Theoretical Population Biology **59**, 175–184 (2001)
doi:10.1006/tpbi.2001.1512, available online at http://www.idealibrary.com on IDEAL®

MINIREVIEW

Quantitative Genetics in the Age of Genomics

Bruce Walsh

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721
E-mail: jbwalsh@u.arizona.edu

Received November 2, 2000

C. elegans Genome Project	... $100	million
Drosophila Genome Project	... $1	billion
Human Genome Project	... $10	billion
Working knowledge of multivariate statistics	...	Priceless

—T-shirt designed by Mike Wade for the Evolution 2000 meetings

Mike Lynch and I recently attempted to summarize the current state of quantitative genetics (Lynch and Walsh, 1998; reviewed for this journal by Baston, 2000). Already, parts of our treatment are somewhat dated due to the explosive growth and refinement of methods for mapping loci underlying complex traits (QTLs, for quantitative trait loci). While such growth is bad news for textbook writers trying to stay current, it is also the hallmark of a scientifically healthy and active field. As the age of genomics ushers in, continued dramatic changes in the field of quantitative genetics are expected. The faster–cheaper trend for sequencing and genotyping that fuels genomics has a parallel trend in computing, and this will also have a significant impact on the field. Finally, a variety of other emerging biotechnologies in such areas as reproductive biology (whole organism cloning, embryo transplantation) and recombinant DNA (transgenic organisms) will also have important consequences.

Given the availability of these new biological and computational tools, what will quantitative genetics become? In particular, if we wrote a second edition in (say) 2020, how much will the field have changed? Ultimately, any attempt to predict the long-term directions of a field results in a paper that future workers will find both somewhat amusing and hopelessly misguided. This caveat being stated for the record, let us press on.

The machinery of quantitative genetics is widely applied in such diverse fields as human genetics, evolution, and breeding. While the broad goals of these different users are the same (determining how genetic and environmental factors contribute to the observed variance, either within or between populations, of particular traits), their specific goals are different. Human geneticists are concerned with finding disease-susceptible genotypes and their associated risk factors (such as particular genotype–environment interactions). Plant and animal breeders are concerned with maximizing selection response and stability. Evolutionary geneticists are concerned with the genetic architecture of particular traits and inferring both their past evolutionary history and their potential for future evolution. Classical quantitative genetics has admirably served these different users in the past. How will it fare in the future and how will it be different?

CLASSICAL VS NEOCLASSICAL QUANTITATIVE GENETICS

The strength of classical (Fisherian) quantitative genetics is that one can use the variance component machinery developed by Fisher (1918) to analyze essentially any character in any organism. All that is needed is a collection of known relatives (often not a trivial task to obtain) and some very general knowledge about the genetic and mating systems (such as ploidy level and whether the species is an outcrosser or a selfer). Under

0040-5809/01 $35.00
Copyright © 2001 by Academic Press
All rights of reproduction in any form reserved.

classical quantitative genetics, we observe only the phenotypic value z of an individual, which we regard as the sum of an unseen genetic (g) and environment (e) value,

$$z = g + e \tag{1}$$

Using the covariance structure for the phenotypic resemblances between sets of known relatives, we can decompose the genotypic value into further components (e.g., $g = a + d + i$, an additive, dominance, and interaction term) and subsequently estimate the variance associated with each of these components. The flexibility inherent in classical quantitative genetics arises because (for many problems) knowledge of just the variance components is sufficient, with any further knowledge of the fine genetic details being largely irrelevant. For example, variance components are sufficient to predict the short term response to selection and to estimate the increase in disease risk for different sets of relatives. This focus on variance components has left most of my molecular colleagues with an uneasy feeling about quantitative genetics, as the notion of having one's experimental analysis be heavily dependent on statistics has been anathema to many molecular biologists (Indeed, I had a colleague who cheerily informed students in my undergraduate class that "If you need statistics to analyze an experiment, you have not designed it correctly"). This is most unfortunate because if we indeed knew all the fine details (for example, the values of all relevant genotypes and their population frequencies), in many cases we would first translate these into genetic variance components before using this information to make population-level predictions. This classical world of quantitative genetics where genotypes are unknown is dominated by random effects models, where any particular genotype (and its associated genotypic value g) is assumed to be a random draw from a population and our interest is in estimating the variance components of g.

In the emerging, more generalized framework for quantitative genetics, at least some genotypes at loci influencing trait variation are assumed known. Under this new framework (which I will refer to as neoclassical quantitative genetics), we know both the phenotype z and the multilocus genotype m for specific genes of interest. The recorded genotypes are either actual segregating sites contributing to character variance or markers tightly linked to these sites. If G_m is the genotypic value associated with this genotype, then the simplest model is

$$z = G_m + g + e \tag{2}$$

We can estimate the genotypic value G_m associated with m (if unknown) by treating it as a standard fixed effect. Since G_m likely depends on the distribution of background genotypes and environments in the population of interest, we expect its value to change (and thus have to be re-estimated) as the population of interest changes. The genotypic value contributed by unmeasured genes is given by g, whose variance components are estimated in the standard Fisherian framework. This basic model can easily be expanded to accommodate a known-genotype x background-genotype epistatic interactions by adding a $G_m \times g$ term. Since g is a random effect, this term is also random, and we can estimate its associated variance (provided the genotype m is sufficiently frequent in our sample) again by standard Fisherian approaches. Similar modifications are also easily envisioned for other potential interactions between G_m, g, and e.

The view of many of my molecular colleagues is that if the appropriate set of genes are known, then G_m accounts for a very significant fraction of the total genetic variance, and hence quantitative genetics (the contribution g) just deals with trivial residual variation. This presupposes that G_m is known either without error or has a very small associated error variance. It further supposes the stability of G_m across individuals. In fact, G_m is the expected value for the genotype m over the distribution of background genotypes and environments. If the loci scored for m have either significant epistatic interactions with the background genes and/or genotype-environment interactions, this can generate significance variance in G_m around its expected value (most of this error would be incorporated into e, although contributions from additive epistatis would be incorporated into g).

These important caveats aside, then my molecular colleagues are partly correct in their assertion. Suppose the phenotypic trait of interest has not yet been displayed in an individual. For example, we are trying to determine the probability that an individual will get a late-onset disease, so that z is eventually coded as either 0 (disease free) or 1 (diseased). In the classical framework, we would use information on the disease status of relatives of this individual coupled with estimates of variance components to estimate the risk. In the neoclassical framework, we could look directly at the genotype m for candidate disease risk genes and use this to predict z. Clearly, if most of the genetic variation is accounted for by G_m, then inclusion of g (via variance components) will have little effect on our prediction. For many problems of interest, however, even if we were able to identify all the relevant genes underlying a trait, we would still not know the genotype of particular individuals. For example,

suppose we are trying to predict the phenotypic values in the offspring from two parents, as would occur if we are trying to predict response to selection or to inform a couple about the disease risk their offspring face. Once such individuals are realized, we can genotype them, but before their realization, we are strongly in the realm of quantitative genetics. Let m_i and m_j be the (potentially multilocus) genotypes of the two parents. Given these known genotypes, the expected frequencies of gametes for each are easily computed (even when loci are linked), and the resulting distribution of genotypic values (over the known loci) in their offspring is given by

$$G(x_i, x_j)\, p(x_i \mid m_i)\, p(x_j \mid m_j) \qquad (3)$$

where $G(x, y)$ is the function giving the resulting genotypic value for an individual formed by gametes of types x and y, and $p(x \mid m)$ is the distribution of gametes of type x given a parent of type m. Under the classical framework, it is generally assumed that genotypic values are normally distributed and hence the offspring distribution is completely specified by the mean and variance (mean vector and covariance matrix in the multivariate case). Assuming the infinitesimal model, the mean is the average breeding value of the parents and the segregation variance is a constant value, independent of the parents.

When genotypes are indeed known, the segregation kernel from any given family (Eq. 3) allows for distributions that depart significantly from normality. In such cases, variance components alone are not sufficient to describe the complete distribution. Note from Eq. (3) that the segregation variance (the offspring variance of G around the offspring mean) is clearly a function of the parental genotypes. With known genotypes, the segregation variance is heteroscedastic (taking different values for different parental genotypes), in contrast to classical quantitative genetics which assumes a homoscedastic segregation variance (a constant value, independent of parental genotypes). This is obviously a point to be exploited by breeders, with some families having a smaller offspring variance, and hence being more predictable, than others. Likewise, families with larger than average segregation variance are of importance to breeders trying to select extreme individuals.

Moving from this fictitious world, where all relevant genotypes are known, to the real world where (at best) only a subset are known, quantitative genetics further increases in importance. As mentioned, the genotypic values for particular loci are potentially functions of the background genotypes and environments (when epistatis and/or genotype–environment interactions are present). Furthermore knowledge of important genotypes is expected to be fleeting. Mutation will generate new QTLs, and a candidate locus that works well in one population may be a very poor predictor (at best) in another (a particuarly interesting example is given by Winkelman and Hodgetts, 1992, who examined candidate locus association in a mouse line before and after selection). Indeed, if even 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. Variance components allow one to quantify just how much of the variation is accounted 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).

Ironically, with the age of genomics we are now at the point where it may be much easier to extract and sequence potential candidates genes from an organism of interest than it is to obtain a collection of known relatives. Under neoclassical quantitative genetics, we are potentially less dependent on known sets of relatives, replacing them to some extent with known genotypes at particular candidate loci. We can also use marker information to make inferences about the degree of relatedness between sampled individuals, and this information can be used in a classical quantitative genetics framework to estimate the associated variance components. The downside to this approach is that the estimation error for degree of relatedness rapidly increases once we pass beyond first and second degree relatives.

GENOMICS, MODEL SYSTEMS, 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 from natural or human populations have been fully isolated, we are still very far from achieving this goal. An often stated hope is that genomics projects will provide candidate loci, as candidate–trait association tests are inherently more powerful than marker-based whole genome scans for QTLs. Even with a list of such potential candidates in hand, one is still

faced with the difficult task of first testing for candidate genotype-trait associations, and then demonstrating that any such associations are not sampling artifacts. We will examine these two issues—the impact of genomic projects on sharpening the search for potential candidates and the subsequent tests for association—in turn.

While a variety of tools are currently available to search for candidates, at present they are restricted by economical (rather than biological) constraints to important traits in just a few organisms. One of the major trends expected over the next decade will be to make these tools economically feasible for just about any organism of interest. Clearly, the trend toward fast and cheaper genotyping and sequencing will continue, as will the continued expansion of newer technologies such as expression chips. Simply extrapolating this trend out (a dangerous thing to do, for past history suggests we will likely *underestimate* it), genomic projects for one's favorite organisms become a reality in the immediate future. With a genomic sequence in hand, the ability to score individuals rapidly (and cheaply) at a very large number of loci also becomes quite feasible.

What are the basic off-the-shelf genomic tools currently available in model systems (e.g., *Drosophila*, yeast, *C. elgans*, and humans)? Perhaps the most useful single tool to quantitative geneticists are dense marker maps, usually constructed from single nucleotide polymorphisms (SNPs) and/or microsatellites. It is these maps that allow for QTL mapping, association studies, and estimation of the relatedness between individuals in a random sample from the population. Complete genomic sequences offer the hope of fishing for candidate genes simply based on sequence information (we will have more to say about this shortly). With a complete genome sequence in hand, one can construct any number of DNA chips—microarrays of a large number of chosen DNA sequences for looking at gene expression in particular tissues (expression array analysis), probing a related genome for homologous genes of interest and many other interesting possibilities we are only beginning to consider. To get somewhat of a feeling for the potential of such chips, the technology now exists to choose a very large number of specific DNA sequences (for example, 10,000 genes of interest in humans) and simply have a computer pull the sequences out of a database and print a custom chip optimized for our particular problem using what amounts to fairly standard ink-jet printer technology. With such a chip in hand one can examine the tissue- and developmental-specific patterns of expression in genes of interest, as well as examine their expression in the tissues of related species. Chips can also be used to rapidly genotype individuals at literally tens of thousands of SNP sites.

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, a fraction of microsatellite markers found in *Drosophila* may extend over to (say) a particular moth of interest. Likewise, a DNA chip from a model organism can be used to extract homologous genes of interest, and subtraction methods allow the unique genomic sequences of one organism (relative to a target model system) to be extracted (or at least enriched). There is thus a level of acceleration with genomics, in that. as more organisms are investigated, the phylogenetic space of model systems becomes increasingly denser, further facilitating genomic projects by providing systems with even closer sequence homology. For example, using results from *Drosophila*, one lab may obtain genomic sequences from the silkworm (*Bombyx mori*) and another from the corn earworm (*Helicoverpa zea*). With both these lepidopteran sequences in hand, workers using other moth species can use these more closely related species in place of *Drosophila* as their model system benchmarks. The World Wide Web greatly facilitates rapid exchange of this growing web of genomic information.

In the immediate future, it is thus clear that large amounts of genomic sequences can be obtained for just about any organism of interest. The impact this will have on quantitative-genetic studies in these (and related) organisms are twofold. First, obvious candidates for microsatellite markers can be chosen from sequence data. While scanning for population variation in potential markers is very straightforward, constructing the genetic map using these markers requires the ability to recover and score meiotic products. The ability to use the polymerase chain reaction (PRC) to sequence single haploid gametes allows such maps to be constructed even for long-lived, or otherwise difficult to breed, species. Likewise, methods for rapid detection of SNPs continue to be developed at a rapid pace. For example, the SNP Consortium (http://snp.cshl.org) announced in August 2000 that it had detected almost 300,000 human SNPs, while (not to be outdone) Celera (http://www.Celera.com) announced a month later that it had detected 2.4 million human SNPs.

The second impact from genomic sequences is the potential to intelligently suggest candidate genes for the character(s) of interest. This is not nearly as straightforward as using sequence data to find potential markers, and a variety of strategies have been suggested to find suitable candidates. The most obvious is to choose candidates based on homologies to genes known to be

involved in trait variation in some well-studied model system. At the coarsest level, this could simply entail examining all genes that are known to have mutations that affect the trait of interest. For example, in *Drosophila* there are over 130 genes with known mutations affecting bristle number (T. Mackay, personal communication). Not only are these loci candidates for bristle number QTLs in *Drosophila*, but their homologues are bristle candidates for other insects as well. With such a collection of potential candidates in hand, one can do additional screening, for example by using microarrays to see which (if any) are indeed expressed in tissues that affect the character(s) of interest.

The next generation of strategies for candidate inclusion–rejection requires the ability to read a DNA sequence in much more biological detail than is currently possible. At present, we can translate any particular coding region of DNA into an amino sequence and by looking for specific protein motifs can make some general (and often somewhat intelligent) statements about its function (i.e., the resulting protein spans a membrane, or binds DNA, etc.). As more proteins are examined, the catalogue of structural motifs will grow, allowing us to make even more informed statements about protein function simply from its amino acid sequence. Results from such proteomic studies may suggest possible genes that otherwise would be overlooked as candidates. Equally important, it may allow other genes to be excluded as candidates. The third generation of candidate detection–exclusion strategies (which may be achieved in the next decade) will be deciphering the regulatory aspects of a DNA sequence, in particular being able to detect tissue- or developmental- specific regulatory signals. Such information will obviously be greatly informative for adding and removing genes as candidates. The most far-removed prospect is in understanding the gene circuits and pathways to the extent of predicting which genes are likely to have major effects on the flux through a pathway. If we can determine that a particular gene can have only a very minor effect on a pathway of interest, it may indeed be a QTL, but it is expected to account for very little trait variation.

The above collection of approaches, some fully developed, the others perhaps nothing more than pipe dreams, can suggest a set of candidates. Of course, we will have to tolerate inclusion of a large number of failed candidates to avoid exclusion of candidates that are indeed QTLs in the population of interest. One approach is to start with a collection of candidates, test for associations, and determine how much residual genetic variation remains after the successful candidates are included. If only a small fraction of genetic variation is accounted for, the search for additional candidates has the potential to be rewarding. The converse is not necessarily true. Even if the model based on the successful candidates accounts for most of the genetic variation, important alleles at other loci (having a large effect on the genotypic value, but a small effect on the variance because they are rare) can still be present.

The final point about candidates is that while much of the focus is on those genes contributing to within-population variation, genes contributing to between-population differences are equally (if not more) important to breeders and evolutionary geneticists. Sequencing projects offer the potential for rapidly scanning a number of candidates for such between-population differences. In populations where the character divergence is thought to have been at least partially driven by selection, one can search for signatures for selective sweeps around candidate genes (provided the divergence is sufficiently recent). Indeed, such an approach (using a cloned maize QTL) suggests that a single regulatory change in a QTL is responsible for some of the major differences in plant architecture between maize and its undomesticated ancestor teosinte (Wang *et al*., 1999).

While a number of clever strategies can be used to enrich the pool of candidate genes, it must be stressed that most candidates are likely to fail to show associations. Candidates influencing a character in one model system may have essentially no effect in a related species (some examples of this striking lack of correlation for mice vs humans are reviewed by Guo and Lange, 2000). Even when a significant correlation is found between candidate genotypes and trait values, it still must be demonstrated that this is a true association, as opposed to being an artifact of population stratification (subdivision). A classic example of such a stratification-induced correlation is that a marker highly associated with diabetes in Pima Native Americans is also a marker for admixture with the Caucasian population (which is a lower risk population for diabetes). In a restricted sample looking only at full-heritage individuals, this association vanishes (Knowler *et al*., 1988). Association tests have been developed to control for admixture (e.g., Ewens and Spielman, 1995; Allison, 1997; Spielman and Ewens, 1998; Allison, *et al*., 1999; Clayton, 1999; George *et al*., 1999; Monks and Kaplan, 2000). Unfortunately, these require collections of relatives. Ideally, by using sufficiently dense markers, SNP-trait associations can be detected in random samples from the entire population, using unlinked markers to control for population stratification. Such population-level associations are the current hope of many human geneticists trying to dissect complex traits. Further commentary on this approach

was recently offered in this journal by Guo and Lange (2000), and elsewhere by Kruglyak (1999) and Risch (2000).

When a large number of markers are examined, the problem of controlling for false positives while still maintaining statistical power is a difficult one. For example, using DNA chip technology, one could easily score 30,000 SNP sites in humans. Given the roughly 3000 cM map of humans, such markers (if suitably chosen) would span every 0.1 cM, meaning that any potential QTL is less than 0.05 cM from a marker (roughly 50 kb). In sib-pair or other pedigree approaches of QTL mapping, such marker information can be efficiently combined to estimate the probability of identity by descent for any particular chromosomal region. However, the problem of dealing with such a large number of highly correlated tests (especially for markers on the same chromosomes) in population-level association studies is still largely unresolved. The advantage of scanning a small collection of candidates (versus a whole-genome SNP scan) is that far fewer tests are involved. The difficult issue is moving from a candidate-trait association (for example, by a significant variance associated with a candidate locus) to identifying particular SNP sites within that candidate influencing the trait of interest. For any candidate, we expect a population sample to contain a number of SNP sites which are likely to be very highly correlated, greatly complicating association studies at the nucleotide level.

The false-positives versus power problem that complicates the search for candidates is even more problematic in tests of second and higher-order epistasis between candidates. Restricting epistasis tests to only candidates that each show significant marginal effects is one approach, but one can easily conceive of situations where loci have small marginal effects, but specific combinations could have substantial effects.

HUMAN GENETICS

Given the above general considerations about the future of quantitative genetics, what can we say about the future impacts for specific fields?

We will say very little about the potential directions of human quantitative-genetics over the next 20 years, as this subject has received much recent discussion (e.g., Guo and Lange, 2000; Risch, 2000). Briefly, the most obvious, and well-stated, goal of the human genome project as it relates to quantitative genetics is in the search for QTLs for disease susceptibility. As Risch (2000) has stressed, genes of major effect are the low-lying fruit, and hence easily picked, while genes of more modest effect will prove considerably more difficult. Even with the complete human sequence in hand, large and carefully designed association studies are required to demonstrate that particular candidate loci do indeed have an impact on human disease. There is considerable debate as to whether the best approach for demonstrating such associations is to use a large genetically heterogeneous population (such as a random sample from a major cosmopolitan city) or to sample from more genetically homogeneous populations (such as Iceland). One of the unstated assumptions of many QTL mappers is that loci that overall have a modest effect may have a much larger effect in the right environment and/or genetic background. If genetic background is critical (i.e., there are large epistatic interactions between some QTLs), then using a more homogenous population offers more power for detecting certain QTLs (those whose more favorable backgrounds are fixed or are at least at high frequencies) and will entirely miss others (if the background is fixed for the wrong loci). Even if epistasis is not significant, we fully expect different populations to be segregating different alleles with different effects. It is likely that association studies may first start with a collection of homogenous populations to provide higher power for mapping a subset of QTLs and then use the total collection of detected associations to screen large heterogeneous populations for additional associations.

BREEDING

For centuries, selection on complex traits to improve domesticated plants and animals has been entirely on phenotype. While this has proven to be a fabulously successful approach, the age of genomics offers the prospect of shifting selection directly to genotypes. At first glance this appears to offer the potential to greatly enhance selection response, but it must be remembered that a gene of major effect rapidly increases in frequency under simple mass (phenotypic) selection. Thus, in many cases, the ability to select directly on genotypes might not offer dramatic increases in response. One obvious case where selection on genotypes can potentially increase response is when we can select individuals before the phenotypic trait is expressed. With the appropriate markers, favorable genes can be selected even in individuals where they are not expressed (for example, selection on milk production in males), but BLUP (best linear unbiased predictors), prediction of breeding values using information from relatives, can accomplish the same task in many settings

(but not all, for example as pointed out by one of the reviewers, males from within the same full-sib family have the same BLUP as they have the same parents).

There are important reasons that phenotype should always be a component in selection schemes. Mutation is continuously generating a background of new variation, and selection on phenotypes is an easy way to capture much of this variation. A more subtle reason is that selection schemes targeting specific genotypes can result in a much greater reduction in effective population size (N_e) than selection based solely on phenotypic value. Given that a very significant component of selection response after the first few generations arises from new mutations (Hill 1982a, b; Frankham, 1980; Weber and Diggins, 1990), a reduction in N_e reduces the accumulated mutational variance because of random genetic drift, lowering the long term rate of response.

Equally important as the genomic revolution are advances in reproductive technologies. Already, the ability to select plant cells in tissue culture and then grow up the surviving individuals into fertile plants greatly increases the efficiency of selection for certain characters (such as pesticide resistance). Likewise, the relatively recent breakthroughs in cloning large mammals mean that animal breeders will now have to learn the schemes plant breeders have used for ages to exploit asexual reproduction. Further expected advances may allow a breeder to use cell cultures of individuals with elite breeding values, creating clones as needed. With sufficient breakthroughs in cell culture technology, it is conceivable that instead of cloning a series of individuals from (say) a prize bull, one could simply extract sperm from an appropriate cell culture. The downside to these approaches is that rapid evolution for adaptation to culture growth can result in rather dramatic genetic changes (such as chromosomal loss or duplication), which limits the shelf-life of such cultures as a powerful breeding tool. Further, as plant breeders are well aware, Mendelian reassortment and recombination are critical components for continued progress, so that cloning is not a substitute for sexual reproduction.

With embryo transplantation and cloning, we now have the experimental tools to ask very interesting questions about the importance of maternal effects (such as the importance of the maternal pre- and post-natal environment). A much deeper understanding of the importance of maternal effects is thus expected to be developed in the near future. This has important implications for breeders, as one can easily envision implanting embryos from in vitro crosses of elite parents into surrogate mothers selected for high performing maternal environments.

Finally, while the impact of transgenic organisms cannot be underestimated in terms of introducing important genetic variation, we presently have a very poor ability to predict the phenotypic consequences of any particular transgenic construction. Currently, novel genes are essentially inserted at random in the genome, usually under some strong promoter sequence to mitigate position effects. This may result in advantageous changes at one trait, but can also result in deleterious effects on others. While present transgenic technologies work well for major genes, given the variation in expression from position effects its effectiveness for moving genes of modest effects is unclear. A deeper understanding of the regulatory control signals will certainly increase transgenic efficiencies, as will the development of truly homologous recombination vectors. Given the essentially unlimited variation from the introduction of novel genes, should breeders still worry about maintaining sufficient effective population sizes? The answer is clearly yes. Besides reducing the impact of inbreeding depression, larger populations are expected to harbor greater polygenic variation (everything else being equal). A large population has a far better chance of segregating modifier genes to ameliorate deleterious effects associated with the introduction of some major genes.

EVOLUTIONARY GENETICS

The genetic architecture of traits is of key concern to evolutionary geneticists. Many of these architectural issues can be addressed by analysis of a collection of tightly linked markers and the appropriate experimental design. Examples include the relative roles of deleterious recessives vs overdominant loci in inbreeding depression, the number of loci that contribute a significant fraction of variation, whether new mutations occur mainly at the set of already segregating loci or if entirely new loci are involved, etc. Hence, many of these fundamental issues can be examined without ever having to sequence a single QTL.

This being said, one ultimate aim of evolutionary quantitative geneticists is the full sequence analysis of a collection of QTLs in several model systems. Such comparative sequence data will allow us to address the fundamental question of the molecular causes of phenotypic variation (both within a population and between populations or species). Are modest phenotypic differences largely caused by genes very far along a developmental pathway or can subtle differences in expression or function of genes at the base of a complex pathway generate

small differences in the phenotypic end product? Are the loci involved in phenotypic variation fairly trait-specific or are they involved in organism-wide signaling such as hormones, hormonal receptors, ion channel genes, or transcription factors? Is the oft-stated view (e.g., Wilson, 1976; Carroll, 2000) that the majority of phenotypic differences (and hence adaptation) are due to regulatory, as opposed to structural, changes correct? When similar phenotypes arise independently in different species, are the same set of genes involved? Is there a difference in strength of selection on regulatory vs. structural differences contributing to trait variation? Are there genomic hot spots of quantitative trait variation? Are there genomic constraints to evolution that one can quantify? The resolution of these questions for even a small number of model systems will provide considerable insight into evolution. The machinery to attack many of these questions is already in hand. For example, with a population sample of DNA sequences from a QTL, one can use standard tests to look for signatures of selection as well as compare differences between species.

Finally, the ability to construct transgenic organisms offers the possibility (in certain settings) to do the once unthinkable, namely to regenerate the sequence of key ancestral steps during the adaptation of a particular character. If one localizes (say) six major genes between an ancestral population and a derived one adapted to a particular environment, each major gene can be inserted into the otherwise ancestral background, and the fitness effects of these constructs can be examined in the appropriately controlled natural settings.

THE COMPUTATIONAL REVOLUTION: A BAYESIAN FUTURE?

In addition to the new technologies from genomics and molecular biology, the computer revolution (itself a key component for practical genomics) has very significant implications for the future of quantitative genetics. As computers continue to become faster and cheaper, computationally intensive algorithms will continue to gain importance. Resampling approaches (such as randomization tests and bootstrapping) already have had an impact on QTL mapping, allowing for the construction of confidence intervals and proper tests of significance. Likewise, variance component estimation using very large pedigrees, pioneered by animal breeders using the BLUP machinery developed by Henderson (reviewed in Lynch and Walsh, 1998), will continue to grow in importance in the analysis of both human and natural populations. Indeed, such an approach has been suggested as a powerful tool for QTL mapping in humans (Almasy and Blangero, 1998). With a collection of major genes in hand, extensive studies of genotype-environment interactions become practical, and the AMMI (additive main effects, multiplicative interactions) model of plant breeders (Gauch, 1992) offers a powerful, yet usually overlooked, tool for the analysis of genotype by environmental effects.

The most far-reaching of these computationally intensive methods are Markov Chain Monte Carlo (MCMC) approaches for simulating draws from complex probability distributions (e.g., Geyer, 1992; Tierney, 1994; Tanner, 1996). In particular, the Gibbs sampler (Geman and Geman, 1984) allows for the efficient exploration of very complex likelihood surfaces and calculation of Bayesian posterior distributions (Smith and Roberts, 1993). These resampling approaches (Gibbs and MCMC) allow for relatively easy analysis in models where standard likelihood calculations, at best, are extremely difficult (such as likelihoods over complex pedigrees, e.g., Thompson, 2000). A more profound result of these resampling methods is that, akin to the shift over the past 20 years in quantitative genetics towards more likelihood-based analysis, the next 20 years will likely be marked by a similar influx of Bayesian methods replacing their likelihood counterparts. For example, resampling-based Bayesian methods for multiple QTL mapping have recently been proposed (Sillanpaa and Arjas, 1998, 1999; Stephens and Fisch, 1998).

As most readers are undoubtedly aware, there is a close connection between likelihood and Bayesian approaches. Under a Bayesian analysis, the posterior probability density for a vector $\mathbf{u}$ of parameters is proportional to the prior density assumed for $\mathbf{u}$ times the likelihood function L for $\mathbf{u}$ given the observed data vector $\mathbf{x}$,

$$\text{Posterior}(\mathbf{u} \mid \mathbf{x}) = \text{constant} * L(\mathbf{u} \mid \mathbf{x}) * \text{Prior}(\mathbf{u})$$

A likelihood ratio analysis is concerned with the local geometry of the likelihood surface around its maximal value. A Bayesian analysis is concerned with the entire shape of the posterior, which depends on both the likelihood function and the assumed prior. A key feature of such an analysis is that, by suitable integration of the full posterior density, marginal posterior distributions for any component(s) of $\mathbf{u}$ can be obtained.

While there are deep philosophical differences between Bayesian and frequentist statisticians (which at times takes on the air of religious warfare), this predicted shift from likelihood to Bayesian approaches will be driven by

practical and applied concerns. Many quantitative geneticists may not be deeply worried about the subtle differences between Bayesian and frequentist notions of probability (much to the anguish of both sides), but will immediately appreciate that a Bayesian marginal posterior for an unknown parameter gives a distribution that incorporates any uncertainty in estimating other nuisance parameters as well. For example, an estimate of an epistatic variance involves estimates of additive, dominance, and environmental variances as well as potential fixed effects. The marginal posterior for the epistatic variance component, by integrating the posterior over all other parameters, automatically incorporates the uncertainty in these nuisance parameters into the uncertainty of the epistatic variance.

Besides debates over philosophical considerations (such as the choice of a prior), the major stumbling block to Bayesian approaches has been their mathematical intractability for the complexity inherent in most quantitative genetics models. This was also the case for likelihood methods 30 years ago, when faster computers and better algorithms made these approaches feasible. Likewise, ever faster computers and the Gibbs sampler (and other MCMC methods) have made Bayesian methods very widely applicable, and they are expected to play a major role in quantitative genetics. This is especially true in the analysis of the coming generation of high-dimensional models that must deal simultaneously with phenotypic, molecular, and environmental information. A concern expressed by one of the reviewers (in which I am in complete agreement) is that a negative consequence of resampling methods may well be an increased use of inappropriate models. The (relative) ease of these methods can tempt the unwary into using a highly complex model that does not fit the observed data. Resampling methods allow for extensive exploratory data analysis (for example, through jackknifing and bootstrapping), and it is hoped that quantitative biologists will be trained to fully examine data sets before proceeding with any particular underlying model of analysis.

SUMMARY

Quantitative genetics is indeed very healthy in this coming age of genomics, and will play an even greater role as genotypes of potential interest are investigated by human geneticists breeders, and evolutionary geneticists. While we have (or soon will have) the ability to do experiments that the founders of quantitative genetics could never envision in their wildest imagination, the basic machinery they developed is easily adaptable to the new analyses that will be required.

Far from "freeing" molecular biologists from mathematics, the age of genomics has forced an appreciation of the importance of quantitative methods. As we start to mine this genomic information and attempt to map molecular variation into trait variation, quantitative genetics will move even more to the forefront. I'm afraid my molecular colleagues will have to develop a deeper appreciation of Fisher (1918).

ACKNOWLEDGMENT

Thanks to Sam Karlin and the anonymous reviewers for their excellent comments and suggestions.

REFERENCES

Allison, D. B. 1997. Transmission-disequilibrium tests for quantitative traits, *Amer. J. Hum. Genetics* **60**, 676–690.

Allison, D. B., Heo, M., Kaplan, N., and Martin, E. R. 1999. Sibling-based tests of linkage and association for quantitative traits, *Amer. J. Hum. Genetics* **64**, 1754–1764.

Almasy, L., and Blangero, J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees, *Amer. J Hum. Genetics* **62**, 1998–1211.

Basten, C. 2000. Review of genetics and analysis of quantitative traits, *Theor. Pop. Biology* **57**, 307.

Carroll, S. B. 2000. Endless forms: the evolution of gene regulation and morphological diversity, *Cell* **101**, 577–580.

Clayton, D. 1999. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission, *Amer. J. Hum. Genetics* **65** 1170–1177.

Ewens, W. J., and Spielman, R. S. 1995. The transmission/disequilibrium test: history, subdivision, and admixture, *Amer. J. Hum. Genetics* **57**, 455–464.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance, *Trans. Royal Soc. Edinburgh* **52**, 399–433.

Frankham, R. 1980. The founder effect and response to artificial selection in *Drosophila*, *in* "Selection Experiments in Laboratory and Domestic Animals" (A. Roberston, Ed.), pp. 87–90. Commonwealth Agricultural Bureaux.

Gauch, H. G. 1992. "Statistical Analysis of Regional Yields Trails: AMMI Analysis of Factorial Designs," Elsevier, Amsterdam.

Geman, S., and Geman, D. 1984. Stochastic relaxation, Gibbs distribution and Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741.

George, V., Tiwari, H. K. Zhu, X., and Elston, R. C. 1999. A test for transmission/disequilibrium for quantitative traits in pedigree data, by multiple regression, *Amer. J. Hum. Genetics* **65**, 236–245.

Geyer, C. J. 1992. Practical Markov chain Monte Carlo (with discussion), *Stat. Sci.* **7**, 73–511.

Guo, S.-W., and Lange, K. 2000. Genetic mapping of complex traits: promises, problems and prospects, *Theor. Popul. Biol.* **57**, 1–11.

Hill, W. G. 1982a. Rates of change in quantitative traits from fixation of new mutations, *Proc. Natl. Acad. Sci.* **79**, 142–145.

Hill, W. G. 1982b. Predictions of response to artificial selection from new mutations, *Genet. Res.* **40**, 255–278.

Knowler, W. C., Williams R. C., Pettitt, D. J., and Steinberg, A. G., 1988. $Gm^{3;5,13,14}$ and type 2 diabetes mellitus: an association in American Indians with genetic admixture, *Am. J. Hum. Genet.* **43**, 520–526.

Kruglak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes, *Nature Genetics* **22**, 139–144.

Lynch, M. and Walsh, B. 1998. "Genetics and Analysis of Quantitative Traits," Sinuaer Associates, Sunderland, MA.

Monks, S. A., and Kaplan, N. L. 2000. Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus, *Amer. J. Hum. Genetics* **66**, 576–592.

Risch, N. J. 2000. Searching for genetic determinants in the new millennium, *Nature* **405**, 847–856.

Sillanpaa, M. J., and Arjas, E. 1998. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data, *Genetics* **148**, 1373–1388.

Sillanpaa, M. J., and Arjas, E. 1999. Bayesian mapping of multiple quantitative trait loci from incomplete oubred offspring data, *Genetics* **151**, 1605–1619.

Smith, A. F. M., and Roberts, G. O. 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte-Carlo methods (with discussion), *J. Royal. Statist. Soc. Ser. B* **55**, 3–23.

Spielman, R. S., and Ewens, W. J.. 1998. A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test, *Amer. J. Hum. Genetics* **62**, 450–458.

Stephens, D. A., and Fisch, R. D. 1998. Bayesian analysis of quantitative trait loci data using reversiible jump Markov chain Monta Carlo, *Biometrics* **54**, 1334–1347.

Tanner, M. A. 1996. "Tools for Statistical Analysis," Springer-Verlag, Berlin/New York.

Tierney, L. 1994. Markov chains for exploring posterior distributions (with discussion), *Ann. Statist.* **22**, 1701–1762.

Thompson, E. A. 2000. "Statistical Inferences from Genetic Data on Pedigrees," IMS Books.

Wang, R.-L., Stec, A., Hey, J., Lukens, L., and Doebley J. 1999. The limits of selection during maize domestication. *Nature* **398**, 236–239.

Weber, K. E. and Diggins, L. T. 1990. Increased selection response in larger populations. II. Selection for ethanol vapor resistance in Drosophila melanogaster at two population sizes, *Genetics* **125**, 585–597.

Wilson, A. C. 1976. Gene regulation in evolution, *in* "Molecular Evolution" (F. J. Ayala, Ed.), pp. 225–236, Sinauer Associates, Sunderland, MA.

Winkelman, D. C., and Hodgetts, R. B. 1992. RFLPs for somatotropic genes identify quantitative trait loci for growth in mice, *Genetics* **131**, 929–937.