

# PATTERNS OF GENETIC VARIATION IN MENDELIAN AND COMPLEX TRAITS

---

Michael E. Zwick, David J. Cutler and Aravinda  
Chakravarti

*Department of Genetics, and Center for Human Genetics, Case Western Reserve  
University School of Medicine, and University Hospitals of Cleveland, Cleveland, Ohio  
44106; e-mail: axc39@po.cwru.edu*

**Key Words** single nucleotide polymorphisms (SNPs), human complex traits,  
*Drosophila* sensory bristles

■ **Abstract** This review discusses the prospects for understanding the genetic basis of complex traits in humans. We take the view that work done on *Drosophila melanogaster* can serve as a model for understanding complex traits in humans, and the literature on this model system, as well as on humans, is reviewed. The prospects for success in understanding the genetic basis of complex traits depend, in part, on the nature of the forces acting on genetic variation. We suggest that different experimental approaches should be undertaken for traits caused by common genetic variants versus those arising from rare genetic variants.

## INTRODUCTION

As we begin the twenty-first century, human geneticists will soon be presented with a reference human genome sequence (17, 81), an historic event that will undoubtedly transform genetics and, in fact, all of the biological sciences. In a similar fashion, a revolution at the beginning of the twentieth century led to the origin of the field of genetics. From the rediscovery of Gregor Mendel's work in 1900, through the elucidation and extension of Mendel's laws in *Drosophila melanogaster* by the Morgan school and ultimately through the discovery of the structure of DNA by Watson and Crick, Mendelian genetics has profoundly influenced all areas of biological research. The most successful applications of Mendelian genetics have characterized the inheritance of traits in which single genotypic changes result in large, discrete phenotypic differences, that is, "Mendelian traits." But the segregation of phenotypic traits in a Mendelian fashion is the exception, rather than the rule. Most phenotypic differences between individuals in morphology, physiology, disease, growth, and behavior are quantitative in nature, exhibiting a continuous nearly normal phenotypic distribution (19). These complex traits arise from the interaction of multiple segregating genetic variants together with environmental

contributions. The central question for all geneticists is, How can we apply the tools and concepts of the Mendelian research paradigm to understand the genetic basis of complex traits?

Most attempts to characterize complex traits have generally relied on a largely descriptive statistical methodology. Although this approach has proven adequate for describing and predicting the response to artificial selection, remarkably little is known about the genetic architecture underlying complex traits (3), especially complex traits in humans. Statistical models, however, are not the sole methodology one can apply to complex traits. As early as 1920, Altenburg & Muller (2) demonstrated how to use Mendelian tools and concepts to genetically dissect a complex trait. Despite this early success in mapping the genetic variants underlying a complex trait in *D. melanogaster*, most geneticists focused on discrete Mendelian traits that were far easier to analyze.

In a seminal review nearly 50 years later and over 30 years ago, Robertson identified a suite of properties of the genes underlying complex traits that are of fundamental interest. These properties include their numbers and their additive, dominance, and pleiotropic effects, as well as their frequency, linkage relationships, and fitness effects (74). To determine these properties of genetic variants, however, it is first necessary to identify the genetic variants segregating in a population. Generating variation data across an entire genome has proven to be an extraordinarily difficult task. With the near completion of a human genome reference sequence and the development of technologies that promise the rapid identification of single-nucleotide polymorphisms (SNPs), generation of variation data might no longer limit the analysis of complex traits (10).

Consequently, how might human geneticists best apply Mendelian tools and concepts to characterize the genetic basis of human complex traits and diseases? Can the characterization of complex traits in any model organism guide the analysis of human studies? Closely examining studies of *D. melanogaster* can aid in this pursuit for three main reasons. First, the long history of studies of *D. melanogaster* genetics, development, evolution, and patterns of naturally occurring DNA sequence variation have generated a body of knowledge that is unmatched in any other model system. Second, natural *D. melanogaster* populations appear to have an evolutionary history that is grossly similar to that of humans. Patterns of genetic variation show that African populations are the most polymorphic (5) and suggest that all other worldwide populations may have recently expanded from an ancestral African population. Third, a number of studies have successfully characterized genetic variation underlying a model complex trait, *Drosophila* sensory bristles, by methodologies directly applicable to the study of human complex traits. With this in mind, the critical issues discussed in this review include the following:

1. How are observed patterns of genetic variation in Mendelian traits related to expected patterns of genetic variation in complex traits?
2. What has been empirically observed about patterns of genetic variation in the human and *Drosophila* genomes? Given the nature of previous empirical studies, what do we expect to have missed?

3. How can *Drosophila* sensory bristles act as a model complex trait to guide the investigation of human complex traits and diseases?
4. What approaches are likely to identify the genetic basis of complex traits and diseases in humans?

Because complex traits are expected to arise from the interaction of multiple segregating genetic variants, their study will inevitably be dependent on both the patterns of extant genetic variation and the evolutionary forces that act to maintain this variation in natural populations. Thus, the different evolutionary forces acting on naturally occurring genetic variation are likely to profoundly influence the methodologies and models that lead to an understanding of the genetic basis of complex traits.

## GENETIC ARCHITECTURE OF MENDELIAN AND COMPLEX TRAITS

### Patterns of Genetic Variation Underlying Mendelian Traits

Mendelian and complex traits can be considered as opposite extremes of a continuous spectrum of relationships between genotypes and phenotypes. Mendelian traits exist when a single mutation at a locus results in large, often discrete phenotypic effects. This simple pattern of cause and effect might suggest that, whenever one observes a Mendelian trait in multiple unrelated individuals, this trait shares a common genetic basis among those individuals. Unfortunately, this is not generally true. Furthermore, both *Drosophila* and humans depart from this expectation in a similar fashion. Among the earliest well-characterized Mendelian traits was the white-eye phenotype in *D. melanogaster*. The white-eye phenotype segregates as a single *X*-linked gene, and it certainly appeared possible that a single genetic change accounted for all white-eyed flies. Early genetic dissection of this phenotype showed that this was not the case. Between March 1915 and April 1942, there were 27 independent mutations recorded at the *white* locus in *D. melanogaster* (26). Currently, 149 independent spontaneous mutations at the *white* locus are identified in FlyBase (20), a database of the *Drosophila* genome projects and community literature. Similarly, in humans, although the cystic fibrosis (CF) locus harbors a single allele that accounts for 70% of all disease alleles, >550 additional disease alleles with various phenotypic effects have been identified (85). This great molecular diversity, in both *Drosophila* and humans, is characteristic of most loci that underlie Mendelian traits. Furthermore, many of these mutations are individually rare and likely deleterious owing to their loss of gene function. This pattern of genetic variation at the genes underlying Mendelian traits is predicted for loci at mutation-selection balance, where individual mutations are expected to be rare, found at low frequency in natural populations, and gradually eliminated by natural selection. Thus, within a population, a typical gene underlying a Mendelian trait is expected to harbor many distinct mutations.

## Patterns of Genetic Variation Underlying Complex Traits

In contrast to Mendelian traits, the expected patterns of genetic variation at the genes underlying complex traits are far murkier. Although an extensive body of theoretical literature describes different quantitative genetic models of complex traits, relatively little is known about the specific genetic variants that underlie these traits (3, 10). Among the universe of possible models, two classes of models make opposite predictions about the nature of genetic variation underlying complex traits. At one extreme, the “*complex trait-rare variant*” model predicts that phenotypic variation in complex traits will be caused by numerous, individually rare genetic variants at multiple loci (16, 44). Natural populations, however, will harbor a great many distinct genetic variants, and these variants taken together may have a total frequency that is substantial. However, because any individual variant is rare, it is expected to be shared infrequently among subpopulations of a given species. This pattern of great allelic diversity is similar to that seen at loci underlying Mendelian traits. The evolutionary dynamics of such variants are likely to result from a balance between creation of new variation by mutation with loss of variation by purifying selection (i.e. mutation-selection balance). At the other extreme, the “*complex trait-common variant*” model predicts that common disease variants will be few at any particular locus, but be relatively common in natural populations and be shared across multiple subpopulations. These variants are more but not exclusively likely to be maintained through the action of some form of balancing selection. Obviously, any complex trait might well arise from the interaction of both rare and common genetic variants, but, before one can hope to differentiate between these complex trait models, one must first characterize extant patterns of genetic variation in natural populations.

## PATTERNS OF GENETIC VARIATION IN DROSOPHILA AND HUMANS

The earliest molecular studies of genetic variation in both *Drosophila* (47) and humans (33) concentrated on detecting electrophoretic variants in genes. Subsequent studies expanded on these approaches to address fundamental questions in both *Drosophila* (46) and humans (9). With advances in technology, more recent work has focused on determining patterns of DNA sequence variation. This approach has been extensively used to analyze patterns of DNA sequence variation among nuclear loci sampled from *D. melanogaster* natural populations (67). Until recently, the application of similar population genetic methods to analyze human samples has suffered from the lack of relevant data from nuclear genes. The initial study characterizing human nuclear DNA sequence variation analyzed published cDNA and genomic sequences in Genbank and noted low diversity (48). However, this study, being limited to data included in Genbank, could not sample variation in an unbiased fashion and subsequent studies have generally found higher levels

of DNA sequence variation. These studies of single nuclear genes/regions include  $\beta$ -globin (31), dystrophin (86–88), elastin (34), lipoprotein lipase (13, 69), angiotensin-converting enzyme (72), the X-chromosome-specific zinc finger protein (ZFX) locus (37), a non-coding region of the X chromosome (39), and *PDHAI* (32). One limitation of single-region studies is that each of these studies uses different numbers and types of samples. Another limitation is that results may be locus specific. If different loci have had differing evolutionary histories or have experienced differing patterns of selection (34), it may be difficult to draw conclusions that are relevant to all humans from studies of individual loci. An ideal study might simultaneously analyze multiple genomic regions in a common set of samples.

Two recent studies (8, 29) approach this ideal by characterizing patterns of DNA sequence variation in the coding and noncoding regions of a large number of human genes. Because both studies surveyed a large number of loci from a well-defined set of samples, these studies provide the best estimates of patterns of genetic variation in human genes. Comparing the results from these papers with the data generated from *D. melanogaster* leads to a number of conclusions (Table 1).

First, as is evident from the total nucleotide diversity, the absolute level of genetic variation in humans is roughly 10-fold lower than that observed in *D. melanogaster*. Second, the data suggest that the ratio of noncoding to coding variations in *D. melanogaster* is significantly higher than that seen in the human data.

**TABLE 1** Comparison of genetic variation in humans and *Drosophila melanogaster*<sup>a</sup>

Parameter	Values by species (reference) <sup>b</sup>		
	Human		<i>D. melanogaster</i>
	Halushka et al (29)	Cargill et al (8)	Moriyama & Powell (67)
Number of loci	75	106	24
Total nucleotide diversity ( $\times 10^4$ )	8.3 $\pm$ 1.9	5.3 $\pm$ 1.3	70 $\pm$ 58
Coding	8.0 $\pm$ 1.9	5.4 $\pm$ 1.3	40 $\pm$ 31
Synonymous	15.1 $\pm$ 3.6	11.7 $\pm$ 2.9	130 $\pm$ 92
Nonsynonymous	5.7 $\pm$ 1.4	3.4 $\pm$ 0.9	15 $\pm$ 14
Noncoding	8.5 $\pm$ 2.0	5.2 $\pm$ 1.3	105 $\pm$ 80
5' UTR	6.8 $\pm$ 1.7	*	*
Intron	10.5 $\pm$ 2.6	*	*
3' UTR	8.4 $\pm$ 2.0	*	*

<sup>a</sup>Total nucleotide diversity is defined as Watterson's (82) estimator of  $\theta = 4N\mu$ , where  $N$  is population size and  $\mu$  is the per-site, per-generation mutation rate. Values from Halushka et al (29) taken directly from their Table 2. Values from Cargill et al (8) and Moriyama & Powell (67) were recalculated from the data that these authors presented in their Table 2 (8) and Table 1 (67), respectively.

<sup>b</sup>\*, Values not computed.

However, two lines of evidence argue that this is probably an experimental artifact of the human studies. Both human studies largely focused on coding regions and surveyed only limited portions of noncoding sequences. It seems likely that these studies failed to survey large noncoding genomic regions where levels of variation might be expected to be higher. Also, single-locus studies that have surveyed larger noncoding regions, such as at the *LPL* locus (69), reported observed ratios of noncoding to coding variation ( $21 \times 10^{-4} \pm 10 \times 10^{-4}$ ;  $5 \times 10^{-4} \pm 5 \times 10^{-4}$ ) that were similar to the ratios seen in *D. melanogaster*. Better surveys of noncoding DNA surrounding multiple human genes could confirm or reject this interpretation.

The third main result arises from the observation that the frequency of silent-site polymorphism is higher than that seen at nonsynonymous sites. Rates of silent-site polymorphism are also higher than that seen at noncoding sites in both the multiple- and single-locus studies. This pattern of variation suggests that noncoding sequences have functional constraints, such as the presence of regulatory sequences, that influence their pattern of genetic variation. One implication of this observation is that a thorough analysis of both coding and noncoding genetic variants around a candidate locus may be required to identify the specific variants that underlie complex traits.

The fourth main result concerns the role of natural selection and demography in shaping patterns of extant genetic variation in *Drosophila* and humans. Most worldwide *D. melanogaster* populations appear to be derived from East African populations (5). If worldwide populations of *D. melanogaster* recently expanded from East Africa in a single event, this demographic pattern would be expected to influence the patterns of genetic variation at all loci in a similar fashion. In *Drosophila*, significant departures from neutrality have been observed for approximately half of the loci examined (67). It is important that large differences among the patterns of genetic variation at different loci are not compatible with a single population expansion out of East Africa. Humans, like *Drosophila*, are thought to have originated from an ancestral African population (79). Initial comparisons among human single-locus studies observed substantial variation among genes (32) and would appear not to be consistent with recent human expansion from a single ancestral population in Africa. Studies that characterize largely coding regions of genes (8, 29) are unlikely to have sufficient statistical power to make a meaningful conclusion, because the total number of segregating sites at any one locus is usually small. Therefore, the effect of human demographic history and natural selection on extant levels of human genetic variation remains an open question. An ideal study to address this issue would identify SNPs in multiple genomic regions from a set of samples representative of diversity in the human population and then empirically determine the extent of variation among genomic regions.

Finally, perhaps the single best characterized factor influencing patterns of genetic variation in natural populations is the rate of recombination in a genomic region. A number of studies have observed that recombination rates are positively

correlated with extant levels of genetic variation in a wide variety of organisms, including *Drosophila* and humans (1, 4, 30, 68, 78). Two different models suggest how this pattern of low DNA sequence variation in regions with low rates of recombination might arise. The first, referred to as the “hitchhiking effect” (28, 40) supposes that selectively favored substitutions sweep out linked neutral variation. The second model, referred to as the “background selection model” (12), supposes that selection against deleterious alleles maintained by mutation causes a reduction in the amount of DNA sequence variation at linked neutral sites. It remains unclear which model best accounts for this observed pattern of DNA sequence variation in *Drosophila*. However, because the human data are limited to a few loci from the X chromosome, the magnitude of this effect throughout the human genome remains unknown. A study that identified SNPs in multiple genomic regions with various rates of recombination per physical distance would better be able to determine the extent to which recombination influences patterns of human genetic variation.

## The Frequency of Segregating Sites

In the previous description of the expected patterns of genetic variation underlying complex traits, the two models contrasted rare versus common alleles. But what are the definitions of rare and common alleles? Different conceptions of the meaning of rare and common are certainly possible. However, from a population genetics perspective, the definitions can be made for the relevant mutation rates and the type of selection acting on the variant.

The expected frequency of a simple, recessive, deleterious allele, is

$$\sqrt{\frac{u}{s}}, \quad 1.$$

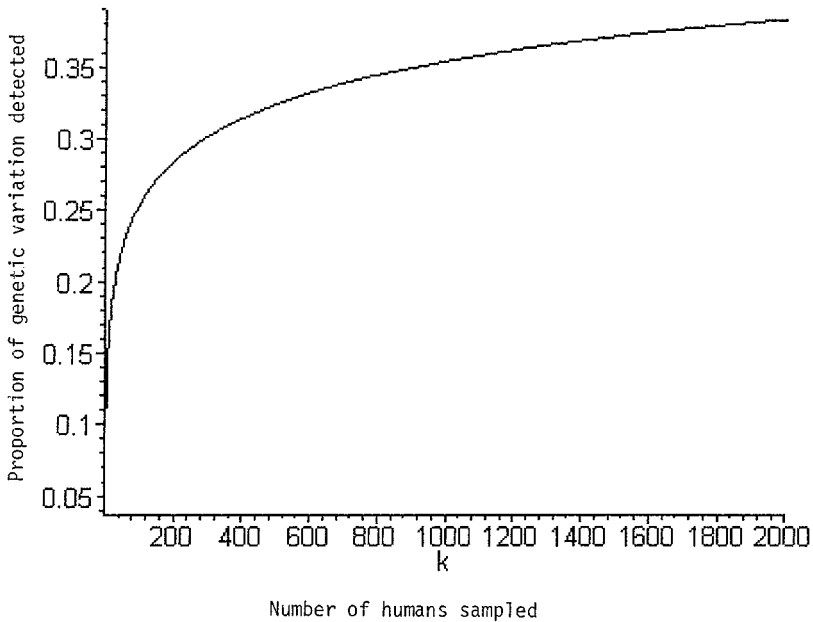
where  $u$  is the mutation rate to this allele and  $1 - s$  is the fitness of a homozygote with this allele [the other homozygote is assumed to have fitness 1.0 (18, p. 259)]. Therefore, if we focus our attention on SNPs and assume that a particular site has a mutation rate of  $\sim 10^{-9}$ /site per generation and that the homozygote is 1% less fit, then the expected frequency of this site is  $< 0.00032$ . The chance that this site will ever reach a frequency of 1% in a large population is astronomically small. Given that the definition of “common” is necessarily somewhat arbitrary, for the purposes of this paper, we regard sites with frequencies  $> 1\%$  as common, and we usually assume that a site that is common cannot be strictly deleterious [whether the site should be thought of as neutral, i.e. no selective effect (42), or under some form of balancing selection (24) remains an open question].

Common and rare sites are expected to be detected at different rates. If most genetic variation is maintained by some form of balancing selection, sites contributing to that variation are likely to be common. If most variation is maintained by mutation-selection balance, these sites are likely to be rare. Therefore, to assess a study’s chance of detecting the sites underlying this variation, it is important to know what variation might be detected in any particular study. If one assumes

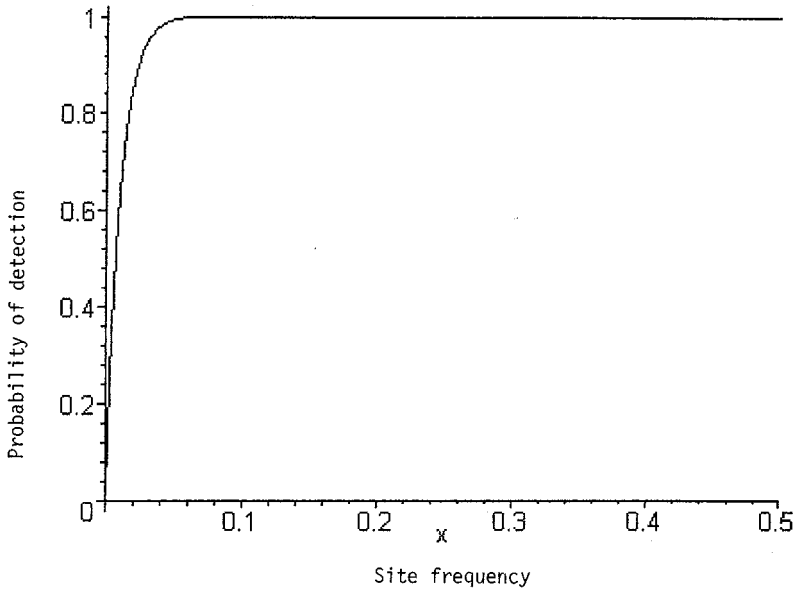
that all sites are neutral, that diploid individuals have formed a single, large population of stable size for a long time, that samples are collected at random, and that sequencing of the samples is accurate, then the expected percentage of all polymorphisms in a given region that are detected by a sequencing study of that region is given by (82):

$$\frac{\sum_{i=1}^{2k-1} \frac{1}{i}}{\sum_{j=1}^{2N-1} \frac{1}{j}}, \quad 2.$$

where  $N$  is population size and  $k$  is the number of individuals sequenced. The basic idea is simple. If one sequences, for example, 100 individuals out of the total population, some fraction of the sites that are polymorphic in the whole population, but happen, by chance, to be monomorphic in the sample, will be missed. If all of these assumptions had been true, the studies of Cargill et al (8) and Halushka et al (29) could have been expected to detect  $\sim 23\%$  of all polymorphic sites in the regions that they studied (this assumes a human population size of 6 billion; if the population in question is much smaller, e.g. 10 million, the figures increase somewhat to 33%). Figure 1 plots the percentage of human variation detected in studies of various sizes. The most striking aspect is the relatively small gains made by increasing the sample size from 100 to 2000.



**Figure 1** Proportion of all variation detected in a sequencing study of  $k$  humans (total human population size,  $N = 6$  billion). (Vertical axis) Proportion of genetic variation detected; (horizontal axis) number of humans sampled.



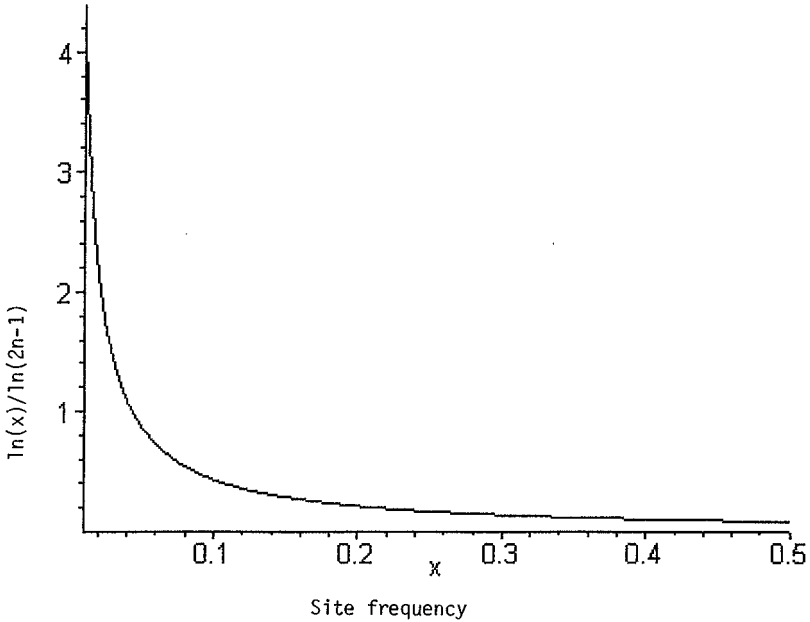
**Figure 2** The probability of detection of a site with frequency  $x$  in a survey of 50 individuals. (*Vertical axis*) Probability of detection; (*horizontal axis*) site frequency.

Thus, it is not unreasonable to suppose that a large sampling study might discover 20%–30% of all human variation in a given region. Of course, the next question is, What proportion of this variation is from common sites, and what proportion is from rare sites? In general, the probability that any given site with frequency  $x$  is observed in a sample of size  $k$  humans is

$$1 - (1 - x)^{2k}. \quad 3.$$

Figure 2 suggests that a study of 50 random individuals is virtually guaranteed to find any site with a 10% frequency (>99.99% probability of detection), but has only a moderate chance (63%) of discovering a site with a frequency of 1%. Such a study has very little chance to discover any truly rare sites (<1% chance of finding a site with frequency of 0.0001). This points out a subtle distinction between the frequency of a site in a sample versus that same site's frequency in the whole population. A site with a frequency of 0.01% in the whole population might, by chance, be detected in a sample of just 50 individuals, but it probably will be missed. A site with a frequency of 10% in the whole population will probably be detected, but very rarely it too might be missed. In general, one might ask what relative proportion of sites with frequency  $x$  is expected in the whole population (75, 83).

Figure 3 shows clearly that the vast majority of sites in the general population are rare. In particular, one expects twice as many sites with frequencies between 0.0001% and 1% as there are sites with frequencies between 1% and 99%.



**Figure 3** The expected proportion of all segregating sites with frequency between  $a$  and  $b$  ( $a < x < b$ ) is given by  $\ln(b/a)/\ln(2N - 1)$ , where  $N$  is the total population size. This proportion can be visualized as the area under the curve between any two points  $a$  and  $b$ .  $N$  was assumed to be 6 billion. (Vertical axis)  $\ln(x)/\ln(2N - 1)$ ; (horizontal axis) site frequency.

It is important to note that these results are certainly dependent on the assumptions of the model. Two assumptions in particular deserve further comment. First, these calculations assumed that population size has remained constant for a long period of time. If human population size has undergone a recent expansion, as is surely the case, then an even higher proportion of the total variation will be rare. Second, these calculations assumed that there was no selection at all on the sites. If most sites contributing to complex variation are in mutation-selection balance, an even larger percentage of sites will be rare. On the other hand, if most sites contributing to complex traits are maintained by balancing selection, then there will be an excess of common sites.

The implications of these simple calculations can profoundly affect the outcome of studies whose aim is to identify genetic variants underlying complex traits. Consider a study design in which 100 individuals drawn at random from the general population are first sequenced to discover a large number of single-nucleotide polymorphisms (SNPs). Suppose that these SNPs are then genotyped in a much larger sample containing variation for a trait of interest (e.g. hypertension). The prospects for the genotyping study to show any association between SNPs and

the trait of interest depend very strongly on the nature of the variation underlying the trait. If most of the phenotypic variation is caused by common sites, then the initial sequencing study is likely to have discovered these sites, and the subsequent genotyping study might be highly successful. On the other hand, if most of the phenotypic variation is caused by rare sites, the initial screen probably missed the vast majority of them, and there is a much lower chance for success in the genotyping study. In other words, a researcher could correctly identify a gene as having a significant effect on a trait, and then identify a large number of SNPs in the region containing this gene, but, because the SNP discovery process was likely to miss most of the rare SNPs, the subsequent genotyping study may fail to show any association between the SNPs discovered and the trait of interest, even though other variants at this gene site do contribute to the trait.

Thus, if most of the genetic variation contributing to complex traits is maintained by mutation-selection balance, there is very little hope that variation in the complex trait is caused by variation in SNPs as ascertained from a general population survey. If most of the contributing variation is rare, a better strategy for a sequencing/genotyping design is to conduct the sequencing study in a population artificially enriched for alleles that cause variation in the trait of interest. With a model organism, this enrichment is accomplished by selective breeding (see next section). Halushka et al (29) hoped to accomplish this enrichment by avoiding "random" individuals, but instead focusing their sequencing efforts on individuals with unusually high or unusually low blood pressure. Unfortunately, this enrichment is presumably trait specific. Thus, although these samples may be enriched for sites contributing to hypertension, there is no necessary reason to suspect that they are enriched for sites contributing to any other trait.

On the other hand, if most variation is maintained by some form of balancing selection, so that many sites causing variation are at high frequency (e.g. >1%), virtually any sequencing/genotyping design might be successful, and, moreover, the SNP discovery portion might be applicable to many traits that are influenced by the assayed genes. This was the approach of Cargill et al (8). They found SNPs from a wide range of individuals selected at random from the population. If most variation is caused by common sites in the regions they surveyed, then they probably detected those sites; subsequent association studies with these sites might be highly successful, and it should be possible to show associations to multiple traits.

Thus, the central question of quantitative genetics, "What is the nature of the forces maintaining quantitative variation?" is of obvious importance to the study of disease in humans. If most variation is maintained by mutation-selection balance, a sequencing/genotyping design must first enrich the sequenced population for alleles of interest, and this process will probably have to be repeated for each separate trait studied. On the other hand, if most variation is maintained by balancing selection, SNPs can be ascertained from a random sample of humans, and some of these SNPs might subsequently show association with any particular trait of interest.

## A MODEL COMPLEX TRAIT: *DROSOPHILA* SENSORY BRISTLES

### Why Study *Drosophila* Sensory Bristles?

The single best studied complex trait in any organism is probably the number of sensory bristles of *D. melanogaster*. *Drosophila* bristles are sensory organs of the peripheral nervous system. *Drosophila* abdominal and sternopleural bristle numbers are a metrical trait that is easily scored and is genetically highly variable. Heritabilities are generally high ( $\sim 1/2$ ) with largely additive genetic variation (19). The genetic pathways involved in the development of bristles are well characterized, with a large number of genes involved in the development of the nervous system and sensory organs acting as possible candidate loci underlying the complex trait (7, 36). The analysis of bristle number in *D. melanogaster* has a long history and has been used to analyze any number of quantitative genetic principles. These include short- (14, 21) and long- (15, 38) term response to selection, selection limits (84); spontaneous mutation rates (reviewed in 35, 41); the effects of recombination (66), population structure (63), and environmental heterogeneity (62) on genetic variation; and response to selection (see review in 59).

### *Drosophila* Sensory Bristles: Mapping Quantitative Trait Loci

Sensory bristles were the first complex trait for which a comprehensive effort was made to estimate map positions of quantitative trait loci (QTLs) ((6, 55, 76, 77, 80), reviewed in (58, 59)). Early experiments to map bristle number QTLs suffered from three major limitations. First, because of the difficulty in obtaining genetic markers, early studies used visible morphological markers that led to a widely spaced, sparse map. Furthermore, the markers themselves often affected bristle number. Second, because of the paucity of markers, a method of progeny testing was used that is expected to be accurate only when the two parental chromosomes are fixed for loci either increasing or decreasing the number of bristles (65). This assumption is unlikely to be met, because, if a large number of loci contributing to bristle number all harbor variations that can either raise or lower bristle numbers, it is unlikely that any genomic region is truly fixed for QTLs that, for example, exclusively raise bristle numbers. Finally, the *Drosophila* stocks examined were either drawn from a small initial population or sampled from lines that had undergone a long-term selection experiment. These experiments often attempted to enrich the sample for alleles that contribute to bristle variation by creating two lines, one that had undergone extremely strong selection to increase bristle number and the other that had undergone similar selection to lower bristle number. With *D. melanogaster* derived from small populations or having experienced strong selection, it is impossible to differentiate between those QTLs that might be

representative of QTLs in natural populations and those that might have arisen within the stocks and/or increased to frequencies unobtainable in nature because of the intensity of the selection. A particular concern is that highly deleterious alleles with large effects on the trait may survive well in an artificial selection experiment, but represent very little of the quantitative variation in nature.

To overcome these problems, more recent studies have used two main approaches (27, 55, 70). First, high and low artificial selection lines were derived from a large natural population sample, and weaker selection was stopped before new mutations were likely to arise. Therefore, the selection lines should have selected among variants segregating in the original natural population. The objective of these experiments was to identify variants that were segregating in the original population and to avoid both newly arising mutations and extraordinarily deleterious variants. To overcome the problems introduced by visual markers, cytological insertion sites of *roo* transposable elements, visualized by in situ hybridization to polytene chromosomes, were used as high-density, highly polymorphic neutral markers, with an average spacing of 4 cM. Because the map was dense, relative to a visual marker map, it is more likely that all QTLs between a pair of markers contribute uniformly to either increase or decrease bristle number. Thus, "high" and "low" chromosomes had unique patterns of *roo* insertion that allowed unambiguous identification of regions with high/low effects on bristle number. These techniques were applied to map QTLs underlying abdominal bristles (55) and sternopleural bristles (27). Both experiments were then repeated (using the same set of base chromosomes, but slightly different QTL mapping algorithms) to determine the generality of the results (70).

These experiments led to two major conclusions. First, a relatively small number of QTLs of large effect could be identified for both abdominal and sternopleural bristle characters. The locations of these QTLs largely mapped to the approximate positions of likely candidate loci that had been previously thought to have effects on bristle number. However, some QTLs failed to map to regions with obvious candidate loci—suggesting that not all loci influencing bristle phenotypes had been identified. The repeatability of both of the original studies was excellent (70), with QTLs in similar regions being mapped in different experiments. This excellent agreement was similar to that seen in older studies and suggests that subsets of loci that affect bristle number are polymorphic in many populations (59), and it strongly suggests that the sites detected in these studies were common in natural populations. The second major conclusion from the QTL mapping studies was that each of the studies mapped a relatively small number of factors with relatively large effects. It is more surprising that the analysis revealed significant sex-specific effects and epistatic interactions between mapped factors that were of the same order of magnitude as the additive effects. This result was quite unexpected given the apparently additive nature of the genetic variance in bristle number that has been inferred from classical breeding studies (19).

## Does Variation in Candidate Loci Cause Quantitative Variation in Bristle Number?

There are several lines of evidence that support the hypothesis that genetic variation at candidate loci causes quantitative variation in bristle number. First, alleles of candidate loci with large effects on bristle number have occasionally been found segregating in natural populations (22, 64). Second, as noted previously, the positions of many QTLs map to positions very near those of candidate loci. However, given the uncertainty associated with QTL location, this must be considered relatively weak evidence for such an association. A third line of evidence comes from complementation to strong alleles and deficiencies of candidate loci. Both naturally occurring (54, 57) and spontaneously mutant (60) high and low alleles at bristle number QTLs genetically interact with mutations at candidate bristle number loci. Fourth, P-element insertional alleles at neurogenic loci have quantitative phenotypic effects on bristle number (56).

The strongest evidence that naturally occurring quantitative variation in bristle number is caused by segregation of genetic variants at candidate loci arises from association studies that have correlated phenotypic variation in bristle number with molecular variation in the regions of candidate neurogenic genes among chromosomes extracted from natural populations. Three main conclusions can be made from these experiments. First, DNA insertion variation in the *achaete-scute* region was found to be strongly associated with variation in bristle number (53, 61; Table 2). Because DNA insertion variants arising from P elements are rare (i.e. never found more than once) and are likely to be deleterious (11), these types of variants are consistent with the rare variants expected under mutation-selection balance. Second, a number of experiments have identified associations between intermediate-frequency, large-effect genetic variants at bristle candidate loci such

**TABLE 2** Proportion of the total genetic variance accounted for by sites significantly associated with bristle number<sup>a</sup>

Locus	Chromosome	Proportion of total abdominal genetic variance	Proportion of total sternopleural genetic variance
<i>achaete-scute</i> , rare variants (61)	X	5%	5%
<i>achaete-scute</i> , common variants (53)	X	22%	25%
<i>Scabrous</i> , common variants (43)	2	13%	8%
<i>Delta</i> , common variants (52)	3	6%	12%

<sup>a</sup>Estimates assume an additive genetic model.

as *scabrous* (43), *Delta* (52), and *achaete-scute* (53) and differences in bristle number (Table 2). This pattern of variation—intermediate-frequency variants with large phenotypic effects—is more consistent with models that incorporate balancing selection than it is with mutation-selection balance models. Alternatively, these variants might be neutral with respect to fitness, but this possibility appears somewhat less likely because variation in bristle number has consistently appeared to be under strong stabilizing selection (23, 49, 71). Third, each of the association studies identified genetic variants that affected males quite differently than females (sex by marker effects). The difference between the effects in males versus females was as large as the main effect of the marker. Sex-specific effects of genetic variants can act as a form of balancing selection to maintain genetic variation in natural populations (25).

Several conclusions can be drawn from these studies. First, genetic variants with an effect on a complex trait can be identified. Second, when detected, this variation often, but not always, appears to reside at previously identified candidate loci. Third, some of the detected variation appears to be maintained strictly by mutation-selection balance. Fourth, other variants appear to be maintained by balancing selection, and a potential mechanism (sex-specific effects) has been identified.

## IDENTIFYING THE GENETIC BASIS OF HUMAN COMPLEX TRAITS AND DISEASES

Determining the genetic basis of complex human traits with DNA sequence variation information is likely to be an important, interesting, and, yet, formidable task. Given what is known about the patterns of human genetic variation and the genetic basis of complex traits such as sensory bristles in *Drosophila*, are there certain experiments that should guide the identification of the genetic variants underlying human complex traits? We believe that the collected observations suggest two main conclusions.

First, the optimal data for identifying the genetic basis of complex traits will consist of the entire DNA sequence (not merely the genotype of several SNPs) of the genomic region under study for all individuals in that study. These data are optimal because they allow investigators to characterize all of the variants in a given sample, not merely the ones that happen to be common in the population at large. Scalable technologies, such as microarrays, may make such an approach feasible in the future. Absent this technological advance, the alternative approach has been to discover SNPs in one sample and then genotype them in an experimental population. The success of this approach depends on two important factors: the choice of the SNP discovery samples and the genetic basis of complex traits. If the genetic variants underlying complex traits are relatively common, then one is likely to discover and genotype variants that are relevant to the complex trait. If complex traits consist largely of rare variants at mutation-selection balance, then it is more likely that the initial SNP discovery will miss these variants, and there

is a much lower chance of success, even when the researchers have successfully identified the correct region to study.

The *Drosophila* sensory bristle experiments exemplify these conclusions in two ways. First, one of the major technical challenges of each study was to efficiently identify a collection of genetic variants in the regions under study. Techniques allowing rapid identification of all sequence variants in a genomic region would have significantly aided these studies. Second, the *Drosophila* sensory bristle experimental results more often identified large-effect, intermediate-frequency variants than they did rare variants, but this could be an artifact of the difficulty in identifying all genetic variations in a region. Nevertheless, if this pattern of genetic variation is similar to that in humans, it suggests that at least some high-frequency variants contribute to complex traits. Therefore, association-based approaches that determine the genotype of common SNPs will at least sometimes successfully detect variants underlying human complex traits (50, 51, 73). Furthermore, when an association was shown between rare, probably deleterious variants (as with P-elements at *achaete-scute*), it was accomplished only on variants that had previously been identified to be rare, putatively deleterious, and specifically studied because they were known to have these properties. Identifying large numbers of human genetic variants that are known to be deleterious and individually rare may prove challenging.

A second main conclusion arising from the *Drosophila* studies is the seeming importance of candidate loci and noncoding genetic variants at those loci. Most but not all QTL markers have been found at or near candidate loci, which suggests that focusing our attention on human candidate loci may prove profitable. It also suggests that focusing on candidate loci to the exclusion of all other genetic regions will almost certainly result in missing some of the variation in complex traits. Furthermore, all of the association studies identified noncoding genetic variants that increased or decreased bristle number. Although this association does not prove that these noncoding variants cause the phenotypic difference (i.e. are the quantitative-trait nucleotides), there were no detected missense mutations segregating in the samples that could account for these associations. These results imply that the genetic variants causing the phenotypic difference might well be noncoding. As a consequence, a complete picture of the patterns of genetic variation of genomic regions that contain candidate loci is likely to be required to identify genetic variants underlying complex traits. Focusing solely on missense mutations, which may be more likely to be population specific and whose dynamics might be more consistent with mutation-selection balance, may lead researchers to miss common variants that underly complex traits. In *Drosophila*, even at a locus such as *ADH* that harbors a common amino acid polymorphism, a noncoding variant with an effect on enzyme activity as large as the missense mutation has been observed to be segregating in natural populations (45). Also, the repeated observation of large sex-specific effects of the markers showing association was quite unexpected, based on earlier breeding studies, and needs to be considered in future studies of complex human traits.

Thus, three possible scenarios emerge. If most variation in complex traits is maintained by balancing selection, then an approach that first discovers common SNPs in a group of randomly selected individuals, followed by genotyping of those SNPs in a population, which shows variation for the complex trait, is highly likely to be successful, because the initial survey was likely to have discovered SNPs that contribute to the variation. If, on the other hand, most variation is maintained by mutation-selection balance, there are two possible approaches. If the initial SNP discovery can occur in a sample that is already enriched for sites causing variation, then SNPs contributing to the variation may be directly discovered, despite their rarity in the general population. Alternatively, if the SNP discovery occurs in randomly selected individuals, the success of the genotyping survey will depend on the amount and pattern of linkage disequilibrium between the discovered SNPs and the sites that actually contribute to variation in the trait of interest. Even if most variation is maintained by mutation-selection balance, common SNPs that are discovered in a random sample of individuals still might be in linkage disequilibrium with sites that cause variation in complex traits.

The challenges facing geneticists to understand the genetic basis of complex human traits are enormous. The genetic dissection of complex traits in model systems, such as *Drosophila*, might well be expected to significantly aid the discovery of the genetic basis of complex traits in humans, where manipulative experiments are impossible. Indirect genetic methods used to understand human complex traits will be profoundly dependent on the patterns of human genetic variation and the different evolutionary forces acting to maintain variation in human populations.

**Visit the Annual Reviews home page at [www.AnnualReviews.org](http://www.AnnualReviews.org)**

#### LITERATURE CITED

1. Aguade M, Miyashita N, Langley CH. 1989. Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. *Genetics* 122:607–15
2. Altenburg E, Muller HJ. 1920. The genetic basis of truncate wing, an inconstant and modifiable character in *Drosophila*. *Genetics* 5:1–59
3. Barton NH, Turelli M. 1989. Evolutionary quantitative genetics: how little do we know? *Annu. Rev. Genet.* 23:337–70
- 3a. Bate M, Arias AM, eds. 1993. *The Development of Drosophila melanogaster*. Plainview, NY: Cold Spring Harbor Lab. Press. 1564 pp.
4. Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356:519–20
5. Begun DJ, Aquadro CF. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* 365:548–50
6. Breese EL, Mather K. 1957. The organization of polygenic activity within a chromosome in *Drosophila*. I. Hair characters. *Heredity* 11:373–95
7. Campos-Ortega JA. 1993. Early neurogenesis in *Drosophila melanogaster*. See Ref. 3a, pp. 1091–129
8. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, et al. 1999. Characterization of single-nucleotide polymorphisms in

- coding regions of human genes. *Nat. Genet.* 22(3):231–38
9. Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton Univ. Press. 1088 pp.
  10. Chakravarti A. 1999. Population genetics: making sense out of sequence. *Nat. Genet.* 21(Suppl.):56–60
  11. Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* 23: 251–87
  12. Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134(4):1289–303
  13. Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, et al. 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.* 63(2):595–612
  14. Clayton GA, Morris JA, Robertson A. 1957. An experimental check on quantitative genetical theory. I. Short-term responses to selection. *J. Genet.* 55:131–51
  15. Clayton GA, Robertson A. 1957. An experimental check on quantitative genetical theory. II. The long-term effects of selection. *J. Genet.* 55:152–70
  16. Collins FS, Guyer MS, Charkravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1580–81
  17. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, et al. 1998. New goals for the U.S. Human Genome Project: 1998–2003. *Science* 282:682–89
  18. Crow JF, Kimura M. 1970. *An Introduction to Population Genetics Theory*. New York: Harper & Row. 591 pp.
  19. Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*. Burnt Mill, Harlow, Essex, England: Longman. 4th ed. 464 pp.
  20. FlyBase Consortium. 1999. The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* 27(1):85–88
  21. Frankham R. 1968. Sex and selection for a quantitative trait in *Drosophila*. II. The sex dimorphism. *Aust. J. Biol. Sci.* 21:1225–37
  22. Frankham R, Nurthen RK. 1981. Forging links between population and quantitative genetics. *Theor. Appl. Genet.* 59:251–63
  23. Garcia-Dorado A, Gonzalez JA. 1996. Stabilizing selection detected for bristle number in *Drosophila melanogaster*. *Evolution* 50:1573–78
  24. Gillespie JH. 1991. *The Causes of Molecular Evolution*. New York: Oxford Univ. Press. 336 pp.
  25. Gillespie JH, Turelli M. 1989. Genotype-environment interactions and the maintenance of polygenic variation. *Genetics* 121:129–38
  26. Green MM. 1996. The “genesis of the white-eyed mutant” in *Drosophila melanogaster*: a reappraisal. *Genetics* 142(2):329–31
  27. Gurganus MC, Nuzhdin SV, Leips JW, Mackay TFC. 1999. High-resolution mapping of quantitative trait loci for sternopleural bristle number in *Drosophila melanogaster*. *Genetics* 152(4):1585–604
  28. Haigh J, Smith JM. 1976. The hitch-hiking effect—a reply. *Genet. Res.* 27(1):85–87
  29. Halushka MK, Tan JB, Bentley K, Hsie L, Shen NP, et al. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* 22(3):239–47
  30. Hamblin MT, Aquadro CF. 1999. DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. *Genetics* 153(2):859–69
  31. Harding RM, Fullerton SM, Griffith RC, Bond J, Cox MJ, et al. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* 60(4):772–89
  32. Harris EE, Hey J. 1999. X chromosome

- evidence for ancient human histories. *Proc. Natl. Acad. Sci. USA* 96(6):3320–24
33. Harris H. 1966. Enzyme polymorphisms in man. *Proc. R. Soc. London Ser. B* 164: 298–310
34. Hey J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. *Mol. Biol. Evol.* 14(2):166–72
35. Houle DB, Morikawa B, Lynch M. 1996. Comparing mutational variabilities. *Genetics* 143(3):1467–83
36. Jan YN, Jan LY. 1993. The peripheral nervous system. See Ref. 3a, pp. 1207–44
37. Jaruzelska J, Zietkiewicz E, batzer M, Cole DEC, Moisan JP. 1999. Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics* 152(3):1091–101
38. Jones LP, Frankham R, Barker JSF. 1968. The effects of population size and selection intensity in selection for a quantitative character in *Drosophila*. II. Long-term response to selection. *Genet. Res.* 12: 249–66
39. Kaessman H, Heissig F, von Haeseler A, Poabo S. 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat. Genet.* 22(1):78–81
40. Kaplan NL, Hudson RR, Langley CH. 1989. The “hitchhiking effect” revisited. *Genetics* 123(4):887–99
41. Keightley PD, Mackay TF, Caballero A. 1993. Accounting for bias in estimates of the rate of polygenic mutation. *Proc. R. Soc. London B Biol. Sci.* 253:291–96
42. Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge Univ. Press. 367 pp.
43. Lai C, Lyman RF, Long AD, Langley CH, Mackay TFC. 1994. Naturally occurring variation in bristle number and DNA polymorphisms at the scabrous locus of *Drosophila melanogaster*. *Science* 266:1697–702
44. Lander ES. 1996. The new genomics: global views of biology. *Science* 274: 536–39
45. Laurie CC, Stam LF. 1994. The effect of an intronic polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* 138(2):379–85
46. Lewontin RC. 1974. *The Genetic Basis of Evolutionary Change*. Columbia Biological Series No. 25. New York: Columbia Univ. Press. 346 pp.
47. Lewontin RC, Hubby JL. 1966. A molecular approach to the study of genic heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595–609
48. Li WH, Sadler LA. 1991. Low nucleotide diversity in man. *Genetics* 129(2):513–23
49. Linney R, Barnes BW, Kearsey MJ. 1971. Variation for metrical characters in *Drosophila* populations. 3. The nature of selection. *Heredity* 27(2):163–74
50. Long AD, Grote MN, Langley CH. 1997. Genetic analysis of complex diseases. *Science* 275:1328–30
51. Long AD, Langley CH. 1999. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* 9(8): 720–31
52. Long AD, Lyman RF, Langley CH, Mackay TFC. 1998. Two sites in the delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* 149(2):999–1017
53. Long AD, Lyman RF, Morgan AH, Langley CH, Mackay TFC. 2000. Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. *Genetics* 154(3):1255–69
54. Long AD, Mullaney SL, Mackay TFC, Langley CH. 1996. Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number

- in *Drosophila melanogaster*. *Genetics* 144(4):1497–510
55. Long AD, Mullaney SL, Reid LA, Fry JC, Langley CH, Mackay TFC. 1995. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* 139(3):1273–91
56. Lyman RF, Lawrence F, Nuzhdin SV, Mackay TFC. 1996. Effects of single P-element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* 143(1):277–92
57. Lyman RF, Mackay TF. 1998. Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in *Drosophila melanogaster*: the delta-hairless gene region. *Genetics* 149(2):983–98
58. Mackay TF. 1995. The genetic basis of quantitative variation: numbers of sensory bristles of *Drosophila melanogaster* as a model system. *Trends Genet.* 11(12):464–70
59. Mackay TF. 1996. The nature of quantitative genetic variation revisited: lessons from *Drosophila* bristles. *BioEssays* 18(2):113–21
60. Mackay TF, Fry JD. 1996. Polygenic mutation in *Drosophila melanogaster*: genetic interactions between selection lines and candidate quantitative trait loci. *Genetics* 144(2):671–88
61. Mackay TF, Langley CH. 1990. Molecular and phenotypic variation in the achaete-scute region of *Drosophila melanogaster*. *Nature* 348:64–66
62. Mackay TFC. 1981. Genetic variation in varying environments. *Genet. Res.* 37:79–93
63. Madalena FE, Robertson A. 1975. Population structure in artificial selection: studies with *Drosophila melanogaster*. *Genet. Res.* 24:113–26
64. McBride G, Robertson A. 1963. Selection using assortative mating in *Drosophila melanogaster*. *Genet. Res.* 4:356–69
65. McMillan I, Robertson A. 1974. The power of methods for the detection of major genes affecting quantitative characters. *Heredity* 32(3):349–56
66. McPhee CP, Robertson A. 1970. The effect of suppressing crossing-over on the response to selection in *Drosophila melanogaster*. *Genet. Res.* 16:1–16
67. Moriyama EN, Powell JR. 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* 13(1):261–77
68. Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* 150(3):1133–41
69. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinsons RG, et al. 1998. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat. Genet.* 1998. 19(3):233–40
70. Nuzhdin SV, Dilda CL, Mackay TF. 1999. The genetic architecture of selection response: inferences from fine-scale mapping of bristle number quantitative trait loci in *Drosophila melanogaster*. *Genetics* 153(3):1317–31
71. Nuzhdin SV, Fry JD, Mackay TF. 1995. Polygenic mutation in *Drosophila melanogaster*: the causal relationship of bristle number to fitness. *Genetics* 139(2):861–72
72. Rieder MJ, Taylor SL, Clark AG, Nickerson DA. 1999. Sequence variation in the human angiotensin converting enzyme. *Nat. Genet.* 22(1):59–62
73. Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* 273:1516–17
74. Robertson A. 1967. The nature of quantitative genetic variation. In *Heritage from Mendel*, ed. A Brink, pp. 265–80. Madison, WI: Univ. Wisc. Press
75. Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. *Genetics* 132(4):1161–76
76. Shrimpton AE, Robertson A. 1988. The isolation of polygenic factors controlling

- bristle score in *Drosophila melanogaster*. I. Allocation of third chromosome bristle effects to chromosome sections. *Genetics* 118:437–43
77. Shrimpton AE, Robertson A. 1988. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. *Genetics* 118:445–59
78. Stephan W, Langley CH. 1998. DNA polymorphism in lycopersicon and crossing-over per physical length. *Genetics* 150(4):1585–93
79. Stringer CB, Andrews P. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* 239:1263–68
80. Thoday JM. 1979. *Polygene Mapping: Uses and Limitations*, ed. JM Thoday, JN Thompson, pp. 219–33. New York: Academic
81. Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M. 1998. Shotgun sequencing of the human genome. *Science* 280:1540–42
82. Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7(2):256–76
83. Wright S. 1938. The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. USA* 24:253–59
84. Yoo BH. 1980. Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster*. I. Response to selection. *Genet. Res.* 35: 1–17
85. Zielenski J, Tsui LC. 1995. Cystic fibrosis: genotypic and phenotypic variations. *Annu. Rev. Genet.* 29:777–807
86. Zietkiewicz E, Akalin N, Labuda D. 1995. Neutral polymorphisms in the deletion-prone regions of the dystrophin gene. *Hum. Hered.* 45(2):80–83
87. Zietkiewicz E, Yotova V, Jarnik M, Korablaskowska M, Kidd KK, et al. 1997. Nuclear DNA diversity in worldwide distributed human populations. *Gene* 205 (1–2):161–71
88. Zietkiewicz E, et al. 1998. Genetic structure of the ancestral population of modern humans. *J. Mol. Evol.* 47(2):146–55