation, and that subpopulation also carries alleles for a trait of interest at high frequencies, then if the population structure is not accounted for, the marker can show an association with the trait simply by being a predictor of the population from which an individual is drawn. Human geneticists account for any potential population structure by using the transmission-disequilibrium test (or TDT) that compares whether an allele is transmitted or not transmitted from a parent to an offspring showing the trait of interest (Spielman et al., 1993; Knapp, 1999a,b). Another powerful tool of human geneticists is fine-mapping of genes by linkage disequilibrium, using the historical recombinations (as reflected in the decay of disequilibrium) that occur between a tightly linked marker and a gene of interest to fine-map that locus (Hastbacka et al., 1992; Graham and Thompson, 1998; Slatkin, 1999).

A final important tool with its roots in human genetics is random-effects models to map QTLs in complex pedigrees (e.g., Amos, 1994; Gessler and Xu, 1996; Xie et al., 1998; Yi and Xu, 1999, 2000). The idea behind a random-effects model is to simply estimate the trait

variance associated with any particular genomic region using anonymous markers than span the genome. As with BLUP/REML, this approach can accommodate both arbitrary pedigrees and numerous fixed effects. It is certainly an approach to consider for QTL mapping in many settings.

4. Conclusions

The age of genomics is a very exciting time for quantitative geneticists. While the view is often suggested that genomics will reduce the importance of quantitative genetics, in fact the opposite is true. Straightforward modifications of classical quantitative genetic models provide the natural framework for handling both phenotypic and genotypic information. Equally important for breeders to consider are powerful tools developed in other fields of quantitative genetics, only a few of which have been discussed here.

5. References

Abel, L. and Muller-Myhsok, B. (1998) Robustness and power of the maximum-likelihood-binomial and maximum-likelihood-score methods, in multiple linkage analysis of affected-sibship data. *American Journal of Human Genetics* 63: 638-647.

Abney, M., McPeck, M. S. and Ober, C. (2000) Estimation of variance components of quantitative traits in inbred populations *American Journal of Human Genetics* 66: 629-650.

Amos, C. I. (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. *American Journal of Human Genetics* 54:535-543.

Arnold, S. J. and Wade, C. (1984a) On the measurement of natural and sexual selection: theory. *Evolution* 38: 709-719.

Arnold, S. J. and Wade, C. (1984b) On the measurement of natural and sexual selection: applications. *Evolution* 38: 720-734.

Barton, N. H. and Turelli, M. (1987) Adaptive landscapes, genetic distances and the evolution of quantitative characters. *Genetical Research* 49: 157-173.

- Burger, R. (2000) *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, New York, 409 pp.
- Cockerham, C. C. (1983) Covariances of relatives from self-fertilization. *Crop Science* 23: 1177--1180.
- Crespi, B. J. and Bookstein, F. L. (1989) A path-analytic model for measurement of selection on morphology. *Evolution* 43: 18-28.
- Elston, R. C. and Cordell, H. J. (2001) Overview of model-free methods for linkage analysis. *Advances in Genetics* 42: 135-150.
- Fisher, R. A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52: 399-433.
- Fu, X.-Y. and Li, W.-H. (1999) Coalescing into the 21st century: An overview and prospects of coalescent theory. *Theoretical Population Biology* 56: 1-10.
- Gabriel, K. R. (1971) Biplot display of multivariate matrices with applications to principal component analysis. *Biometrika* 58: 453-467.
- Gauch, H. G., Jr. (1988) Model selection and validation for yield trials with interaction. *Biometrics* 44: 705-715.
- Gauch, H. G., Jr. (1992) *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*. Elsevier, Amsterdam, the Netherlands, 278 pp.
- Gauch, H. G., Jr. and Zobel, R. W. (1988) Predictive and postdictive success of statistical analysis of yield trials. *Theoretical and Applied Genetics* 76: 1-10.
- Geyer, C. J. (1992) Practical Markov chain Monte Carlo (with discussion). *Statistical Science* 7: 473--511.
- Gessler, D. D. G. and Xu, S. (1996) Using the expectation or the distribution of identical-by-descent for mapping quantitative trait loci under the random model. *American Journal of Human Genetics* 59:1382-1390

- Gianola, D. and Fernando, R. L. (1986) Bayesian methods in animal breeding theory. *Journal of Animal Science* 63: 217--244.
- Gollob, H. F. (1968) A statistical model which combines features of factor analysis and analysis of variance techniques. *Psychometrika* 33: 73-115.
- Graham, J. and Thompson, E. A. (1998) Disequilibrium likelihoods for fine-scale mapping of a rare allele. *American Journal of Human Genetics* 63: 1517-1530.
- Hastbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. and Lander, E. (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics* 2: 204-211.
- Henderson, C. R. (1984) *Applications of Linear Models in Animal Breeding*. Univ. Guelph, Guelph, Ontario, 462 pp.
- Hudson, R. R. (1991) Gene genealogies and the coalescent process. In: Futuyama, D. J. and Antonovics, J. (Eds) *Oxford Surveys in Evolutionary Biology*. Oxford University Press, Oxford, pp. 1-44
- Kempton, R. A. (1984) The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science* 103: 123-135.
- Knapp, M. (1999a) The transmission/disequilibrium test and parental-genotype reconstruction: The reconstruction-combined transmission/disequilibrium test. *American Journal of Human Genetics* 64: 861-870.
- Knapp, M. (1999b) A note on power approximations for the transmission/disequilibrium test. *American Journal of Human Genetics* 64: 1177-1185.
- Kreitman, M. (2000) Methods to detect selection in populations with application to the human. *Annual Review of Genomics and Human Genetics* 1: 539-559.
- Lande, R. and Arnold, S. J. (1983) The measurement of selection on correlated characters. *Evolution* 37: 1210-1226.
- Lande, R. and Thompson, R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743--756.

Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA, 980 pp.

Mandel, J. (1971) A new analysis of variance model for non-additive data. *Technometrics* 13: 1-8.

McPeck, M. S. (1999) Optimal allele-sharing statistics for genetic mapping using affected relatives. *Genetic Epidemiology* 16: 225-249.

Monks, S. A., Kaplan, N. L. and Weir, B. S. (1998) A comparative study of sibship tests of linkage and/or association. *American Journal of Human Genetics* 63: 1507-1516.

Mrode, R. A. (1996) *Linear models for the prediction of animal breeding values*. CAB International, Wallingford, UK, 187 pp.

Schluter, D. (1988) Estimating the form of natural selection on a quantitative trait. *Evolution* 42: 849-861.

Schluter, D. and Nychka, D. (1994) Exploring fitness surfaces. *American Naturalist* 143: 597--616.

Searle, S. R., Casella, G. and McCulloch., C. E. (1992) *Variance Components*. John Wiley and Sons, NY, 501 pp.

Shaw, R. G. and Waser, N. M. (1994) Quantitative genetic interpretations of postpollination reproductive traits in plants. *American Naturalist* 143: 617-635.

Slatkin, M. (1999) Disequilibrium mapping of a quantitative-trait locus in an expanding population. *American Journal of Human Genetics* 64: 1765-1773.

Spielman, R. S., McGinnis, R. E. and Ewens, W. J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* 52: 506-516.

Tanner, M. A. (1996) *Tools for Statistical Analysis*, 3rd edn. Springer-Verlag, New York, 207 pp.

- Tavare, S. and Balding, D. J. (1995) Coalescents and genealogical structure under neutrality. *Annual Review of Genetics* 29: 410-421.
- The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815.
- Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* 22: 1701--1762.
- Turelli, M. (1988) Population genetic models for polygenic variation and evolution. In: Weir, B. S., Eisen, E. J., Goodman, M. M. and Namkoong, G. (eds.), *Proceedings of The Second International Conference on Quantitative Genetics*, Sinauer Associates., Sunderland, MA, pp. 601-618.
- Turelli, M. and Barton, N. H. (1990) Dynamics of polygenic characters under selection. *Theoretical Population Biology* 38: 1-57.
- Turelli, M. and Barton, N. H. (1994) Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? *Genetics* 138: 913-941.
- Walsh, B. and Lynch, M. 2002. *Evolution and Selection of Quantitative Traits*. Sinauer Associates, Sunderland, MA (In prep). Draft chapters can be found on the web at http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html
- Wang, R.-L., Stec, A., Hey, J., Lukens, L. and Doebley, J. (1999) The limits of selection during maize domestication. *Nature* 398: 236-239.
- Wendel, J. G. (2000) Genome evolution in polyploids. *Plant Molecular Biology* 42: 225-249.
- Willis, J. H. (1996) Measures of phenotypic selection are biased by partial inbreeding. *Evolution* 50: 1501-1511.
- Xie, C., D. Gessler, D. G. and Xu, S. (1998) Combining data from different line crosses for mapping quantitative trait loci using the identical-by-descent based variance component method. *Genetics* 149:1139-1146.
- Yi, N. and Xu, S. (1999) A random model approach to mapping quantitative trait loci for complex binary traits in outbred populations. *Genetics* 153:1029-1040.

Yi, N. and Xu, S. (2000). Bayesian mapping of quantitative trait loci under the IBD-based variance component model. *Genetics* 156:411-422

Zobel, R. W., Wright, M. J. and Gauch, H. G., Jr. (1988) Statistical analysis of a yield trial. *Agronomy Journal* 80: 388-393.

Zollner, S. and von Haeseler, A. (2000) A coalescent approach to study linkage disequilibrium between single-nucleotide polymorphisms. *American Journal of Human Genetics* 66: 615--628.