

POPULATION GENOMICS: Genome-Wide Sampling of Insect Populations

William C. Black IV,¹ Charles F. Baer,²
Michael F. Antolin,³ and Nancy M. DuTeau¹

Departments of ¹Microbiology and ³Biology, Colorado State University, Fort Collins, Colorado 80523; e-mail: wcb4@lamar.colostate.edu, antolin@lamar.colostate.edu, nduteau@cvmb.colostate.edu, ²Ecology and Evolution Program, University of Oregon, Eugene, Oregon 97403-1210; e-mail: cbaer@darkwing.uoregon.edu

Key Words adaptation, DNA, gene mapping, neutral evolution, population genetics

■ **Abstract** Modern population genetics underwent a major paradigm shift during the last decade of the 20th century with the discovery that thousands of genes of known function and position in a genome can be analyzed simultaneously in a single individual. The impact of this technology on insect population genetics is potentially profound. Sampling distributions of genetic statistics can now be derived from many individual loci or among many segregating sites within a gene. Inferences regarding random mating, gene flow, effective population sizes, disequilibrium, and relatedness among populations can now be based on patterns of variation at many loci. More importantly, genome-wide sampling enables population geneticists to distinguish effects that act on the whole genome from those that act on individual loci or nucleotides. We introduce the term “population genomics” to describe the process of simultaneous sampling of numerous variable loci within a genome and the inference of locus-specific effects from the sample distributions. The four critical assumptions implicit in the population genomics approach are explained in detail. Studies adopting this paradigm are reviewed, and the steps necessary to complete a population genomics study are outlined.

CONTENTS

INTRODUCTION	442
BACKGROUND	443
A Hypothetical Example of the Population Genomics Approach	444
Detecting Selection and Neutral Evolution	445
EXAMPLES OF THE POPULATION GENOMICS APPROACH	448
CRITICAL ASSUMPTIONS OF POPULATION GENOMICS	450
Assumption 1: Genes Evolve Independently	450
Assumption 2: Genome-Wide Effects Influence All Parts of the Genome Equally	453

Assumption 3: Analysis of Single-Nucleotide Polymorphisms at Many Loci	455
Assumption 4: Relative Locations of Genetic Markers	455
INITIATING POPULATION GENOMIC STUDIES	456
Data Collection	456
Data Analysis	457
Nucleotide Polymorphisms	458
Detecting Selection	459
Sample Size and Design Are Critical in Population Genomics	460
CONCLUSION	461

INTRODUCTION

The study of adaptation remains a primary focus of entomology. The genetics of adaptation is fundamentally about gene substitution within populations over time: alleles favored by natural selection replace other alleles and at the same time change traits within populations. Historically, the genetics of adaptation has been approached in three ways: by quantitative genetic analysis of phenotypic variation (85, 112), by marker-based population genetics (42), and by analyses of single-gene enzyme polymorphisms (38). Differences between these approaches are rapidly disappearing with the advent of “genome projects” that map and sequence the entire genomes of insects [e.g. *Drosophila melanogaster* (1)], the development of computer-based statistical procedures (bioinformatics) for examining the resulting data sets, and the creation of genetic maps that identify loci affecting phenotypic variation [quantitative trait loci (QTL)]. On one hand, entomologists interested in phenotypic evolution can apply a wide array of molecular techniques to dissect phenotypes into their underlying genetic causes. On the other hand, population genetic theory provides statistical tools to infer the evolutionary forces acting on genes. This suggests an integrated strategy capable of simultaneously describing the genetic basis of phenotypic variation and the evolution of those genes that underlie that variation. Successful application of this strategy can provide insights into problems in evolutionary entomology such as insecticide resistance, pathogen transmission, plant-herbivore interactions, sex determination, pheromone use, and the genetics of speciation.

We use the term “population genomics” to describe this process of simultaneous sampling of numerous variable loci within a genome and the inference of locus-specific effects from the sample distributions. Locus-specific effects include selection, mutation, recombination, and assortative mating, processes that should affect one or a few genes at a time. Locus-specific effects contrast with genome-wide effects like genetic drift (including founder events and population bottlenecks), migration, and inbreeding, which should affect all parts of a genome in the same way. We are careful to distinguish the terms “gene” and “locus.” By “gene” we refer to a region of DNA that codes for RNA, regardless of whether the RNA is ultimately modified and translated into enzymes and proteins or has other

functions within the cell (e.g. transfer RNA or ribosomal RNA). “Locus” refers to the location of a particular DNA sequence in a genome, whether it codes for RNA or not. Most importantly for this review, in the age of DNA sequencing, locus also refers to a single nucleotide among the hundreds of millions that compose the entire genome of a typical insect. A nucleotide position in a gene that varies among individuals is referred to as a segregating site.

We acknowledge that insect genomes are complex structures that include non-transcribed heterochromatin, centromeres and telomeres, various kinds of repetitive DNA, and the DNA found in organelles such as mitochondria. However, our discussion of population genomics is generally restricted to the single-copy nuclear (scnDNA) genes that form the genetic basis for the majority of adaptations. We avoid detailed discussion of the genetics of multigene families, like nuclear ribosomal-RNA genes, because their molecular evolution is greatly affected by the forces of molecular drive and therefore is necessarily correlated. The dependent, concerted evolution of repetitive DNA loci violates the first critical assumption of population genomics as discussed below. The lack of recombination in mitochondrial DNA also violates the first assumption so that population genomics cannot be applied towards identifying adaptation in individual mitochondrial genes. However, as illustrated in some of the examples that follow, the mitochondrial genome is an excellent tool for studying genome-wide effects (27, 111).

BACKGROUND

The paradigm shift toward population genomics began during the last decade of the 20th century with the discovery that hundreds or thousands of genetic markers, of known position in a genome, can be analyzed simultaneously in a single organism. Four technologies were critical in developing population genomics in insects. First, the polymerase chain reaction (PCR) allowed the amplification of many loci from the small amounts of genomic DNA typically isolated from a small insect (e.g. *Drosophila* spp., parasitic wasps, mosquitoes, and aphids). Second, a variety of highly polymorphic markers, like microsatellites, random amplified polymorphic DNAs, and amplified fragment length polymorphisms, were discovered to be abundant in arthropod genomes. Third, techniques for detecting single-nucleotide polymorphisms (SNPs) within genes allowed genetic variation to be examined in even the most conserved genes. Finally, statistical algorithms were developed to analyze recombination among the many genetic markers simultaneously segregating in one or a few genetic crosses, thereby allowing insect geneticists rapidly and inexpensively to construct linkage maps of the relative positions of markers in an insect genome (50, 72, 137).

The impact of these technologies on genetic studies of adaptation in insects is profound. Entomologists can now choose genetic markers with which to analyze variation throughout an entire genome; it is no longer necessary to use one or

a few loci to draw inferences regarding selection, random mating, gene flow, effective population sizes, multilocus disequilibrium, and relatedness among insect populations. Population genetic analyses can now be conducted by sampling insect genomes and estimating population statistics at two levels: across variable loci dispersed throughout the genome and among polymorphic-nucleotide sites within individual genes.

On the whole-genome level, a large number of loci dispersed throughout the genome can be sampled, and distributions of genetic statistics can be estimated for the sampled loci. Statistics will have sampling distributions based on locus-to-locus variation with extreme values at the tails of the distributions. The extreme values may represent locus-specific effects—forces that directly influence genes and their surrounding loci within those regions of the genome. Locus-specific effects include selection, mutation, recombination, and nonrandom mating. Estimates from the majority of loci will fall into the centers of the distributions and represent loci affected the same way across the genome. These are loci influenced by genome-wide effects that include genetic drift (founder effects and population bottlenecks), migration, and inbreeding.

A Hypothetical Example of the Population Genomics Approach

The power of population genomics is that locus-specific and genome-wide effects can be distinguished by examining variation at many loci or even many segregating nucleotides within a locus. As a hypothetical example, consider a situation in which we are attempting to identify genomic regions in a crop pest that confer virulence to a novel host plant cultivar. We examine variation at 1000 cDNA loci distributed throughout the genome of the insect. In practice, examination of this many loci would be impractical. Instead we would identify “candidate” loci that we suspect may control the phenotype either for a priori mechanistic reasons (e.g. a digestive enzyme), or because they have been found to cosegregate with the phenotype in linkage mapping studies. At the level of individual genes, we wish to make estimates of variation in allele and genotype frequencies within and among populations differentially exposed to the novel cultivar.

The statistical measurements that are frequently discussed in this review (and in the population genomics literature) for describing and analyzing this type of variation are F_{is} and F_{st} (148). F_{is} is a measure of the observed numbers of heterozygotes (H_{obs}) relative to the expected number of heterozygotes (H_{exp}) under assumptions of the Hardy-Weinberg model. F_{is} can vary from -1 to 1 . If $H_{obs} > H_{exp}$, then $F_{is} < 0$. This might occur if, for example, heterozygotes for a virulence allele have a higher fitness than insects homozygous for that allele. When $H_{obs} < H_{exp}$, then $F_{is} > 0$, and this might occur if, for example, heterozygotes for a virulence allele have a lower fitness than insects homozygous for that allele. F_{st} is a measure of the standardized variance in the frequency of an allele among populations and

can vary from 0 (allele frequencies are approximately equal among all populations) to 1 (alternate alleles fixed in different populations). We might expect $F_{st} \gg 0$ at a locus that overcomes host plant resistance when compared among populations differentially exposed to the novel cultivar.

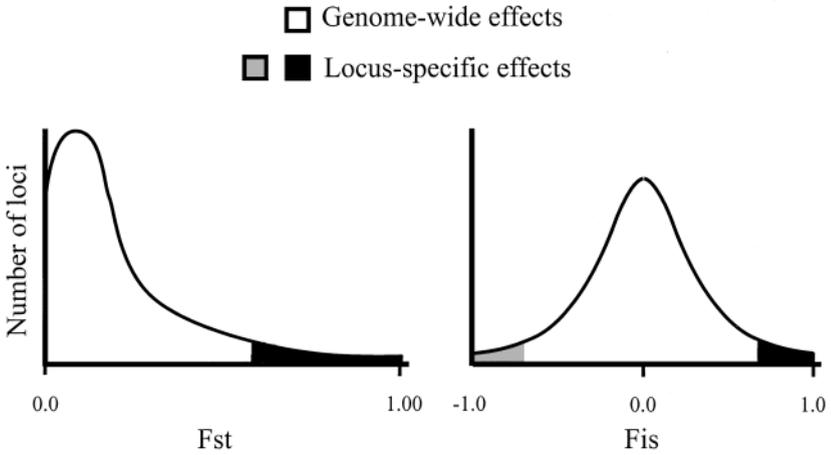
Assume that alleles at 1 of the 1000 cDNA loci that we examined actually confer a higher fitness on insects in the presence of the novel cultivar. F_{st} at the 999 other loci will follow a Poisson (random) distribution (Figure 1A) owing largely to genome-wide effects. However, F_{st} at the locus that confers virulence to the novel cultivar will exceed the expected variance because it is subject to locus-specific selection and will fall within the black region of the sampling distribution. Assume that, at that same locus, heterozygotes for the virulence allele have a higher fitness than insects homozygous for that allele, which might occur if homozygotes for either virulence or wild-type alleles have a low fitness relative to virulence heterozygotes. If we also assume that insects mate at random and that alleles at the remaining 999 loci have no major impact on survival or reproduction, then F_{is} at these loci will approximate 0 (Figure 1A) owing to genome-wide effects. However, $H_{obs} > H_{exp}$, and $F_{is} < 0$ at the locus conferring virulence. This reflects locus-specific selection, and F_{is} for this situation will fall within the gray region of the F_{is} curve. Alternatively, assume that, at the same locus, homozygotes for the virulence allele have a higher fitness than heterozygotes or wild-type homozygotes. Then $H_{obs} < H_{exp}$, and $F_{is} > 0$ at the virulence locus. F_{is} will fall within the locus-specific black region of the F_{is} curve.

At a finer scale, population genomics can identify locus-specific effects at individual nucleotide positions. After screening our 1000 cDNA clones, we decided to focus on the one cDNA at which F_{is} and F_{st} behave as indicated in Figure 1A. We examined nucleotide variation across 400 nucleotides in that cDNA. Assuming that the actual nucleotide substitutions that confer virulence to the novel cultivar are segregating at site 200, we can now consider the outcomes of several different types of selection.

Detecting Selection and Neutral Evolution

The neutral theory of molecular evolution provides a context for testing hypotheses in population genomics, wherein existing genetic variation is compared with what should arise purely by mutation. The alternative hypothesis is that genome-wide or locus-specific variation within and between populations is adaptive, arising from selection that reduces variation (i.e. is purifying or directional) or that maintains genetic variation (i.e. is balancing or diversifying) (7). Neutral evolution assumes that most mutations in a gene are lethal but that the negative effects on fitness of the remaining mutations will be small enough that they will be neutral or nearly neutral. Neutral theory (66) shows that the majority of neutral or nearly neutral mutations will become extinct through random chance but that a few, over many generations, will become fixed.

A. Sampling distributions among loci



B. Sampling distributions within a gene

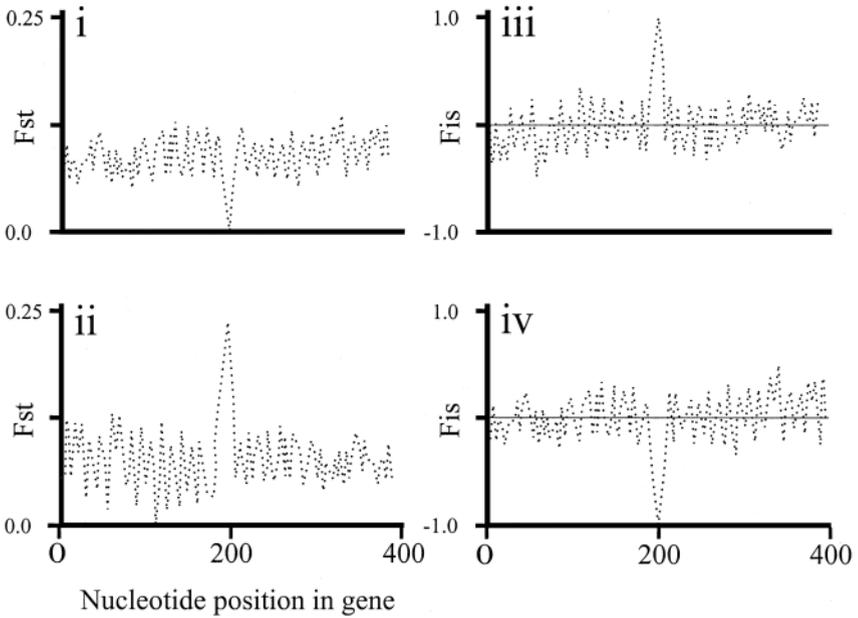


Figure 1 Hypothetical sampling distributions of F_{is} and F_{st} among (A) loci distributed throughout a genome and (B) segregating sites in a gene. See text for further description.

Purifying selection is the component of the neutral evolution model that rapidly eliminates any nucleic acid substitutions that reduce the fitness of an organism, and this concept is often used to explain the absence of nucleotide variation observed at many loci. The combined influences of neutral evolution and purifying selection are frequently cited to explain the overall variation observed in a gene (42, 56). Purifying selection homogenizes populations by removing alleles that might arise in a population by migration or mutation and by preventing divergence in allele frequencies between populations by genetic drift. Assume that, in the absence of the newly resistant plant cultivar, fitness is maximized in insects with a particular nucleotide at site 200. Under purifying selection (Figure 1*Bi*), levels of differentiation (F_{st}) are lower at nucleotide position 200 than at surrounding nucleotides.

Positive selection occurs when an allele, new to a species through mutation or new to a population through migration, confers a relative fitness advantage to an organism and thereby increases in frequency. Positive selection can explain clines of variation among geographic or seasonal populations or differences in allele frequencies among nearby populations that differ in some important environmental variable. Positive selection might occur in populations exposed to the novel resistant cultivar. Insects with a novel nucleotide at position 200 might have a greater fitness than insects with the original nucleotide at that position. Under positive selection, F_{st} at nucleotide position 200 would be greater than the average F_{st} at surrounding nucleotides (Figure 1*Bii*). Furthermore, there would be an apparent excess of homozygotes for the novel nucleotide at position 200, and the inbreeding coefficient (F_{is}) would exceed the average F_{is} measured at other variable nucleotide sites across the gene (Figure 1*Biii*).

Balancing selection maintains genetic variation and can occur in a number of distinct ways (131). Most commonly, balancing selection is thought of as favoring heterozygous individuals (overdominance). Correlations between overall heterozygosity and fitness have been described in a broad number of organisms (94), but the data can be explained generally by partial dominance (30, 56), and direct evidence of overdominance at individual loci is lacking (7, 28).

A more plausible mechanism of balancing selection is through frequency-dependent selection, in which individual alleles confer their highest fitness when rare and become less favored at higher frequencies. This is thought to be the mechanism that maintains extremely high allelic diversity at the major sex-determining locus of some Hymenoptera (35, 51, 125). In this genetic system of sex determination, fertilized (diploid) eggs develop as females only if they are heterozygous for any two alleles segregating at the locus. If diploid individuals are homozygous at the sex locus, then they develop as diploid males with extremely low fitness (35, 54, 125). As a consequence, new alleles are always favored when they first appear in populations but confer lower fitness when they are more common and have a higher probability of becoming homozygous. Frequency-dependent selection is thought to affect a wide variety of insect adaptations including sex

ratio variation (11), female mate choice, male mating tactics (31), and mimicry in butterflies (88).

A third mechanism of balancing selection arises when mutations occur that confer a fitness advantage in an organism but only in certain environments or circumstances. For example, a novel allele for insecticide resistance arises and is subject to positive selection in one environment but has a lower fitness in the absence of that insecticide and is subject to negative selection in an insecticide-free environment. This form of balancing selection, when coupled with habitat selection or limited mobility, is thought to be the basis of local adaptation in insects (96). Balancing selection may also have different effects according to sex, and there is increasing evidence that this may maintain variation in *Drosophila* spp. (28, 138).

Under balancing selection, F_{st} at nucleotide position 200 would be greater than the average F_{st} at surrounding nucleotides (Figure 1Bii) among populations differentially exposed to the novel plant cultivar. However, there would be an apparent excess of heterozygotes at nucleotide position 200. F_{is} would be lower than the average F_{is} measured at other segregating sites across the gene (Figure 1Biv).

Diversifying selection applies to proteins that encode highly variable proteins and therefore appear to be adapted to a potential diversity of environmental conditions. Such genes might encode salivary gland proteins susceptible to host immunity or be associated with overcoming host plant resistance, or they could be involved in conferring resistance to a parasite or pathogen. As with frequency-dependent selection, insects with new alleles and genotypes at a gene might have a higher fitness than individuals with more common genotypes and alleles. Models of diversifying selection have been examined in a wide variety of different systems (33).

In practice, without understanding the mode of action of all or parts of an enzyme, it may be difficult to differentiate the forces of diversifying selection and neutral evolution. However, comparison of the rate and spectrum of mutations in proximal genes may help differentiate these two forces (32, 135). Population genomics predicts that nucleotide diversity indices (described below) at the segregating sites in the gene subject to diversifying selection are much greater than the average diversity at segregating sites in nearby genes or flanking loci. However, this prediction becomes invalid if adjacent sites are also rapidly evolving.

EXAMPLES OF THE POPULATION GENOMICS APPROACH

Four recent studies serve as excellent examples of the population genomics approach. In *Heliothis virescens*, a major lepidopteran pest of cotton, variation was examined at 13 allozyme and two cDNA loci [*Hpy* (the sodium channel locus), and *Hejs* (the juvenile hormone esterase locus)] among five populations from Georgia to Arizona (129)]. The allozyme and the *Hejs* loci indicated little

genetic differentiation among population ($F_{st} = 0.001$) whereas large differences were detected at the *Hpy* locus ($F_{st} = 0.041$). Mutations at the *Hpy* locus are known to confer resistance to pyrethroid insecticides; a negative correlation was demonstrated between allelic diversity and the degree of resistance to pyrethroid insecticides. Populations from Louisiana are subject to intensive insecticide applications and had the lowest allelic diversity, suggesting that selection removed susceptible alleles. These results indicate the power of the population genomics approach for uncovering selection acting upon a single locus. While the majority of markers indicate free gene flow among *H. virescens* populations throughout the southern United States and Mexico, local positive selection from pyrethroid insecticides caused large locus-specific variation at the gene conferring resistance.

A study of gene flow and adaptation in the American oyster (*Crassostrea virginica*) was conducted along the Atlantic and Gulf coasts of the United States (65). The study compared variation at six anonymous scnDNA markers to variation at six allozyme loci and the mitochondrial DNA among nine coastal collections extending from Massachusetts to Louisiana. The anonymous, single copy DNA marker and the mitochondrial markers consistently indicated barriers to gene flow among oysters distributed in the Gulf of Mexico as compared with oysters collected along the Atlantic coast. In fact, a large shift in frequencies occurred in populations close to the tip of Florida. In contrast, none of the allozyme markers differed in frequency among the populations. Thus, despite evident barriers to gene flow indicated by the anonymous scnDNA and mitochondrial markers, allozymes were maintained in equal frequencies among collections. The most parsimonious explanation for these observations is that while populations were genetically differentiated owing to an apparent barrier to gene flow, allozyme frequencies were maintained at equal frequencies among populations by some form of balancing selection (but see 91).

Population genomics is especially powerful for addressing hypotheses regarding individual nucleotide substitutions. The frequency of fast and slow allozymes at the alcohol dehydrogenase locus (*Adh*) in *D. melanogaster* follows a latitudinal cline on three continents. The distribution of 113 *Adh* alleles containing 44 segregating sites in 1533 flies was examined from 25 collections throughout the North-South cline (19, 68). These studies measured variation at each of the 44 segregating sites and found a consistent level among 42. However, variation at two sites was excessive. One of these sites corresponded to the codon encoding the amino acid differences between the fast and slow allozymes and the other corresponded to an upstream insertion/deletion site. Further testing among fast and slow alleles showed that slow alleles were maintained at similar frequencies within all populations, probably by continuous gene flow. However, polymorphic sites within fast alleles varied greatly in frequency among populations. These results suggest that, although gene flow maintains genetic homogeneity among *Drosophila* populations, positive selection along the North-South cline acts specifically on nucleotides associated with the fast allele.

An ultimate understanding of phenotypic evolution will require characterization of the molecular genetics of the trait(s) of interest. Mackay, Langley, and their colleagues have focused efforts on understanding the molecular genetic basis of the quantitative traits abdominal and sternopleural bristle number in *D. melanogaster* (70, 80, 82, 87). Bristles are neurosensory organs, and there are many genes in *Drosophila* spp. known to affect neural development (26); such genes are a priori candidates for quantitative variation in bristle number. In addition, there are many genes that have been identified on the basis of major mutations affecting bristle number (78). Both classes of genes often map to regions containing quantitative trait loci for bristle number (81, 83, 100), and in a number of cases, bristle quantitative trait loci have been shown to either complement or be dominant modifiers of candidate genes (81, 86).

Relationships between molecular and phenotypic variation at three chromosome regions containing the candidate genes *Achaete-Scute* (80, 87), *Scabrous* (70, 82), and *Delta* (79) have now been assessed, and, in every case, naturally occurring molecular variants have been found to be significantly associated with phenotypic variation in bristle number. Although it remains to be demonstrated that the molecular variation associated with phenotypic variation occurs at the same loci as the known major mutants, the evidence is the most suggestive yet that loci with known functions in the development of a quantitative trait are also segregating for alleles that contribute to natural quantitative variation in that trait.

CRITICAL ASSUMPTIONS OF POPULATION GENOMICS

The conceptual basis for population genomics depends on four critical assumptions. First and most important, we assume that genes and nucleotide substitutions evolve independently through extensive and continuous recombination. Second, we assume that migration, drift, and inbreeding influence all parts of the genome equally. Third, we assume that an investigator has the means to assay SNPs at many loci rapidly and inexpensively. Fourth, we assume that the investigator has access to information on the relative linkage and/or physical positions of genetic markers. Below, we fully explain each assumption and provide empirical data that highlight the issues associated with these assumptions.

Assumption 1: Genes Evolve Independently

Detection of locus-specific effects critically depends on independent evolution of genes or nucleotide substitutions within genes, achieved through extensive and continuous recombination. If separate mutations are correlated either because they do not recombine or they are involved in creating essential secondary structure in a gene product or if their gene products interact within the same metabolic pathway, then we cannot differentiate forces acting on single genes or segregating sites. For example, suppose we have identified a genomic region conferring adaptation to a

particular host plant. The genomic region contains 10 genes in a linear order, but these genes never recombine. Furthermore, suppose that an allele of gene 6 confers the ability to use an alternative host plant. Under this scenario, we would never be able to determine which of the 10 genes confer host adaptation. The advantageous allele at locus 6 would increase in frequency due to selection, but so would all of the other nine genes, even though they do not influence host adaptation. If all of the polymorphic sites within the *Adh* gene failed to recombine then investigators would not have been able to determine which polymorphisms were associated with clinal variation (19).

Another useful way of thinking about independent evolution among genes or substitutions within genes is that extensive and continuous recombination creates a natural linkage-mapping experiment. In most mapping studies, recombination occurs in a single round of meioses in heterozygous parents and is detected among F_2 offspring. If we extend the family into $F_3, F_4 \dots F_n$ generations, then opportunities arise for additional recombination among individual genes. As the number of generations approaches infinity, recombination eventually occurs between individual nucleotides. In principle, this should allow one to associate a particular phenotype not only with a particular gene but also with a particular segregating site within that gene.

A large number of studies in *D. melanogaster* have now validated this assumption of independence among loci (2, 13, 19, 29, 39, 61, 71, 74, 76, 82, 95, 101, 108, 117, 132). In many of these studies, the authors plotted the linkage disequilibrium coefficient (D) for all possible pairwise combinations of segregating sites as a function of physical distance between sites in a gene. The linkage disequilibrium coefficient measures the degree to which a pair of nucleotide substitutions at two polymorphic sites is observed together in individuals. The observed frequency is compared with the expected frequency based on the independent frequencies of those substitutions in a population. In general, investigators in the above studies found independent segregation of substitutions, even those <50 nucleotides apart, which suggests that recombination is sufficiently common to maintain independent evolution among polymorphic sites, even at the level of substitutions within a gene.

Several circumstances may, however, preclude independent evolution of loci. First, recombination rates are not uniformly distributed across chromosomes and, in general, are much lower among genes located within or near centromeres and telomeres than among genes in other chromosome regions (59). This observation alone would not affect the assumption of independent evolution. However, a number of studies in *Drosophila* spp. have demonstrated that DNA variation is positively correlated with the amount of recombination (17, 20, 75). This correlation was formally evaluated in terms of three alternative models (57). (a) A selective-sweep model considers the outcome of a gene with a large positive impact on fitness that is closely linked to one or more genes that are neutral or nearly neutral in their effects on fitness. The model predicts that, if a locus is subjected to rapid positive selection, then proximal neutral genes will also be carried along

(will be “swept” or will “hitchhike”) to fixation if selection is rapid enough that recombination cannot independently assort the proximal genes. For example, in the evolution of insecticide resistance, we would expect that genes linked to an insecticide resistance allele would have lower genetic variability than genes that assort independently of the resistance gene. (b) Background selection is a corollary model in which a gene with a negative impact on fitness is closely linked to one or many neutral genes. As the locus under selection is rapidly swept to extinction, so are the linked neutral genes. We would expect the genes linked to an insecticide susceptibility allele to have lower genetic variability after extinction of the original susceptibility alleles. A neutral model was also explored to explain this observation. The study concluded that, in general, patterns of nucleotide substitutions in regions of low recombination are best supported by the selective-sweep model and the background selection model to a lesser degree (24, 57, 124).

An assumption of independent evolution may also be invalid among genes located within chromosomal inversions, because inversions eliminate or severely suppress recombination (78, 146). Inversions can be identified in some insects by cytogenetic analysis of polytene chromosomes in the ovarian nurse cells of partially gravid females or in the salivary glands of 4th-instar larvae. Polytene chromosomes arise as a consequence of endomitosis [chromosome replication without nuclear or cytoplasmic fission (146)]. After endomitosis, somatic pairing of homologous chromosomes [seen only in Collembola and Diptera (146)] results in thick, rosy structures with distinct “bands and puffs” that correspond to different densities and conformations of euchromatin and heterochromatin. Somatic pairing of chromosomes in insects bearing an inversion are seen to form loop structures as homologous regions align.

Recombination among loci within inversions is reduced or eliminated during meiosis. A possible advantage of inversions may lie in their ability to maintain favorable (epistatic) groups of alleles at loci contained in the inversion (34, 37, 140). A number of recent studies have investigated the distribution of nucleotide diversity within and between chromosomal arrangements at loci closely linked to the inversion breakpoints in *Drosophila* spp. (9, 10, 144). These studies arrived at different conclusions regarding the level of exchange between arrangements, suggesting that the rate may vary by genome location, the age of the inversion, and the genes located in proximity to the breakpoints. However, these studies suggest that genetic exchange between genes contained in inversions is suppressed at inversion breakpoints relative to other genomic regions. In general these studies found an absence of shared polymorphism between arrangements, a low diversity of alleles at the junction region, and little evidence for genetic exchange between karyotypic classes.

Finally, an assumption of independent evolution may be invalid within the many insect groups for which recombination rates are slow. Sexual reproduction occurs only once annually in many aphid species, and some species or strains of aphids have lost the ability to sexually reproduce (i.e. are anholocyclic). In coccids and cecidomyiids, sexual reproduction may be facultative (146). In haplodiploid species

and in *D. melanogaster* and many other dipteran species, meiosis occurs in only one sex. In these cases, the degree of independent evolution among genes may be reduced (22). In parthenogenetic species, loci will not evolve independently within a matriarchal lineage because no recombination occurs. In these cases, population genomic analysis can be accomplished only among different matriarchal lineages, and, even among these, we would expect limited recombination.

Assumption 2: Genome-Wide Effects Influence All Parts of the Genome Equally

The extent to which locus-specific effects can be accurately detected depends greatly on the assumption that migration, drift, and inbreeding homogeneously influence variation throughout a genome. Genome-wide processes should affect all loci in a similar way on average, but it should be remembered that the “on average” is just that, in the statistical sense. Although genome-wide processes affect all loci similarly, every population statistic is subject to sampling error. Even loci subject to identical evolutionary processes will vary in their statistical properties because of the stochastic nature of insect populations affected by environmental and demographic variability (14, 99, 109, 110).

A priori, we would expect that (a) genetic variation should be uniformly distributed among populations with high gene flow, (b) genetic drift will cause populations to differentiate when they are maintained or established by one or a few individuals, and (c) inbreeding will cause individuals to become homozygous at the majority of loci throughout their genomes. Although evolutionary processes can in theory be unambiguously classified as either genome-wide or locus-specific, in practice the two are not always easy to disentangle.

Consider the fate of a new allele arising in a population. A novel neutral mutation might enter a population with frequency $1/2N_e$, where N_e is the effective population size (66). Remember that many new alleles enter a population through migration and mutation in every generation. The rate with which neutral mutations become fixed in a population is proportional to $4N_e s$, where s is the magnitude of selection acting on the new allele (66). If the allele is neutral then s equals 0, but alleles with an s that is small relative to N_e are defined as nearly neutral, and their ultimate fate is determined largely by genetic drift. A general expectation is that genetic load due to many alleles with small negative- s values can progressively increase in populations with reduced N_e . In contrast, alleles with an equivalent small s but that occur in much larger populations may be under almost exclusive control of selection because drift will have no effect. The implication is that the influence of s (a locus-specific effect) is not independent of N_e , which is proportional to genetic drift (a genome-wide effect). Some alleles may act as if they are neutral (i.e. not under selection) when they occur or have occurred in small populations (67, 84).

Also consider that, when a new mutation arises in a population, it is in complete disequilibrium with all other segregating sites on the same chromosome. The extent

to which disequilibrium is disrupted is a function of meiosis, generation length, and the physical locations of markers. Any factors that reduce recombination will confound locus-specific and genome-wide effects. Again, locus-specific effects (selective sweeps and background selection) acting on the new allele and associated loci are confounded by genome-wide processes (52).

It is well known that homozygotes become common under inbreeding and drift. Consequently, selection against recessive deleterious mutations also becomes more common (36, 85, 130) as does fixation of slightly deleterious alleles. In these cases, drift and inbreeding may generate locus-specific effects. Numerous studies have documented the large genetic loads that insect genomes carry, including recessive alleles that are lethal or deleterious when homozygous (30, 84). Suppose that lethal or deleterious alleles at several loci reach a high frequency in a new population by genetic drift. A period of rapid purifying selection would occur in the population as individuals homozygous for these recessive alleles died. Proximal genomic regions would be subject to strong background selection, and heterozygosity in that genomic region would be greatly reduced.

Similarly, if an allele is uncommon (<5%) in a natural population, then homozygotes for that allele will be very rare (<0.25%). Suppose again that uncommon alleles at one or several loci reach high frequency in a new population by genetic drift. Homozygotes may have an unusually high fitness in the new population, and the uncommon allele will become fixed through positive selection, causing associated alleles at linked loci to hitchhike to high frequency. The implication for population genomics is that many loci and proximal genomic regions with high loads of deleterious and lethal alleles may be much more susceptible to reduced heterozygosity during inbreeding or population bottlenecks. In practice this could be difficult to distinguish from a general reduction in heterozygosity caused by increased drift. In fact, the interaction between the purifying selection of deleterious alleles and directional selection on favored alleles under conditions that promote inbreeding is thought to form the genetic basis of mating-system evolution (12).

Another important caveat to an assumption of genome-wide effects concerns epistatic interactions among alleles at different loci. In the shifting-balance model, Wright (149) argued that rapid adaptive evolution may result from beneficial combinations of epistatically acting alleles that drift to high frequency in small, subdivided populations. This occurs because normally rare or uncommon alleles may suddenly reach high frequencies, which in turn gives rise to new combinations of multilocus genotypes in individual organisms that encode a novel array of phenotypes. Thus, although we would expect genetic variability to decrease when a population is founded by few individuals, genetic variance caused by epistasis may actually increase genotypic and phenotypic variance after a severe population bottleneck (43, 139). In support of this scenario, Bryant & Meffert (25) and Meffert & Bryant (92, 93) detected increased additive genetic variance after a severe population bottleneck in the common housefly *Musca domestica*. Similar results were achieved in a laboratory simulation of the shifting-balance model among populations of the flour beetle *Tribolium castaneum* (139). The

implications of these results for population genomics are that novel epistatic interactions may cause correlated (nonindependent) shifts among multiple parts of a genome and that proximal genomic regions could be subject to selective sweeps and background selection. Although this has yet to be documented at the level of individual nucleotides, epistasis would be manifested as significant disequilibrium among substitutions at different loci.

Assumption 3: Analysis of Single-Nucleotide Polymorphisms at Many Loci

A wide variety of techniques is available for identification of SNPs. These include restriction fragment length polymorphism analysis, single-strand conformation polymorphism (SSCP) (106), heteroduplex analysis (145), denaturing gradient gel electrophoresis (97), and allele-specific oligonucleotide hybridization (116). In our experience, SSCP analysis is the least expensive, fastest, and most reproducible and sensitive of all of these techniques (23, 53).

Population genomics studies have generally analyzed variation at scnDNA (65, 107) or cDNA loci. The advantage of cDNA markers is that the investigator can often infer the structure and function of that gene by performing a BLAST search on the national databank. Insect cDNA sequences are becoming increasingly common in genetic databases. With the publication of the *Drosophila* genome sequence (1), insect geneticists now have access to a large amount of information on individual genes. We routinely use the BLASTX algorithm available at the National Center for Bioinformatics (NCBI) website to translate our cDNA sequences and compare these to sequences in the *Drosophila* genome. In our work with the mosquitoes *Aedes aegypti*, *Aedes albopictus*, and *Aedes triseriatus*, most cDNA sequences (60%) match with a *Drosophila* gene of known function; however, for the remainder of sequences, BLASTX will not find a gene of similar function or will find a match to a *Drosophila* gene of unknown function.

Assumption 4: Relative Locations of Genetic Markers

The last assumption is unnecessary when examining segregating sites in candidate genes (70, 79, 80, 87). In this case, all that is needed is sequence information along the gene of interest.

For genome-wide analyses, information on the physical distances (in base pairs) among genes can be used, but this information has been determined in only the few arthropods with completed or ongoing genome projects. Otherwise, a formal linkage analysis is required. Modern statistical algorithms that analyze recombination frequencies simultaneously among many loci generally require information on the segregation of alleles in a defined pedigree in which only the four alleles arising from each of the P_1 parents segregate throughout the cross. The four most common pedigrees used for this purpose are an F_1 intercross, a recombinant inbred line, a back cross to a recurrent parent, or a backcross to a parent from an inbred line.

A series of necessary biological conditions must apply to perform any of these four crossing designs. First, genotypes of the P_1 and F_1 parents must be known, which generally requires single-pair matings among P_1 parents. Many insect species mate only in swarms or under unknown field conditions, but, if the genotype of one P_1 (usually the female) is known, it will be possible to infer the genotypes of the other P_1 parent from a collection of F_1 progeny. For the same reasons, it may be necessary to infer F_1 parental genotypes by examining the genotypes of F_2 (intercross or back-cross) progeny. Back-cross designs depend on the ability to mate an F_1 individual to its actual (recurrent) P_1 parent or to any (nonrecurrent) P_1 individual from an inbred line. In most insect species, females store the sperm in the spermathecae so that a recurrent back cross is possible only between an F_1 daughter and her P_1 father (e.g. 55). Assuming that these conditions have been met, the genotypes in the P_1 and F_1 parents and F_2 offspring can be determined. The number of loci is limited only by the amount of DNA that can be extracted from an individual insect. As genotypes are determined at each genetic marker, they are entered into a locus \times F_2 or BC offspring matrix. MAPMAKER and JOINMAP are two programs that utilize this format to (a) identify loci arising on a common linkage group, (b) estimate pairwise recombination frequencies in centimorgans among these loci, and (c) determine the optimal linear order of markers along a linkage group (73, 123). These results provide sufficient information on the relative linkage relationships of markers for a population genomics project.

INITIATING POPULATION GENOMIC STUDIES

Data Collection

Most of the *Drosophila* population genomics studies cited above analyzed variations in isochromosomal lines extracted from natural populations by using balancer chromosomes. However, balancer chromosomes are available for very few other insect species (15). We are discovering that this is not a major impediment. Once we have designed a primer pair and determined the conditions that will reliably amplify all or part of a gene in individual insects, we examine the SSCP patterns in several insects from different, often geographically distinct populations. In very few cases do we find no variation. Alternatively, in a few loci, there is so much variation that no two individuals show the same SSCP genotype. In general, we select loci with an intermediate amount of variability that exhibit a finite number of distinct SSCP genotype patterns. The PCR products from each SSCP genotype are sequenced directly. DNA sequences are then aligned (63) to identify segregating sites.

Because in most cases we examine PCR products from diploid, potentially heterozygous individuals, we carefully inspect the DNA ladder (for manual sequencing) or the chromatographic scan (for automated sequencing) for mixed (heterozygous) sites (in haplodiploid species, it is possible to identify individual

alleles in haploid males, and these precautions are not necessary). Heterozygotes are manifested as bands of equal intensity in one ladder position or as overlapping (or nearly overlapping) peaks in the chromatograph. We score each heterozygous position as R = A and G, Y = C and T, W = A and T, K = G and T, M = A and C, or S = G and C. Sequences from several different individuals with the same SSCP genotype are obtained to confirm the sensitivity and reproducibility of our technique. We then survey, by SSCP, the genotypes of all individuals sampled in a population. The sequences associated with an SSCP pattern are then entered into an individual \times nucleotide position data matrix. A label at the beginning of each row identifies individuals and the collection to which they belong.

Most statistical methods and computer programs (114, 120, 126) use this data structure. However, many methods do not accept R, Y, W, K, M, or S as codes for heterozygotes because they assume complete knowledge of the sequence of an allele or haplotype. Note that, with the SSCP approach, the actual phase of nucleotides at segregating sites are unknown. We have attempted cloning of PCR products from individuals with distinct genotypes to recover the sequences of individual alleles. However, this effort has generally failed because of the appearance of sporadic substitutions during the cloning process. These substitutions are frequently manifested as singletons. We know that most of these are errors because they do not appear in our sequences obtained directly from PCR products, and we find >2 alleles per individual.

Data Analysis

The same statistics traditionally used in population-genetic studies to analyze one or several loci are appropriate in the population genomics approach. Derivation of the sampling distribution of population-genetic statistics requires nothing more than plotting the frequency histograms of estimates from each locus or segregating site within a gene. We also recommend that each statistic be plotted in relation to the physical or linkage position to determine whether extreme values cluster in specific regions in the genome.

It is also important that there have been recent advances in statistical approaches based on inferences and data present in gene genealogies (16, 21, 40, 64, 13, 134, 136, 147). However, these approaches are beyond the scope of this review.

Tests of Recombination and Independence To test our first assumption, it is important to test for evidence of independent evolution among loci or among nucleotide sites. Linkage disequilibrium is frequently analyzed between pairs of variable sites. A χ^2 test is used to detect significant linkage disequilibrium, and a Bonferroni procedure can be used to correct for multiple tests (143). In addition, the sign test on the linkage disequilibrium coefficient D (77) can be applied to search for overall evidence of linkage disequilibrium in the complete data set. Recombination events between polymorphic sites are usually analyzed by the four-gamete test (59). Finally, it is common to plot D as a function of physical

distance between all pairwise combinations of polymorphic sites. A Mantel test can be used to test the correlation between recombination and physical distance. While the correlation is expected to be weak, extreme values of D may indicate sites that interact epistatically. If evidence of epistasis is found, these substitutions should no longer be treated as independent. The authors have written a FORTRAN program (LGENOME) that analyzes linkage disequilibrium at pairs of segregating sites in a gene. The program does not assume knowledge of allele sequences but rather analyzes disequilibrium among genotypes within individuals and among populations using the methods of Ohta (102, 103, 143).

Nucleotide Polymorphisms

As indicated above, patterns of nucleotide variation within and between alleles can leave traces of selection. Five different statistics are commonly used to estimate and report nucleotide polymorphism (44). These are (a) the number of segregating sites (S), (b) the average number of pairwise nucleotide differences (k), (c) the nucleotide diversity (π) or average number of nucleotide differences per site (98), and (d) the heterozygosity per site (μ) expected under the neutral model at mutation-drift equilibrium (141). Frequently the pairwise nucleotide difference distribution or mismatch distribution is also analyzed. According to the neutral model with no recombination, this distribution is Poisson-like in growing populations, but it follows a geometric distribution in constant-size populations (122). The shape of the mismatch distribution can be characterized by the raggedness (r) statistic (48, 49). This statistic measures the smoothness of the observed mismatch distribution, which is smaller in expanding than in constant-size populations. The computer program DnaSP (114) estimates nucleotide polymorphism, linkage disequilibrium, recombination, the raggedness statistic, and genetic differentiation between alleles. The spatial distribution of substitutions along a coding sequence can also be examined (128). This test compares the difference between the observed cumulative distribution of distances between substitutions and a theoretical, homogeneous distribution. Critical values of the test statistic for significance tests are derived by Monte Carlo simulations of the null model, and clumping of substitutions within a gene may indicate selection acting in this region.

Population Structure A number of approaches have been used to compare levels of genetic diversity among sequences within and between populations. The level of genetic differentiation is commonly estimated as the average number of nucleotide substitutions per site (d_{xy}) (98). Genetic differentiation between alleles can be compared by using the permutation test proposed previously (58, 62). The hypergeometric distribution is used to test whether the number of detected, shared silent polymorphisms (sites segregating for the same two nucleotides) can be explained by parallel mutations or whether the mutations were inherited from a common ancestor (113, 115). The authors have written a FORTRAN program (GENOMEST) that calculates F_{is} and F_{st} (as well as Weir and Cockerham's f and θ) at each segregating site in a gene.

Detecting Selection

Neutrality tests are performed to determine whether the observed data conform to the predictions of the neutral model of molecular evolution. Tajima's test (127), based on intraspecific data, compares the observed polymorphism frequency spectrum with that expected under neutrality. Negative values of Tajima's D statistic indicate an excess of polymorphisms segregating at low frequency in the data set. The frequency distributions of silent and replacement polymorphisms in populations can be used to detect effects of weak selection (5, 6, 8). This is done by cataloging the silent substitutions that change a codon from a preferred to an unpreferred one or vice versa. Codons are then classified into preferred and unpreferred codons according to methods described previously (3, 7). Next, the frequency distributions of preferred, unpreferred, and replacement substitutions are determined (4) and compared by Mann-Whitney U tests. There are two different variants of this test: the fdMWU test (8), in which only polymorphisms are included in calculating the frequency distribution of the different mutational classes, and the fddMWU test, which includes fixed differences.

The Fu & Li tests compare independent estimates of μ assuming neutrality (41). Their D statistic is based on the number of mutations in the internal and external branches of the gene genealogy, whereas the F statistic compares the average number of pairwise differences (k) and the number of mutations in external branches of the genealogy. An outgroup species is needed to estimate the number of mutations in external branches. In the Fu & Li test without an outgroup (D^* and F^* statistics), the numbers of mutations in the external branches are inferred from the numbers of singletons or polymorphic variants that are present only once in the sample. Negative values of the different Fu & Li statistics (41) indicate an excess of unique polymorphisms in the data set. A drawback of Tajima's (127) and Fu & Li's tests (41) is that they make the unrealistic assumption of no recombination between sites.

The Hudson-Kreitman-Aguade test (60) requires data from two genomic regions on the intraspecific variation in at least one species and on the amount of interspecific divergence. This test determines whether the level of polymorphism and the level of divergence are proportional in both regions as expected from neutral predictions. The test assumes free recombination between both genomic regions and no recombination between sites of the same region; however, the test is conservative when these assumptions do not hold. The permutation test for genetic differentiation (58, 62) can be performed with the PERMTEST program provided by the authors.

McDonald's statistical test (89) contrasts the heterogeneity in the distribution of polymorphism and divergence across a DNA region. Putative heterogeneity is analyzed by the number of runs detected in the sample, where a run is defined as a set of one or more polymorphic (or fixed) sites preceded and followed by at least one fixed (or polymorphic) site. Positive and balancing selections cause the number of runs detected in the sample to be smaller than that expected under neutrality, which is tested by Monte Carlo simulations. The McDonald & Kreitman (90)

test determines whether the ratio of nonsynonymous to synonymous polymorphisms within species is the same as the ratio of nonsynonymous to synonymous substitutions between species, as expected from neutral predictions.

The F_s test statistic (41) is based on the probability of having no fewer haplotypes or alleles than those observed in the data set. An excess of rare alleles relative to the number expected under neutral predictions is reflected by large negative values of the F_s statistics. Critical values of the observed F_s statistics can be obtained by Monte Carlo simulations. Phylogenetic analysis to reconstruct the gene genealogy of the studied lines can be performed with MEGA or PAUP programs (69, 126).

Sample Size and Design Are Critical in Population Genomics

Most of the tests described above critically depend upon an assumption of neutral evolution among polymorphic characters. However, nucleotide polymorphisms in surveyed loci frequently vary significantly among different populations and therefore invalidate tests of neutrality that assume a mutation-drift equilibrium. Recent papers demonstrate that the ways in which individuals and populations are sampled can severely invalidate assumptions of neutral evolution.

Schmid et al (119) sought to determine whether variation in amino acids sequences observed in three rapidly evolving nuclear genes between *Drosophila melanogaster* and *D. simulans* also exists within populations of these species. They did this by comparing sequences of the three genes sampled from flies collected across the whole geographic distribution of the species. However, they sequenced only small numbers of alleles from a single population or from populations in proximity to one another. They demonstrated that this sample design does not allow an analysis of the geographic population structure of species or identification of different patterns of selection in local populations. Alleles or substitutions detected in small samples from diverse geographic locations do not indicate the relative abundance of these alleles in local populations. This is of concern because population-specific selective sweeps for certain loci were detected in a study of microsatellite variation in separate populations across the world (118). Detailed population genomic analyses of populations of *D. melanogaster* and *D. simulans* have revealed that both species exhibit a considerable amount of local population structure (18, 46, 47). In a study of the *Gld* locus in *D. melanogaster*, the ratio of replacement to silent substitutions was significantly elevated in a Chinese collection (90), but not in two samples from Africa or a third sample from North America (45). These studies point to the need for careful analysis of sequence polymorphisms in local populations or a nested hierarchical approach rather than continued analyses among diverse populations or among species.

The tests of neutral evolution listed above are useful for detecting strong positive selection, but they do not reject the null hypothesis of neutral evolution if selection coefficients are small. Power analyses showed that Tajima's D and Fu & Li's D fail to detect an ancient or very recent selective sweep and that their power is

low with small sample sizes (121). Similar results probably apply to most of the tests listed above. Furthermore, weak and episodic selection can provide patterns of nucleotide substitutions that are indistinguishable from neutral evolution (42). The existence of weak selection and the problems associated with detecting it are now widely acknowledged (4, 104, 105, 142).

CONCLUSION

We have shown how the population genomics paradigm can and is being used to identify adaptive variation in genes or among segregating sites within genes. We echo the findings of a comprehensive review of molecular evolution of *D. melanogaster* by Akashi & Kreitman (7), who concluded that a mosaic of patterns was present in the fly genome. No single population genetic model explained all of the results, and, although most results could be explained by either neutral evolution or purifying selection, some nucleotide variation was consistent with adaptation. Unfortunately, adaptive variation could be equally explained by directional selection for adaptive alleles, purifying selection against deleterious alleles, or balancing selection for different alleles in different environments. Akashi & Kreitman (7) stressed that patterns of variation can be understood only if examined in reference to the form of selection acting on the individual loci under consideration, and they also expressed the need for a priori models based on gene function and the natural history of the organisms. In fact, many of the tests for balancing and directional selection outlined above critically depend on knowledge of phenotypic, environmental, or seasonal factors that are the source of selection. The study described earlier that identified the precise nucleotide substitutions in the *D. melanogaster Adh* gene associated with latitudinal clines (19) illustrates this point perfectly. The segregating sites correlated with latitudinal clines were only identified by partitioning the variation in substitution frequencies with respect to the latitude at which flies were collected. If the variation was not partitioned or was partitioned according to other effects, then F_{st} was uniformly distributed across all segregating sites and no sign of selection was evident.

In this regard, population genomics provides a novel and necessary interface between population genetics and molecular biology. A priori hypotheses concerning the adaptive significance of a segregating site require knowledge of the basic biochemistry and physiology of the gene product. Population geneticists can study variation in individual genes (38) and identify genes and regions of genes that are apparently subject to selection. However without an understanding of the molecular structure, function, and interaction of amino acids and a clear picture of how a gene influences the metabolism or development of an organism, the adaptive significance of a segregating site will remain obscure. Conversely, although the molecular biologist may understand in detail the structure, function, and physiology of a gene product, only insights gained through population genomics can

indicate whether variation found in a gene is subject to natural selection, which parts of a gene or protein are subject to selection, and the mode of selection acting on that gene. Ultimately, this interface will yield a means for predicting mechanisms of molecular adaptation. In turn this will lead to more rational design of insecticides or transgenic host plant immunity, thereby reducing the chances of adaptation to future pest control strategies.

ACKNOWLEDGMENTS

We thank Mike Lynch, Jeff Leips, and colleagues for comments. This work was supported by the MacArthur Foundation for the Network on the Biology of Parasite Vectors and by National Institutes of Health grants AI 41436, AI 45430, ESO5949, Colorado Hatch Project #697, and USDA–NRI grant 99-35302-8020.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–95
- Aguade M, Miyashita N, Langley CH. 1989. Restriction-map variation at the zeste-*tko* region in natural populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* 6:123–30
- Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* 139:1067–76
- Akashi H. 1996. Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* 144:1297–307
- Akashi H. 1997. Codon bias evolution in *Drosophila*: population genetics of mutation-selection drift. *Gene* 205:269–78
- Akashi H. 1997. Distinguishing the effects of mutational biases and natural selection on DNA sequence variation. *Genetics* 147:1989–91
- Akashi H, Kreitman M. 1995. Molecular evidence for natural selection. *Annu. Rev. Ecol. Syst.* 26:403–22
- Akashi H, Schaeffer SW. 1997. Natural selection and the frequency distributions of “silent” DNA polymorphism in *Drosophila*. *Genetics* 146:295–307
- Andolfatto P, Kreitman M. 2000. Molecular variation at the In(2L)t proximal breakpoint site in natural populations of *Drosophila melanogaster* and *D. simulans*. *Genetics* 154:1681–91
- Andolfatto P, Wall JD, Kreitman M. 1999. Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. *Genetics* 153:1297–311
- Antolin MF. 1993. Genetics of biased sex ratios in subdivided populations: models, assumptions, and evidence. *Oxf. Surv. Evol. Biol.* 9:239–81
- Antolin MF. 1999. A genetic perspective on mating systems and sex ratios of parasitoid wasps. *Res. Popul. Ecol.* 41:29–37
- Aquadro CF, Jennings RM Jr, Bland MM, Laurie CC, Langley CH. 1992. Patterns of naturally occurring restriction map variation, dopa decarboxylase activity variation and linkage disequilibrium in the *Ddc* gene region of *Drosophila melanogaster*. *Genetics* 132:443–52
- Baer CF. 1999. Among-locus variation in

- Fst: fish, allozymes and the Lewontin-Krakauer test revisited. *Genetics* 152:653–59
15. Beeman RW, Stuart JJ, Haas MS, Friesen KS. 1996. Chromosome extraction and revision of linkage group 2 in *Tribolium castaneum*. *J. Hered.* 87:224–32
 16. Beerli P, Felsenstein J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–73
 17. Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356:519–20
 18. Begun DJ, Aquadro CF. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* 365:548–50
 19. Berry A, Kreitman M. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. *Genetics* 134:869–93
 20. Berry AJ, Ajioka JW, Kreitman M. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* 129:1111–17
 21. Bertorelle G, Excoffier L. 1998. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* 15:1298–311
 22. Beye M, Hunt GJ, Page RE, Fondrk MK, Grohmann L, et al. 1999. Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics* 153:1701–8
 23. Black WC, DuTeau NM. 1996. RAPD-PCR and SSCP analysis for insect population genetic studies. In *The Molecular Biology of Insect Disease Vectors, A Methods Manual*, ed. JM Crampton, CB Beard, C Louis, pp. 514–31. New York: Chapman & Hall
 24. Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* 140:783–96
 25. Bryant EH, Meffert LM. 1995. An analysis of selectional response in relation to a population bottleneck. *Evolution* 49:626–34
 26. Campos-Ortega JA. 1993. Early neurogenesis in *Drosophila melanogaster*. In *The Development of Drosophila melanogaster*, ed. M Bate, A Martinez Arias, Vol. 1, pp. 1091–129. Plainview, NY.: Cold Spring Harbor Lab. Press
 27. Caterino MS, Cho S, Sperling FAH. 2000. The current state of insect molecular systematics: a thriving tower of Babel. *Annu. Rev. Entomol.* 45:1–54
 28. Charlesworth B, Hughes KA. 1999. The maintenance of genetic variation in life history traits. In *Evolutionary Genetics from Molecules to Morphology*, ed. R. Singh, CB Krimbas, pp. 369–92. Cambridge, UK: Cambridge Univ. Press
 29. Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* 70:155–74
 30. Charlesworth D, Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* 18:237–68
 31. Choe JC, Crespi B. 1997. *The Evolution of Mating Systems in Insects and Arachnids*. Cambridge, UK: Cambridge Univ. Press
 32. Clark AG, Aguade M, Prout T, Harshman LG, Langley CH. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* 139:189–201
 33. Clark AG, Wang L. 1997. Molecular population genetics of *Drosophila* immune system genes. *Genetics* 147:713–24
 34. Coluzzi M. 1982. Spatial distribution of chromosomal inversions and speciation in Anopheline mosquitoes. *Prog. Clin. Biol. Res.* 96:143–53
 35. Cook JM, Crozier RH. 1995. Sex

- determination and population biology in the Hymenoptera. *Trends Ecol. Evol.* 10:281–86
36. Crow JF. 1993. Mutation, mean fitness, and genetic load. *Oxf. Surv. Evol. Biol.* 9:3–42
 37. Dobzhansky TG. 1970. *Genetics of the Evolutionary Process*. New York: Columbia Univ. Press
 38. Eanes WF. 1999. Analysis of selection on enzyme polymorphisms. *Annu. Rev. Ecol. Syst.* 30:301–26
 39. Eanes WF, Ajioka JW, Hey J, Wesley C. 1989. Restriction-map variation associated with the G6PD polymorphism in natural populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* 6:384–97
 40. Excoffier L, Smouse PE. 1994. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics* 136:343–59
 41. Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709
 42. Gillespie JH. 1994. Alternatives to the neutral theory. In *Non-neutral Evolution*, ed. B Golding, pp. 1–17. New York: Chapman & Hall
 43. Goodnight CJ. 1988. Epistasis and the effect of founder events on the additive genetic variance. *Evolution* 42:441–54
 44. Graur D, Li W-H. 1999. *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer. 2nd ed.
 45. Hamblin MT, Aquadro CF. 1997. Contrasting patterns of nucleotide sequence variation at the glucose dehydrogenase (Gld) locus in different populations of *Drosophila melanogaster*. *Genetics* 145:1053–62
 46. Hamblin MT, Aquadro CF. 1999. DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. *Genetics* 153:859–69
 47. Hamblin MT, Veuille M. 1999. Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. *Genetics* 153:305–17
 48. Harpending HC. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* 66:591–600
 49. Harpending HC, Sherry ST, Rogers AR, Stoneking M. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.* 34:483–96
 50. Heckel DG. 1993. Comparative genetic linkage mapping in insects. *Annu. Rev. Entomol.* 38:381–408
 51. Heimpel GE, Antolin MF, Strand MR. 1999. Diversity of sex-determining alleles in *Bracon hebetor*. *Heredity* 82:282–91
 52. Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* 8:269–94
 53. Hiss RH, Norris DE, Dietrich CH, Whitcomb RF, West DF, et al. 1994. Molecular taxonomy using single strand conformation polymorphism (SSCP) analysis of mitochondrial ribosomal DNA genes. *Insect Mol. Biol.* 3:171–82
 54. Holloway AK, Heimpel GE, Strand MR, Antolin MF. 1999. Survival of diploid males in *Bracon sp. near hebetor* (Hymenoptera: Braconidae). *Ann. Entomol. Soc. Am.* 92:110–16
 55. Holloway AK, Strand MR, Black WC, Antolin MF. 2000. Linkage analysis of sex determination in *Bracon sp. near hebetor* (Hymenoptera: Braconidae). *Genetics* 154:205–12
 56. Houle D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* 130:195–204
 57. Hudson RR. 1994. How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? *Proc. Natl. Acad. Sci. USA* 91:6815–18
 58. Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9:138–51
 59. Hudson RR, Kaplan NL. 1985. Statistical

- properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–64
60. Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–59
61. Hudson RR, Saez AG, Ayala FJ. 1997. DNA variation at the Sod locus of *Drosophila melanogaster*: an unfolding story of natural selection. *Proc. Natl. Acad. Sci. USA* 94:7725–29
62. Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132: 583–89
63. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23:403–5
64. Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, et al. 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* 64:817–31
65. Karl SA, Avise JC. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* 256:100–2
66. Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge Univ. Press
67. Kimura M, Ohta T. 1971. *Theoretical Aspects of Population Genetics*. Princeton, NJ: Princeton Univ. Press
68. Kreitman M, Hudson RR. 1991. Inferring the evolutionary histories of the Adh and Adh-dup loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* 127:565–82
69. Kumar S, Tamura K, Nei M. 1994. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput. Appl. Biosci.* 10:189–91
70. Lai C, Lyman RF, Long AD, Langley CH, Mackay TF. 1994. Naturally occurring variation in bristle number and DNA polymorphisms at the scabrous locus of *Drosophila melanogaster*. *Science* 266:1697–702
71. Laitinen T, Kauppi P, Ignatius J, Ruotsalainen T, Daly MJ, et al. 1997. Genetic control of serum IgE levels and asthma: linkage and linkage disequilibrium studies in an isolated population. *Hum. Mol. Genet.* 6:2069–76
72. Lander ES, Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–99
73. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, et al. 1987. MAP-MAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174–81
74. Langley CH, Aquadro CF. 1987. Restriction-map variation in natural populations of *Drosophila melanogaster*: white-locus region. *Mol. Biol. Evol.* 4:651–63
75. Langley CH, MacDonald J, Miyashita N, Aguade M. 1993. Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* 90:1800–3
76. Langley CH, Tobari YN, Kojima KI. 1974. Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* 78:921–36
77. Lewontin RC, Kojima K. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution* 14:458–72
78. Lindsley DL, Zimm GG. 1992. *The Genome of Drosophila melanogaster*. San Diego, CA: Academic
79. Long AD, Lyman RF, Langley CH, Mackay TF. 1998. Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* 149:999–1017

80. Long AD, Lyman RF, Morgan AH, Langley CH, Mackay TF. 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* 154:1255–69
81. Long AD, Mullaney SL, Mackay TF, 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* 144:1497–510
82. Lyman RF, Lai C, MacKay TF. 1999. Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genet. Res.* 74:303–11
83. 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* 149:983–98
84. Lynch M, Blanchard J, Houle D, Kibota T, Schultz S, et al. 1999. Perspective: spontaneous deleterious mutation. *Evolution* 53:645–63
85. Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer
86. Mackay TF, Fry JD. 1996. Polygenic mutation in *Drosophila melanogaster*: genetic interactions between selection lines and candidate quantitative trait loci. *Genetics* 144:671–88
87. Mackay TF, Langley CH. 1990. Molecular and phenotypic variation in the achaete-scute region of *Drosophila melanogaster*. *Nature* 348:64–66
88. Mallet J, Joron M. 1999. Evolution of diversity in warning coloration and mimicry: polymorphisms, shifting balance, and speciation. *Annu. Rev. Ecol. Syst.* 30:201–33
89. McDonald JH. 1996. Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol.* 13:253–60
90. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–54
91. McDonald JH, Verrelli BC, Geyer LB. 1996. Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster, *Crassostrea virginica*. *Mol. Biol. Evol.* 13:1114–18
92. Meffert LM, Bryant EH. 1991. Mating propensity and courtship behavior in serially bottlenecked lines of the housefly. *Evolution* 45:293–306
93. Meffert LM, Bryant EH. 1992. Divergent ambulatory and grooming behavior in serially bottlenecked lines of the housefly. *Evolution* 46:1399–407
94. Mitton JB. 1993. Theory and data pertinent to the relationship between heterozygosity and fitness. In *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives*, ed. NW Thornhill, pp. 17–41. Chicago, IL: Univ. Chicago Press
95. Miyashita NT, Aguade M, Langley CH. 1993. Linkage disequilibrium in the white locus region of *Drosophila melanogaster*. *Genet. Res.* 62:101–9
96. Mopper S, Strauss SY. 1998. *Genetic Structure and Local Adaptation in Natural Insect Populations: Effects of Ecology, Life History, and Behavior*. New York: Chapman & Hall
97. Myers RM, Maniatis T, Lerman LS. 1987. Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* 155:501–27
98. Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia Univ. Press
99. Nei M, Maruyama T. 1975. Letters to the editors: Lewontin-Krakauer test for neutral genes. *Genetics* 80:395

100. 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* 153:1317–31
101. Ober C, Cox NJ, Abney M, Di Rienzo A, Lander ES, et al. 1998. Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma. *Hum. Mol. Genet.* 7:1393–98
102. Ohta T. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc. Natl. Acad. Sci. USA* 79:1940–44
103. Ohta T. 1982. Linkage disequilibrium with the island model. *Genetics* 101:139–55
104. Ohta T. 1996. The current significance and standing of neutral and neutral theories. *BioEssays* 18:673–77
105. Ohta T, Gillespie JH. 1996. Development of neutral and nearly neutral theories. *Theor. Popul. Biol.* 49:128–42
106. Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86:2766–70
107. Orti G, Hare MP, Avise JC. 1997. Detection and isolation of nuclear haplotypes by PCR-SSCP. *Mol. Ecol.* 6:575–80
108. Riley MA, Hallas ME, Lewontin RC. 1989. Distinguishing the forces controlling genetic variation at the Xdh locus in *Drosophila pseudoobscura*. *Genetics* 123:359–69
109. Robertson A. 1975. Gene frequency distributions as a test of selective neutrality. *Genetics* 81:775–85
110. Robertson A. 1975. Letters to the editors: remarks on the Lewontin-Krakauer test. *Genetics* 80:396
111. Roderick GK. 1996. Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annu. Rev. Entomol.* 41:325–52
112. Roff DA. 1997. *Evolutionary Quantitative Genetics*. New York: Chapman & Hall
113. Rozas J, Aguade M. 1994. Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 91:11517–21
114. Rozas J, Rozas R. 1999. *DnaSP* version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–75
115. Rozas J, Segarra C, Ribo G, Aguade M. 1999. Molecular population genetics of the rp49 gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics* 151:189–202
116. Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. 1986. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324:163–66
117. Schaeffer SW, Aquadro CF, Anderson WW. 1987. Restriction-map variation in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* 4:254–65
118. Schlotterer C, Vogl C, Tautz D. 1997. Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* 146:309–20
119. Schmid KJ, Nigro L, Aquadro CF, Tautz D. 1999. Large number of replacement polymorphisms in rapidly evolving genes of *Drosophila*: implications for genome-wide surveys of DNA polymorphism. *Genetics* 153:1717–29
120. Schneider S, Roessli D, Excoffier L. 2000. ARLEQUIN ver 2.000, A software for population genetics data analysis. Geneva, Switzerland: Univ. Geneva

- Genet. Biometry Lab., Dep. Anthropol. Ecol. 2.0 ed.
121. Simonsen KL, Churchill GA, Aquadro CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141:413–29
 122. Slatkin M, Hudson RR. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–62
 123. Stam P, Van Ooijen JW. 1996. Joinmap version 2.0: software for the calculation of genetic linkage maps. Wageningen, Netherlands: Centre for Plant Breeding and Reproduction Research
 124. Stephan W, Xing L, Kirby DA, Braverman JM. 1998. A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proc. Natl. Acad. Sci. USA* 95:5649–54
 125. Stouthamer R, Luck RF, Werren JH. 1992. Genetics of sex determination and the improvement of biological control using parasitoids. *Environ. Entomol.* 21:427–35
 126. Swofford DL. 1998. *Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Sunderland, MA: Sinauer
 127. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–95
 128. Tang H, Lewontin RC. 1999. Locating regions of differential variability in DNA and protein sequences. *Genetics* 153:485–95
 129. Taylor MF, Shen Y, Kreitman ME. 1995. A population genetic test of selection at the molecular level. *Science* 270:1497–99
 130. Thornhill NW. 1993. *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives*. Chicago, IL: Univ. Chicago Press
 131. Travis J. 1989. The role of optimizing selection in natural populations. *Annu. Rev. Ecol. Syst.* 20:279–96
 132. Tsaour SC, Ting CT, Wu CI. 1998. Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of *Drosophila*. II. Divergence versus polymorphism. *Mol. Biol. Evol.* 15:1040–46
 133. Tufto J, Engen S, Hindar K. 1996. Inferring patterns of migration from gene frequencies under equilibrium conditions. *Genetics* 144:1911–21
 134. Tufto J, Raybould AF, Hindar K, Engen S. 1998. Analysis of genetic structure and dispersal patterns in a population of sea beet. *Genetics* 149:1975–85
 135. Turelli M, Clark AG, Spofford JB. 1984. Linkage data supporting a mathematical explanation for some empirical cis-trans effects. *Heredity* 52:145–47
 136. Turner TF, Trexler JC, Harris JL, Haynes JL. 2000. Nested cladistic analysis indicates population fragmentation shapes genetic diversity in a freshwater mussel. *Genetics* 154:777–85
 137. Van Ooijen JW, Maliepaard C. 1996. MapQTL version 3.0: software for the calculation of QTL position on genetic maps. Wageningen, Netherlands: Centre for Plant Breeding and Reproduction Research
 138. Vieira C, Pasyukova EG, Zeng ZB, Hackett JB, Lyman RF, et al. 2000. Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. *Genetics* 154:213–27
 139. Wade MJ, Goodnight CJ. 1991. Wright's shifting balance theory: an experimental study. *Science* 253:1015–18
 140. Wasserman M. 1968. Recombination-induced chromosomal heterosis. *Genetics* 58:125–39
 141. Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7:256–76
 142. Wayne ML, Simonsen KL. 1998. Statistical tests of neutrality in the age of weak selection. *Trends Ecol. Evol.* 13:236–40
 143. Weir BS. 1979. Inferences about linkage disequilibrium. *Biometrics* 35:235–54
 144. Wesley CS, Eanes WF. 1994. Isolation

- and analysis of the breakpoint sequences of chromosome inversion In(3L)Payne in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 91:3132–36
145. White MB, Carvalho M, Derse D, O'Brien SJ, Dean M. 1992. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 12:301–6
146. White MJD. 1972. *Animal Cytology and Evolution*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.
147. Wiehe T, Mountain J, Parham P, Slatkin M. 2000. Distinguishing recombination and intragenic gene conversion by linkage disequilibrium patterns. *Genet. Res.* 75:61–73
148. Wright S. 1951. The genetic structure of populations. *Ann. Eugenics* 15:323–54
149. Wright S. 1982. The shifting balance theory and macroevolution. *Annu. Rev. Genet.* 16:1–19