

Modeling the Genetic Architecture of Complex Traits With Molecular Markers

Rongling Wu^{1,2,*}, Wei Hou³, Yuehua Cui⁴, Hongying Li¹, Tian Liu¹, Song Wu¹, Chang-Xing Ma⁵ and Yanru Zeng²

¹Department of Statistics, University of Florida, Gainesville, FL 32611 USA, ²School of Forestry and Biotechnology, Zhejiang Forestry University, Lin'an, Zhejiang 311300 People's Republic of China, ³Department of Epidemiology and Health Policy Research, University of Florida, Gainesville, FL 32611 USA, ⁴Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA and ⁵Department of Biostatistics, The University at Buffalo, Buffalo, NY 14214-3079, USA

Received: September 29, 2006; Accepted: November 16, 2006; Revised: November 29, 2006

Abstract: Understanding the genetic control of quantitatively inherited traits is fundamental to agricultural, evolutionary and biomedical genetic research. A detailed picture of the genetic architecture of quantitative traits can be elucidated with a well-saturated genetic map of molecular markers. The parameters that quantify the genetic architecture of a trait include the number of individual quantitative trait loci (QTL), their genomic positions, their genetic actions and interactions, and their responsiveness to biotic or abiotic factors. A variety of genetic designs and statistical models have been developed to estimate and test these architecture-modeling parameters. With the availability of very highly abundant single nucleotide polymorphism markers, DNA sequence variants, i.e., quantitative trait nucleotides (QTNs), which contribute to quantitative variation can be identified. A newly emerging active area - functional mapping, has shown its value to unravel the genetic machinery of dynamic traits at the QTL or QTN level that change their phenotypes with time or other variables. Functional mapping provides a quantitative framework for testing the interplay between genetic effects and trait formation and development and, thus, appeals to push statistical genetic analysis and modeling into the context of developmental biology. Some of the statistical methods for genetic mapping have been patented.

Keywords: Quantitative trait loci, linkage, linkage disequilibrium, interval mapping, functional mapping, complex trait, genetic architecture.

INTRODUCTION

Most traits of agricultural, biological and biomedical importance vary continuously because they are often controlled by an unknown number of genes, each with a small effect and segregating in a Mendelian fashion, and are also sensitive to environmental changes [1]. Because of this, the genetic study of these so-called quantitative traits or complex traits has always been one of the most difficult tasks in the entire biological science. A number of quantitative genetic models that combine Mendelian inheritance and traditional statistical approaches, such as analysis of (co)variance, have been developed to separate genetic and environmental effects on a quantitative trait [1]. The experimental results from these models have been instrumental for providing scientific guidance for plant and animal breeding as well as evolutionary predictions for various developmental events.

Under many circumstances, quantitative genetic models have been largely insufficient to unveil the precise genetic architecture of a complex trait because they cannot estimate the number of genes involved and the chromosomal locations and genetic effects of these genes. While quantitative genetics had almost entered a "dead end", it has been activated by two important scientific developments in

late 1970s and early 1980s. First, the rapid development of molecular technologies led to the generation of a virtually unlimited number of markers that specify the genome structure and organization of any organism [2]. Second, improved statistical and computational techniques, such as the Expectation-Maximization algorithm [3], made it possible to tackle complex genetic and genomic problems. Lander and Botstein [4] were the first who integrated molecular genetics and statistics to establish a general conceptual framework for dissecting a quantitative trait into its individual gene components, referred to as quantitative trait loci (QTL), through an association analysis between the markers and phenotypes. Since the publication of Lander and Botstein's seminal idea, a large body of sophisticated statistical methods has been developed for mapping complex traits [5-10]. These methods have been employed to identify and map specific QTL that contribute to phenotypic variation in a wide range of quantitative traits for plants, animals and humans [11-14].

Statistical approaches for QTL mapping rely on the nature of a mapping population, in which the markers and putative QTL are co-segregating. With the availability of DNA sequence technologies, it has now been possible to identify sequence variants that are associated with quantitative variation, which thus helps to gain better insights into the genetic basis of a complex trait. In this article, we will review fundamental approaches for detecting, estimating and locating individual QTL and point out their application to study the genetic architecture of a quantitative

*Address correspondence to this author at the Department of Statistics, University of Florida, Gainesville, FL 32611; USA, Tel: (352)392-3806; Fax: (352)392-8555; E-mail: rwu@stat.ufl.edu

trait. In particular, we will highlight a presently active area of QTL mapping for complex dynamic traits that undergo developmental changes during an organism's lifetime.

LINKAGE MAPPING

The purpose of QTL mapping is to locate specific QTL on the chromosome and quantify their genetic effects on a quantitative trait. However, these QTL underlying the trait cannot be directly observed, and they should be inferred from other known neutral genes (called genetic or molecular markers) based on the co-segregation between the markers and QTL. Genetic markers dispersed over an organism's genome are typed within a mapping population of individuals, for whom phenotype data are available. The prerequisite for QTL mapping is the construction of a genetic linkage map with molecular markers. The greater coverage the map has of the entire genome, the more likely it is that the map can detect a complete suite of QTL for a trait.

By testing the statistical significance between the known markers and observed phenotypes, one can estimate the effects of the underlying QTL. Single marker analyses based on *t* or *F* tests can be used to obtain such information about QTL-phenotype associations. But this analytical approach can only estimate confounded effects due to the markers. We do not know whether such marker effects result from either strong QTL effect, or tight QTL-marker linkage, or both. Interval mapping proposed by Lander and Botstein [4] can separate the influence of the linkage and QTL effects by assuming that a putative QTL is located within an interval bracketed by two flanking markers. The basic idea of interval mapping is the computation of log-likelihood ratio (LR) tests between the null (i.e., there is no QTL) and alternative (i.e., there is a QTL) on an evenly separated grid of possible QTL locations. The maximum LR value calculated within the marker interval is compared to the critical threshold to declare the existence of a QTL. Because it is difficult to analytically determine a critical threshold when the null hypothesis contains a non-identifiable parameter, i.e., QTL position, permutation tests have been widely used for the empirical determination of the chromosome-wide or genome-wide threshold.

To demonstrate the utilization of interval mapping, we performed QTL mapping for body weight in an F_2 population of 535 mice, initiated from two strains, Large and Small [15]. With this population, a linkage map based on 75 microsatellite markers was constructed. The F_2 hybrids were weighted at 10 weekly intervals starting at age 7 days. The raw weights were corrected for the effects of each covariate due to dam, litter size at birth, parity and sex. We used body weight at age 10 weeks and chromosome 1 as an example to draw the profile of the LR values for QTL testing by searching for a QTL at every 2 cM from the left marker for each marker interval across the entire linkage group (Fig. 1). In a comparison between the maximum of the LR values and the threshold, a putative QTL is identified between marker interval D1Mit20 and D1Mit7. The genetic effects of the QTL on body size at age 10 weeks can be estimated, and was further tested to be significant because the peak of the LR value between these two markers is beyond the critical value determined from permutation tests.

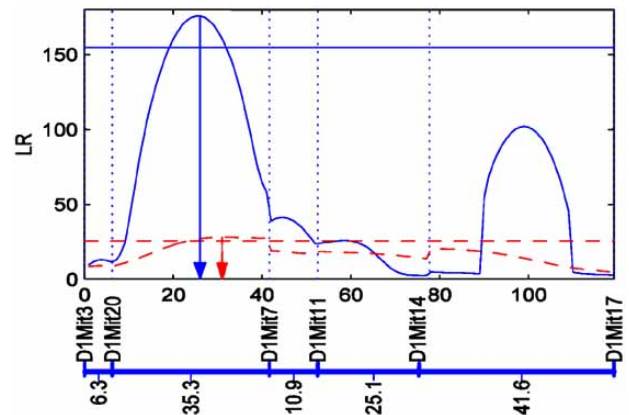


Fig. (1). The profile of the log-likelihood ratios (LR) between the full model (there is a QTL) and reduced (there is no QTL) model for body weight growth across mouse chromosome 1, estimated by interval mapping (broken curve), in a comparison with the LR profile estimated by functional mapping (solid curve). The genomic position corresponding to the peak of a profile is the maximum likelihood estimate of the QTL localization. The vertical broken lines indicate the positions of markers on this chromosome shown beneath. The map distances (in centimorgan) between two markers are calculated using the Haldane mapping function. The thresholds for proclaiming the genome-wide existence of a QTL are obtained from permutation tests [62]. The distribution of each of the LR values over 1000 simulation replicates can be approximated by a chi-square distribution. The 99.9th percentile of the distribution of the maximum is used as an empirical critical value to declare genome-wide existence of a QTL at the 0.001 significance level. These percentiles are indicated at two horizontal lines (the upper one for functional mapping and lower one for interval mapping). The data is from ref. [15].

Interval mapping approach cannot adequately use information from all possible markers on the genome and, therefore, can be biased when the same chromosome contains more than one QTL. Zeng [5] and Jansen and Stam [6] independently proposed a so-called composite interval mapping technique to increase the precision of QTL detection by controlling the chromosomal region outside the marker interval under consideration. This approach that has theoretically proven advantageous for the detection and separation of linked QTL on the same chromosome [16] has been widely adopted in practice. In a simulation study involving three QTL on a linkage group by Zeng [5], interval mapping provided a flat LR profile and did not give adequate power to separate these QTL (Fig. 2), but composite interval mapping is able to detect these QTL with reasonable resolution. Statistically, composite interval mapping is a combination of interval mapping based on two given flanking markers and a partial regression analysis on all markers except for the two ones bracketing the QTL. However, the choice of suitable markers loci (including linked or unlinked) that serve as covariates is still an open problem.

An interesting approach, called multiple interval mapping, is proposed by Kao *et al.* [7]. It is the extension of interval mapping by using multiple marker intervals

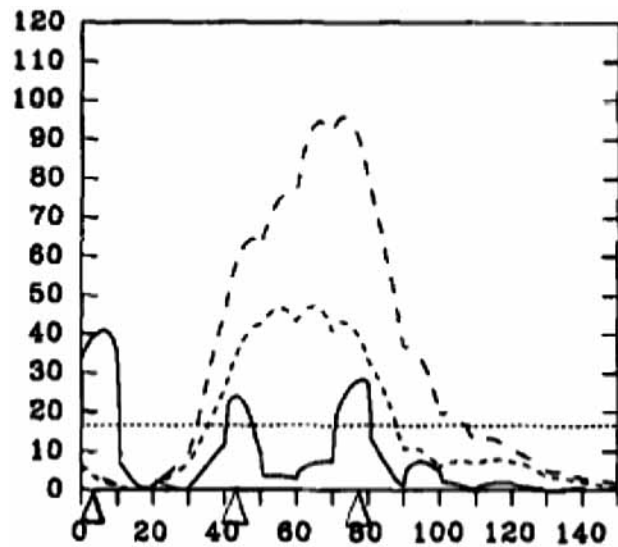


Fig. (2). Comparison in the detection of multiple linked QTL on a chromosome between interval mapping and composite interval mapping. The positions of QTL are indicated by the triangles. The three curves represent the log-likelihood ratio (LR) profiles under three different models, (1) Composite interval mapping involving all the linked markers on this chromosome and unlinked markers on other chromosomes (solid), (2) composite interval mapping involving all the unlinked markers (long-dashed) and (3) interval mapping (short-dashed). The horizontal line is the simulated 5% critical threshold for model (1). Adapted from ref. [5].

simultaneously to fit multiple putative QTL. This method allows to estimate and test both main and interaction effects among all possible QTL on the genome. Theoretically, it can precisely estimate the heritability of a trait because all the underlying QTL can be detected.

LINKAGE DISEQUILIBRIUM MAPPING

The most important goal in genetic research is to identify and characterize the actual genes that are responsible for phenotypic variation. Thus far, only a handful of genes that determine variation in commercially important traits have been described. The reason for the identification of relatively few genes can be attributed to limitations of the techniques used to detect genes. In the last decade, linkage analysis-based mapping approaches have been helpful for detecting QTL for a wide variety of traits, such as seed size and stem wood growth in plants and forest trees, fatness and obesity in animals, and hypertension and cancer in humans. But linkage analysis typically defines the location of a QTL to within a 20-30cM chromosomal interval - perhaps 1% of a species' genome. Given that around 30,000 functional genes in a mammalian genome are estimated in a typical genome, there are roughly 300 genes that are thought to under each QTL "bump" [17]. Thus, identifying the gene (or genes)

influencing the trait of interest based on a linkage analysis is a task hard to achieve.

More recently, an alternative approach based on linkage disequilibrium (LD), i.e., the non-random co-segregation of non-alleles at linked loci, has been shown to be powerful for aiding gene discovery [18]. The basic premise behind LD mapping is that a particular allele at a marker will tend to co-segregate with one allelic variant of the gene of interest, provided the marker and gene are very closely linked. LD mapping potentially has two advantages over conventional linkage mapping. The first is that it may be logistically easier. In theory, breeding schemes such as backcrosses or full-sib matings may not be required, making experimental design more straightforward and saving considerable time. The second, probably greater, advantage offered by LD mapping is that QTL may be mapped to very small regions thus aiding discovery of the underlying gene(s). In order to perform efficient LD mapping, markers must be mapped at a density compatible with the distances that LD extends in the population.

Currently, several consortia and laboratories have undertaken to develop dense maps of molecular markers for a wide variety of species. However, in order to predict how many markers will be required for LD mapping, the extent of non-random associations between markers must first be established. LD has been estimated in humans [19] as well as Holstein cattle [20], sheep [21] and dogs [22]. The disadvantage of linkage disequilibrium mapping is that the associations between different loci are also affected by evolutionary forces such as mutation, drift, selection and admixture and, thus, significant marker-QTL associations detected by a LD analysis may be spurious. This disadvantage can be overcome by a mapping strategy combining linkage and linkage disequilibrium, such as that developed in Wu and Zeng [23] and Wu *et al.* [8].

SINGLE NUCLEOTIDE POLYMORPHISMS AND HAPMAP

One of the fruits of the Human Genome Project is the discovery of millions of DNA sequence variants in the human genome. The majority of these variants are single nucleotide polymorphisms (SNPs), which comprise approximately 80% of all known polymorphisms, and their density in the human genome is estimated to be on average 1 per 1000 base pairs. SNPs, as the newest markers, have been the focus of much attention in human genetics because they are extremely abundant and well-suited for automated large-scale genotyping. A dense set of SNP markers opens up the possibility of studying the genetic basis of complex diseases by population approaches. SNPs are more frequent and mutationally stable, making them suitable for association studies to map disease-causing mutations, especially useful in personalized medicine for their association with disease susceptibility, drug treatment response and nutritional needs.

Several recent empirical studies suggest that SNPs are not evenly distributed over the genome in terms of the extent of LD and that the structure of haplotype (a linear arrangement of nonalleles at linked loci; Fig. 3) on a chromosome can be broken into a series of discrete haplotype blocks [24-28]. In each haplotype block, consecutive sites are in

successfully cloned [14,31], despite a considerable number of QTL reported in the literature.

A more accurate and useful approach for the characterization of genetic variants contributing to quantitative variation is to directly analyze DNA sequences associated with a particular trait [32]. If a string of DNA sequence is known to increase disease risk, this risk can be reduced by inhibiting the expression of this string of DNA sequence with a specialized drug. The control of this disease can be made more efficient if all possible DNA sequences determining its variation are identified throughout the entire genome.

With the recent development of the human genome project, massive amounts of DNA sequence data have been available across the human genome [30], facilitating the complete identification of specific sequence variants responsible for complex traits. Alternatively, candidate gene approaches have also been used for LD-based mapping. It is expected that LD mapping of candidate genes has a greater chance to unravel the genetic architecture of a complex trait because candidate genes are hypothesized to exert an effect on the formation process of the trait [13].

FUNCTIONAL MAPPING: FOUNDATION

Many traits, such as growth, disease (e.g., AIDS) progression and drug response, are dynamic in nature and should be measured in a longitudinal way. To obtain a clear picture of the genetic control of these traits, it is crucial to characterize the change pattern of genetic expression during development. Figure 4 illustrates four representative patterns of time-dependent genetic effects triggered by different QTL. For example, some QTL are permanently expressed, some are only turned on early in development while others are turned on at specific stages in development. For each pattern shown in Figure 4, curve parameters for developmental trajectories can be tested for individual genotypes. If different genotypes at a given QTL correspond to different trajectories, this QTL must affect differentiation of this trait. Thus, by estimating the curve parameters that define the trait trajectory of each QTL genotype and testing the differences in these parameters among genotypes, we can determine whether a dynamic QTL exists and how it affects the formation and expression of a trait during development.

Although, the elucidation of the relationship between genetic control and development for longitudinal traits is statistically a pressing challenge, some of key difficulties have been overcome in the statistical genetics group at the University of Florida [10,33,34]. A general statistical framework, called functional mapping, has been proposed to genome-wide map specific QTL that determine the developmental pattern of a complex trait.

The basic rationale of functional mapping lies in the connection between gene action or environmental effects and development by parametric or nonparametric models. Functional mapping maps dynamic QTL that are responsible for a biological process that is measured at a finite number of time points. A number of mathematical models have been established to describe the developmental process of a biological phenotype. For example, a series of growth

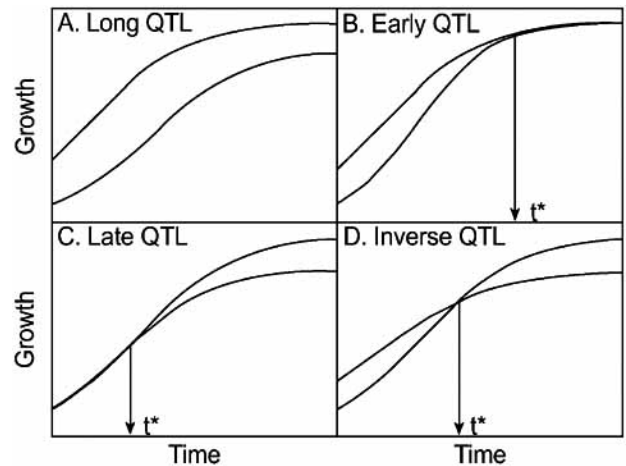


Fig. (4). Four representative patterns for the genetic control of growth trajectories by a dynamic QTL. Each curve is presented by a different QTL genotype. Pattern 1: Permanent QTL. Some QTL are permanently expressed, which gives rise to two parallel growth curves in which one genotype is consistently better than the other throughout the entire growth process (Fig. 1A). The expression of this QTL is not affected by development, and therefore shows no interaction with age. Pattern 2: Early QTL. Some QTL are expressed at early developmental stages and are switched off after a particular age. The two genotypes have different growth at early stages, but tend to converge at later stages (Fig. 1B). This QTL displays interactions with age because there is a change of variance of the QTL effects during development. Pattern 3: Late QTL. Some QTL remain silent during early stages and are expressed only after a particular age. The two genotypes display similar growth at early stages, but tend to diverge at later stages (Fig. 1C). Analogous to Pattern 2, there is a QTL \times age interaction in this case, operating with the conditional neutrality mechanism. Pattern 4: Inverse QTL. One genotype performs better than the other at the early stage of growth, but this changes at the later stage (Fig. 1D). This gene displays inverse effects at a particular age. Genotype \times age interactions occur because there is a change of the direction of the QTL effects during development.

equations have been derived to describe the S-shaped feature of growth in height, size or weight [35,36] that occur whenever the anabolic or metabolic rate exceeds the rate of catabolism (see Fig. 5 for a live example of growth curve). Based on fundamental principles behind biological or biochemical networks and allometrical scaling, West *et al.* [37] have mathematically proven the universality of these growth equations. Using a limited number of mathematical parameters, the shape, form and process of growth can be precisely described. With mathematical functions incorporated into the QTL mapping framework, functional mapping estimates parameters that determine shapes and functions of a particular biological network, instead of directly estimating the gene effects at all possible time points. Because of such connections among these points through mathematical functions, functional mapping strikingly reduces the number of parameters to be estimated and, hence, displays increased statistical power. In addition to the results by interval mapping, Figure 1 also gives an example for functional

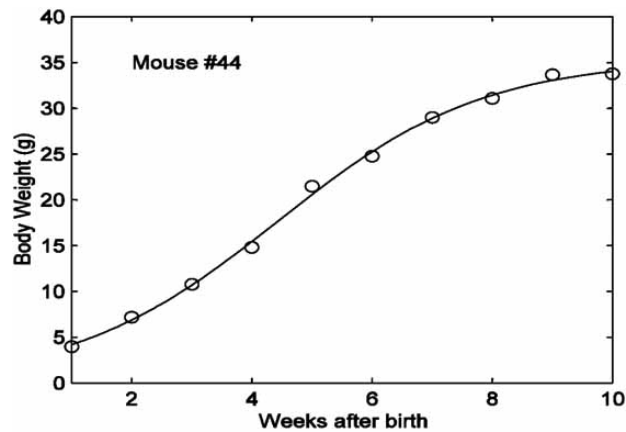


Fig. (5). One typical example of fit to growth curve using a logistic equation from an F_2 population of two mouse inbred strains. The data is from ref. [15].

mapping by simultaneously mapping growth trajectories from ages 1 to 10 weeks on mouse chromosome 1 [15] with the growth equation. The comparison between functional mapping and interval mapping suggests that the former has greater power to detect a growth QTL than the latter in terms of the width of the peak of the LR profile between markers D1Mit20 and D1Mit7. Functional mapping gives a narrower peak and, therefore, can locate the QTL at a smaller region, than interval mapping.

From a statistical perspective, functional mapping is a problem of jointly modeling mean-covariance structures in longitudinal studies, an area that has recently received a considerable interest in the statistical literature [38-41]. However, different from general longitudinal modeling, functional mapping integrates the parameter estimation and test process within a biologically meaningful mixture-based likelihood framework. Functional mapping is thus advantageous in terms of biological relevance because biological principles are embedded into the estimation process of QTL parameters. The results derived from functional mapping will be closer to biological reality.

Using the example shown in Fig. 1, we draw three different growth curves (Fig. 6), each corresponding to a genotype at the QTL detected on mouse chromosome 1, by estimating genotype-specific growth parameters from functional mapping. Statistical tests suggest that the three genotypic curves are significantly different before ages 4 weeks, their differences are reduced between ages 4 and 6 weeks, and then increased differentiation occurs after age 6 weeks. This indicates that the QTL detected triggers varying impacts on growth depending on stages of mouse development. The convergence of two homozygotes implies that the additive effect plays a gradually increased role in shaping growth trajectories, whereas the divergence of the heterozygote from the two homozygotes implies that the dominant effect on body weight growth is increased with age.

The effect of the detected QTL on growth rate can be further examined. The inflection point, at which growth rate is maximal, occurs about one week earlier for the homozygote for the allele from the Large strain than the

homozygote for the allele from the Small strain and about one day earlier for the homozygote for allele from the Small strain than for the heterozygote (Fig. 6). It is possible that this difference in development causes the homozygote for the Large strain to reach the asymptotic growth earlier than the other genotypes. By testing the shape and pattern of the three growth curves, we can investigate possible pleiotropic effects of this growth QTL on many different developmental events, such as the timing of sexually maturity and reproductive fitness, or biomedically important traits, such as metabolic rate and fatness. Functional mapping can, therefore, integrate growth and development, which are historically regarded as two different biological problems, into a comprehensive framework under which their common or unique underlying genetic machineries are identified.

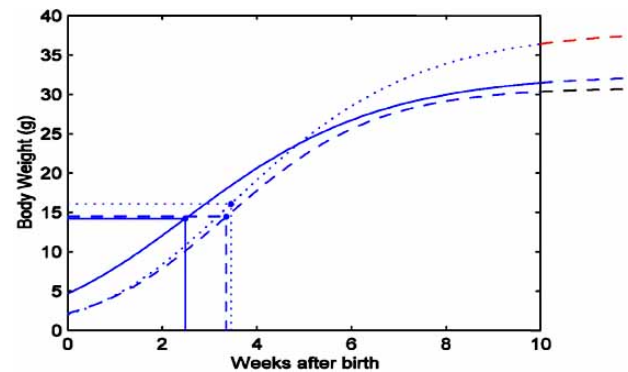


Fig. (6). Three growth curves, each presenting a group of genotypes at the QTL detected on mouse chromosome 1 in the F_2 population, using functional mapping (Fig. 1). The solid, slash and dot curves represent the homozygote for the allele from the Large strain, the homozygote for the allele from the Small strain and the heterozygote, respectively. The timing and growth at the inflection points for the three genotypic curves are indicated by the vertical and horizontal lines, respectively.

FUNCTIONAL MAPPING: DEVELOPMENTS AND CHALLENGES

Although, the idea of functional mapping was originally stimulated from a growth model for a controlled cross [33], this method has been extensively expanded to a wide spectrum of biomedical and evolutionary fields. Wang and Wu [42] and Zhu *et al.* [43] combined the principle of functional mapping and mathematical models for HIV dynamics established by Ho and colleagues [44,45] to map host QTL that govern dynamic changes of HIV viral loads in a human body. Different from the original linkage analysis strategy used to derive functional mapping in a controlled cross, functional mapping in natural human populations is founded on linkage disequilibrium that defines the degree of marker-QTL cosegregation or non-random association at the population level. Wang *et al.* [46] proposed a genetic model that incorporates the interaction between different QTL derived from viral and host genomes. Very recently, Wu *et*

al. [47] have generalized functional mapping for HIV dynamics by considering a genome-wide scan for host QTL based on multilocus linkage disequilibrium mapping.

Functional mapping has been extended to study the genetic control of drug response [48-50] by implementing well-developed pharmacodynamic and pharmacokinetic models [51]. Functional mapping, in conjunction with the DNA sequence data from the Human Genome Project [30], displays great potential to open a novel area for pharmacogenetic or pharmacogenomic research aimed to study the genetic control of drug response, an active cutting edge discipline that has recently received increased attention in general biomedical sciences [52]. Functional mapping allows to identify specific genetic variants that contribute to inter-patient variation in response to different drugs or different dosages of the same drug. The results from functional mapping will provide scientific guidance for the design of personalized medications based on a patient's genetic makeup.

"Naturally-occurring" or "programmed" cell death (PCD) in which the cell uses specialized cellular machinery to kill itself is a ubiquitous phenomenon that occurs early in organ development. Such a cell suicide mechanism that enables metazoans to control cell number and eliminate cells threatening the organism's survival has been thought to be under genetic control. The 2002 Nobel Prize in medicine was awarded to three scientists because of their discoveries of the genetic regulation of PCD [53]. A recently extended functional mapping model has made it possible to identify and detect QTL that are responsible for PCD in an organism [54]. This extended model incorporates the biological mechanisms of PCD that undergoes two different developmental stages, exponential growth and polynomial death. PCD-incorporated functional mapping will open a novel avenue for the study of the genetic architecture of important diseases, such as cancer, that undergo universal PCD processes.

In theory, genetic information contained within any kind of longitudinally measured data can be extracted by functional mapping. The advantages of functional mapping in model flexibility, stability and power result from its statistical formulation including the mathematical approximation of the mean vector and the modelling of the covariance structure by stationary [33] or nonstationary models [55]. Although, original functional mapping was derived for the biological processes that can be described by parametric functions, functional mapping has been modified within the nonparametric context to accommodate the situation in which biologically meaningful mathematical equations do not exist [56]. Analyses of longitudinal data need a number of mathematical manipulations for the covariance matrix, such as the calculation of the matrix determinant and inverse. In modeling the structure of covariance matrix by parametric functions, however, these mathematical manipulations will not be made possible because of the matrix's sparse structure, especially when the dimension of the matrix is too large. To overcome this problem, Zhao and Wu [57] proposed a de-noising approach based on wavelet transformation to reduce the dimensionality of longitudinal data. Preserving the favorable proper-

ties of functional mapping, wavelet thresholding approach creates a new direction in statistical genetics to manipulate high-dimensional multivariate dynamic data in both statistically and biologically meaningful ways [57].

In practice, functional mapping may frequently encounter a type of longitudinal data in agricultural and biomedical research, where repeated measurements are irregularly spaced or sparse and the numbers of repeated measurements vary among subjects. This type of sparse longitudinal data may also consist of noisy measurements with underlying smooth random trajectories for each subject in a sample. Some statistical theory and methods have been established to analyze sparse and irregular longitudinal data [58], although there is still much room for their improvement. To push functional mapping toward a broad range of applications, especially in HIV/AIDS studies, pharmacogenetic and pharmacogenomic research and genetic modelling of programmed cell death processes, there is a pressing need to consider the sparsity and irregularity of longitudinal data within the context of functional mapping.

CURRENT & FUTURE DEVELOPMENTS

With considerable achievements in genetic mapping due to collective efforts of researchers in the past two decades [4-10], we are now in a great position to detail the genetic architecture of a complex, quantitatively inherited trait. In her seminal review, Mackay [13] defined the overall picture of the genetic architecture of a complex trait in terms of its composed elements, i.e., the number of genes involved and their frequencies and pleiotropic effects, gene-gene (epistatic), gene-sex and gene-environment interactions. In a follow-up review, Mackay and colleague [59] further documented the importance of these elements in creating and maintaining the genetic variation of a specific quantitative trait in a population. Wu and colleagues [33,34] pioneered a general statistical framework for functional mapping that can be used to unravel the genetic architecture of dynamic complex traits, as defined by Mackay [13]. With continuing advances in molecular technology and the availability of a complete reference sequence of the entire genome for various species, it is urged to develop more powerful, more sophisticated, more meaningful, and more easily used model frameworks (like patents [60,61]) that help geneticists, breeders and clinical doctors extract useful genetic information from the data sets that are often high-dimensional and have complex dynamic structure.

ACKNOWLEDGEMENTS

We thank Dr. James Cheverud for providing his mouse data used to test interval mapping and functional mapping, the three anonymous referees for their constructive comments on an earlier version of this manuscript and two of the referees for clarifying the number of functional genes that have been estimated in mammals. The preparation of this manuscript is supported by NSF grant (0540745) to R. W.

REFERENCES

- [1] Lynch M, Walsh B. Genetics and Analysis of Quantitative Traits. Sinauer, Sunderland, MA 1998.
- [2] Drayna D, Davies K, Hartley D, Mandel JL, Camerino G, Williamson R, White R. Genetic mapping of the human X-

- chromosome by using restriction fragment length polymorphisms. *Proc Natl Acad Sci USA* 1984; 81: 2836-2839.
- [3] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc Ser B* 1977; 39: 1-38.
- [4] Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using FLP linkage maps. *Genetics* 1989; 121: 185-199.
- [5] Zeng Z-B. Precision mapping of quantitative trait loci. *Genetics* 1994 136; 1457-1468.
- [6] Jansen RC, Stam P. High resolution mapping of quantitative traits into multiple loci via interval mapping. *Genetics* 1994; 136: 1447-1455.
- [7] Kao C-H, Zeng Z-B. General formulas for obtaining the maximum likelihood estimates and the asymptotic variance-covariance matrix in QTL mapping when using the EM algorithm. *Biometrics* 1997; 53: 653-665.
- [8] Wu RL, Ma CX, Casella G. Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 2004; 160: 779-792.
- [9] Xu S-Z. Estimating polygenic effects using markers of the entire genome. *Genetics* 2003; 163: 789-801.
- [10] Wu RL, Ma C-X, Lin M, Casella G. A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics* 2004; 166: 1541-1551.
- [11] Doebley J, Stec A, Gustus C. Teosinte branched1 and the origin of maize: Evidence for epistasis and the evolution of dominance. *Genetics* 1995; 141: 333-346.
- [12] Peltonen L, McKusick VA. Dissecting human disease in the postgenomic era. *Science* 2001; 291: 1224-1229.
- [13] Mackay TFC. Quantitative trait loci in *Drosophila*. *Nat Rev Genet* 2001; 2: 11-20.
- [14] Li CB, Zhou AL, Sang T. Rice domestication by reducing shattering. *Science*; 2006; 311: 1936-1939.
- [15] Cheverud JM, Routman EJ, Duarte FAM, van Swinderen B, Cothran K, Perel C. Quantitative trait loci for murine growth. *Genetics* 1996, 142: 1305-1319.
- [16] Zeng Z-B. Theoretical basis of precision mapping of quantitative trait loci. *Proc Nat Acad Sci USA* 1993; 90: 10972-10976.
- [17] Skarnes WC, von Melchner H, Wurst W, Hicks G, Nord AS, Cox T, Young SG, Ruiz P, Soriano P, Tessier-Lavigne M, Conklin BR, Stanford WL, Rossant J. A public gene trap resource for mouse functional genomics. *Nat Genet* 2004; 36: 543-544.
- [18] Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002; 3: 299-309.
- [19] Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999; 22: 139-144.
- [20] Farnir F, Coppieiers W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res* 2002; 10: 220-227.
- [21] McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J. Linkage disequilibrium in domestic sheep. *Genetics* 2002; 160: 1113-1122.
- [22] Liu T, Todhunter RJ, Lu Q, *et al.* Modelling extent and distribution of zygotic disequilibrium: Implications for a multigenerational canine pedigree. *Genetics* 2006; 174: 439-453.
- [23] Wu RL, Zeng Z-B. Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* 2001; 157: 899-909.
- [24] Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; 29: 229-232.
- [25] Patil N, Berno AJ, Hinds DA, *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; 294: 1719-1723.
- [26] Dawson E, Abecasis GR, Bumpstead S, *et al.* A first-generation linkage disequilibrium map of human chromosome 22. *Nature* 2002; 418: 544-548.
- [27] Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in the human genome. *Science* 2002; 296: 2225-2229.
- [28] Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet* 2003; 33: 382-387.
- [29] Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 2003; 4: 587-597.
- [30] The International HapMap Consortium. The International HapMap Project. *Nature* 2003; 426: 789-94.
- [31] Frary A, Nesbitt TC, Grandillo S, *et al.* fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 2000; 289: 85-88.
- [32] Liu T, Johnson JA, Casella G, Wu RL. Sequencing complex diseases with HapMap. *Genetics* 2004; 168: 503-511.
- [33] Ma C-X, Casella G, Wu RL. Functional mapping of quantitative trait loci underlying the character process: A theoretical framework. *Genetics* 2002; 161: 1751-176.
- [34] Wu RL, Lin M. Functional mapping - How to study the genetic architecture of dynamic complex traits. *Nat Rev Genet* 2006; 7: 229-237.
- [35] von Bertalanffy L. Quantitative laws in metabolism and growth. *Quart Rev Biol* 1957; 32: 217-231.
- [36] Richards FJ. A flexible growth function for empirical use. *J Exp Bot* 1959; 10: 290-300.
- [37] West GB, Brown JH, Enquist BJ. A general model for ontogenetic growth. *Nature* 2001; 413: 628-631.
- [38] Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* 1999; 86: 677-690.
- [39] Pourahmadi M. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* 2000; 87: 425-435.
- [40] Pan JX, Mackenzie G. On modelling mean-covariance structures in longitudinal studies. *Biometrika* 2003; 90: 239-244.
- [41] Wu WB, Pourahmadi M. Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* 2003; 90: 831-844.
- [42] Wang ZH, Wu RL. A statistical model for high-resolution mapping of quantitative trait loci determining human HIV-1 dynamics. *Stat Med* 2004; 23: 3033-3051.
- [43] Zhu Y, Hou W, Wu RL. A haplotype block model for fine mapping of quantitative trait loci regulating HIV-1 pathogenesis. *J Theor Med* 2004; 5: 1-7.
- [44] Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 1999; 373: 123-126.
- [45] Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics *in vivo*: virion clearance rate, infected cell lifespan, and viral generation time. *Science* 1996; 271: 1582-1586.
- [46] Wang ZH, Hou W, Wu RL. A statistical model to analyze quantitative trait locus interactions for HIV dynamics from the virus and human genomes. *Stat Med* 2005; 25: 495-511.
- [47] Wu S, Yang J, Wu RL. Multilocus linkage disequilibrium mapping of quantitative trait loci that affect HIV dynamics: A simulation approach. *Stat Med* 2006; 25: 3826-3849.
- [48] Lin M, Aquilonte C, Johnson JA, Wu RL. Sequencing drug response with HapMap. *Pharmacogenomics J* 2005; 5: 149-156.
- [49] Lin M, Wu RL. Theoretical basis for the identification of allelic variants that encode drug efficacy and toxicity. *Genetics* 2005; 170: 919-928.
- [50] Lin M, Wu RL, Johnson JA. A bivariate functional mapping model for gene identification of drug response for systolic and diastolic blood pressures. *Pacific Symposium on Biocomputing (2005)* (<http://helix-web.stanford.edu/psb06/lin.pdf>).
- [51] Derendorf H, Meibohm B. Modeling of pharmacokinetic/pharmacodynamic (PK/PD) relationships: Concepts and perspectives. *Pharmaceutical Res* 1999; 16: 176-185.
- [52] Watters JW, McLeod HL. Using genome-wide mapping in the mouse to identify genes that influence drug response. *Trends Pharmacol Sci* 2003; 24: 55-58.
- [53] Horvitz HR. Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res* 1999; 59: 1701s-1706s.
- [54] Cui YH, Zhu J, Wu RL. Functional mapping for genetic control of programmed cell death. *Physiol Genomics* 2006; 25: 458-469.
- [55] Zhao W, Chen YQ, Casella G, Cheverud JM, Wu RL. A nonstationary model for functional mapping of complex traits. *Bioinformatics* 2005; 21: 2469-2477.
- [56] Lin M, Wu RL. A joint model for nonparametric functional mapping of longitudinal trajectories and time-to-events. *BMC Bioinformatics* 2006; 7:138-149.

- [57] Zhao W, Wu RL. A wavelet thresholding approach for functional genetic mapping of longitudinal traits. *Genetics* 2006 (revised).
- [58] Yao F, Müller H.-G, Wang J.-L. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc* 2005; 100: 577-590.
- [59] Anholt RR, Mackay TFC. Quantitative genetic analyses of complex behaviours in *Drosophila*. *Nat Rev Genet* 2004; 5: 838-849.
- [60] Schork, N.J., Essioux, L., Cohen-Akenine, A., Blumenfeld, M., Cohen, D.: US20016291182 (2001).
- [61] Frudakis, T.N.: US20067107155 (2006).
- [62] Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics* 1994; 138(3): 963-71.