

Inferring the major genomic mode of dominance and overdominance

Hong-Wen Deng¹, Yun-Xin Fu² & Michael Lynch³

¹ORC, Dept. of Medicine, Creighton University 601 N. 30th ST., Suite 6787, Omaha, NE 68131, USA (Phone: 402-280-5911; Fax: 402-280-5034; E-mail: deng@creighton.edu); ²Human Genetics Center, SPH, University of Texas P.O. Box 20334, Houston, TX 77225, USA; ³Dept. of Biology, University of Oregon, Eugene, OR 97403, USA

Key words: dominance and overdominance, heterosis, hypothesis testing, neutral variation, recombination rate

Abstract

The mode of within-locus gene action in most genomic regions is termed as the major genomic mode, i.e., it is the within-locus allelic effects in most regions of the genome. Determining whether dominance or overdominance is the major genomic mode is important for two long-standing evolutionary genetics issues: 1. How is the genetic variation in most genomic regions maintained? 2. What is the major mechanism for heterosis? Many efforts have been made, but almost all of them suffer some explanatory difficulties. Here we propose an alternative inference approach. It is based on the existent theoretical results on the correlation of the recombination rate and the level of neutral variation in different genomic regions. Positive and negative correlations suggest dominance and overdominance, respectively, as the major genomic mode. Zero correlations imply either few selected sites or about equal composition and distribution of dominant and overdominant regions in the genome, depending on the data distribution. This approach not only avoids all the problems associated with earlier approaches, but it is also particularly useful in organisms where controlled breeding is difficult. Well-corroborated data in *Drosophila* and recently emerging data in mice and humans all suggest dominance as the major genomic mode.

Introduction

Inbreeding depression in outcrossing populations results from mating among relatives, and outbreeding enhancement between inbreeding (or isolated) populations results from mating among usually-inbreeding lines (or small populations). Both phenomena are widely observed (e.g., Wright, 1977; Charlesworth & Charlesworth, 1987; Falconer, 1989; Crow, 1993; Lynch & Walsh, 1997). For the sake of simplicity, they will be collectively referred to as heterosis hereafter. The magnitude of heterosis has many implications, such as the evolution of self-incompatibility systems in monoecious plants (Lande & Schemske, 1985; Schemske & Lande, 1985; Charlesworth & Charlesworth, 1987), the evolution of dispersal mechanisms for inbreeding avoidance in animals (Shields, 1982), the biological conservation of rare and endangered species (Soule, 1986), the improvement of agri-

cultural production (Falconer, 1989), and the protection of human welfare (Cavalli-Sforza & Bodmer, 1971).

Two rival mechanistic genetic hypotheses concerning individual loci have been used to explain heterosis. The first is the dominance hypothesis (Davenport, 1908; Crow, 1952), which argues that heterosis is caused by an enhanced expression of deleterious genes due to an increase in homozygosity with the heterozygote performance being intermediate (but not exactly) between the two corresponding homozygotes. The second is the overdominance hypothesis (East, 1908; Shull, 1908; Crow, 1952), which argues for heterozygote superiority relative to homozygotes. Dominance and/or overdominance may not be necessary for heterosis (Richey, 1942; Minvielle, 1987; Schnell & Cockerham, 1992). For example, additive-by-additive gene interaction may contribute to inbreeding depression in populations at gametic phase disequilibrium and

contribute to outbreeding enhancement between populations even under gametic phase equilibrium within each population (Lynch & Walsh, 1997). However, dominance and/or overdominance are usually the major mechanisms of heterosis (Schnell & Cockerham, 1992).

Discriminating between dominance and overdominance is important for discerning different mechanisms for the maintenance of genetic variability (Crow, 1993), as dominance is compatible with mutation-selection balance and overdominance essentially encompasses all kinds of balancing selection at the allelic level. While most experimental data are consistent with the dominance hypothesis, overdominance cannot be ruled out in many situations (Simmons & Crow, 1977; Charlesworth & Charlesworth, 1987; Barrett & Charlesworth, 1991; Stuber et al., 1992; Mitton, 1993; Crow, 1993). In particular, it is difficult to distinguish dominance and overdominance as the within-locus allelic effects for most genomic regions, i.e., the major genomic mode. Although the evidence for functional overdominance does not seem to be very convincing and most cited examples are compatible with associative overdominance, an artifact of linked deleterious recessive genes (c.f., Houle, 1989, 1994; Crow, 1993), the debate over the relative importance of the two genetic effects continues (Sprague, 1983; Wallace, 1989; Mitton, 1993; Crow, 1993; Houle, Morikawa & Lynch, 1996; Charlesworth & Hughes, 1997) and is unlikely to be settled until an inference of an appropriate method is made.

The purpose of this paper is to propose a new method to infer the mode of the prevailing within-locus gene action, dominance or overdominance, in most regions in the genome. The inference of this method bears directly on two long-standing evolutionary and quantitative genetic issues: 1) How is the genetic variation in most genomic regions maintained in diverse natural environments and during the long-term evolutionary history of populations? 2) What is the major genetic cause of heterosis? The theoretical foundations upon which we base our investigative approach have been well laid-out by previous researchers in molecular population genetics and molecular evolution. The data we rely upon are those well corroborated in *Drosophila*, and those recently emerging in mice and humans. What we try to accomplish here is to link the well-known theoretical and empirical results to the long-standing issues in evolutionary and quantitative genetics. We conclude that dominance is the major genomic mode in *Drosophila*, mice, and humans; polymor-

phism in most genomic regions has been shaped under directional rather than balancing selection; and that dominance is likely the major cause for heterosis. We start by pointing out some of the potential problems with previous approaches to discriminating dominance and overdominance, as this serves as a necessary comparison for the proposed method of inference.

Problems

First, dominance and overdominance are defined by within-locus genetic effects. For a locus with alleles **A** and **a**, let the three genotypic values be:

AA	Aa	aa
1	1-hs	1-s

Then $h < 0.0$ implies overdominance, $h = 0.5$ additivity, $0 \leq h \leq 1.0$ ($h \leq 0.5$) dominance and $h > 1.0$ implies underdominance. Note that, throughout, we use 'dominant' or 'dominance' to refer to the cases with $0 \leq h \leq 1.0$ ($h \neq 0.5$), which includes cases of complete dominant or recessive, and partial dominant or recessive.

However, since the definition of a locus itself is not at all clear, the definition of dominance and overdominance based on it may be problematic. This may be easily demonstrated by the following hypothetical example. For two adjacent regions *A* and *B* of a DNA sequence, let region *A* have two type sequences *A* and *a*, and region *B* have two type sequences *B* and *b*. Let genotypic values for regions *A* and *B* be, respectively: $AA = 1.0$, $Aa = 1 - hs$, $aa = 1 - s$; $BB = 1.0$, $Bb = 1 - hs$, $bb = 1 - s$, and $0 < h < 1.0$. If we regard both regions as different loci, then the genetic effects are dominant. However, if a locus is defined to include both regions, there are four different alleles: $|_B^A$, $|_b^A$, $|_B^a$, $|_b^a$. For alleles $|_b^A$ and $|_B^a$, we will have overdominance. This is because the relevant genotypic values (under multiplicative fitness function) are respectively $|_b^A ||_b^A = 1 - s$, $|_B^a ||_B^a = 1 - s$, $|_b^A ||_B^a = 1 - 2hs + h^2s^2$, and we always have $|_b^A ||_B^a > |_B^a ||_b^A$ and $|_b^A ||_B^a > |_B^a ||_B^a$ for $0 < s < 1.0$. This relationship is sometimes referred to as associative overdominance (Houle, 1989, 1994). Thus, the appearance of dominance and overdominance critically depends on what segment of the genome is under study. As an extreme case, if we define the whole genome as a 'super locus', then overdominance will almost always be the case (as reflected

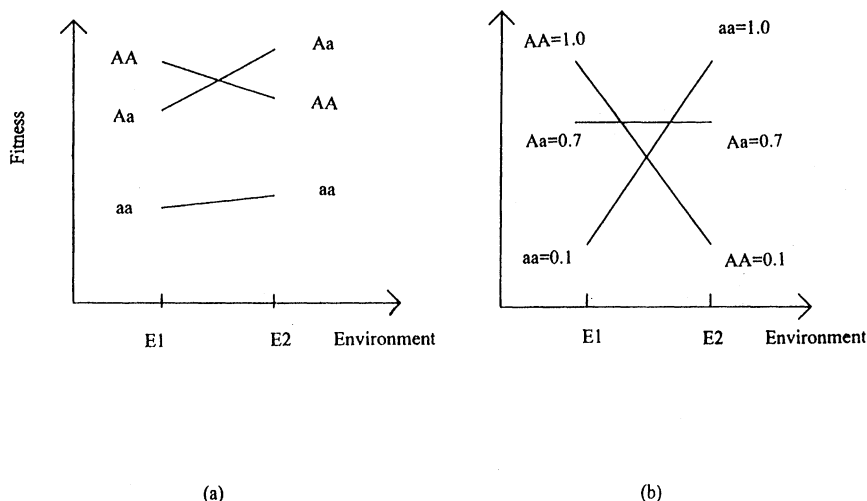


Figure 1. Due to the potential $G \times E$ interaction, a) dominance and overdominance revealed may change from one environment (E1) to another (E2); b) overdominance may result when averaging over environments even if dominance is revealed for both E1 and E2 (marginal overdominance). In b, if fitness is multiplicative across environments E1 and E2 and E1 and E2 are equally-frequently experienced, the overall fitness of Aa is 0.49, while those of AA and aa are 0.1.

by heterosis), even if dominance exists in all individual genomic regions.

Second, distinguishing dominance and overdominance has a bearing on the maintenance of genetic variability, a fundamental issue in evolutionary genetics. Ideally, dominance and overdominance should be studied for fitness *per se*, instead of its components. However, fitness cannot be measured easily (Lewontin, 1974). Studying dominance and overdominance for fitness components and inferring mechanisms for the maintenance of genetic variability could be misleading. This can also be illustrated simply. Assume fitness = viability*fecundity, where locus A underlies total viability with alleles A and a , and locus B controls total fecundity with alleles B and b . As before, let genotypic values for locus A and B be respectively: $AA = 1.0$, $Aa = 1 - hs$, $aa = 1 - s$; $BB = 1.0$, $Bb = 1 - hs$, $bb = 1 - s$, and $0 < h < 1.0$. Given heterosis, studying viability or fecundity, respectively, will reveal dominance. However, for the same reasons demonstrated in the previous paragraph, if studying fitness (i.e., both loci simultaneously), overdominance will sometimes be revealed. This happens when alleles A and b are linked together, and so are alleles a and B .

Third, almost all approaches to inferring dominance and overdominance involve assays of inbreds and outbreds in controlled laboratory conditions, rarely in natural environments. Even if assays are conducted in natural environments, they are in a limited num-

ber of particular natural environments. However, the natural environments that organisms currently experience and those where selection and evolution occurred over the long-term evolutionary history of species are diverse and complex. Due to the common phenomenon of genotype-by-environment interaction (Stearns, 1992; Lynch & Walsh, 1997), dominance revealed in one environment could very well turn out to be overdominance in another (Figure 1). Alternatively, dominance could exist in all individual environments, but overdominance would be revealed when averaging over the environments (marginal overdominance) (Figure 1). Some data on the marker associated analyses of dominance and overdominance bear on this potential problem (Fu & Ritland, 1996; P. Gaffiney pers. comm.). Therefore, inferences on dominance and overdominance from a particular experiment may have little bearing on the long-term issues concerned with balancing selection.

New approach

The existence of these potentially serious problems with almost all previous approaches calls for an alternative and new inference approach that is immune from these problems.

As in most molecular evolutionary studies, we assume that, throughout evolutionary history, mutation

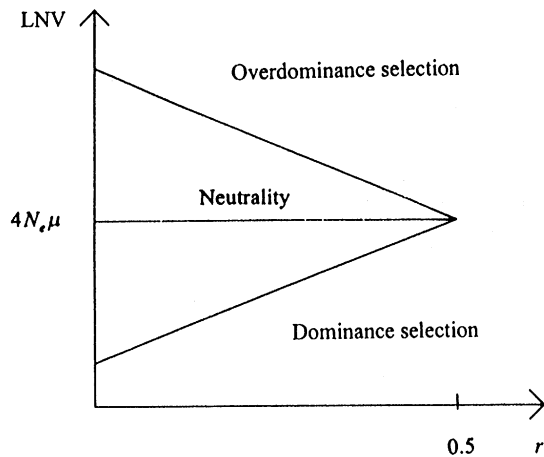


Figure 2. Distinct relationships of the level of neutral variation (LNV) and recombination rate (r) under dominance, overdominance being the major mode of the within-locus genetic effects, and under neutrality. The three lines intersect in regions of free recombination ($r = 0.5$), where the LNV per nucleotide site is $4N_e\mu$.

rates (μ) are more or less constant across neutral sites such as introns, noncoding regions, pseudogenes, and the third positions of codons. The constancy of μ across neutral sites may be checked by comparing molecular polymorphism within species with molecular diversity across species (Hudson, 1990). We also assume that selected sites, if any, are distributed across the whole genome. This is partially supported by the observation that even nongenic genomic DNA may perform essential functions (Zukerkandl, 1976). Then, as well-laid out by the previous theoretical results, distinct patterns exist between the level of neutral variation and the recombination rate across genomic regions, depending on dominance and overdominance. Level of neutral variation can be obtained directly from DNA sequence data or indirectly from RFLP data or from microsatellite loci, and it can be measured by different indices such as the nucleotide diversity (Nei, 1987). All the principles below have been well worked out and are well known; they are listed because they form the theoretical basis for our proposed inference approach.

1. If there are no selected sites in the genome, the level of neutral variation and recombination rate should be uncorrelated (Figure 2). At any time a genomic region in a population is sampled, the average amount of the level of neutral variation is unaffected by recombination (Hudson, 1983; Tajima, 1990). Level of neutral variation is determined by the effective population size (N_e) and μ . On a chromosome, all the neutral regions (or sites) have the

same N_e , and the level of neutral variation is a flat line along regions of different recombination rates (Figure 2). Data from regions on different chromosomes, such as autosomes and sex chromosomes, can be combined for analysis if their respective N_e 's are adjusted to be comparable as in Begun and Aquadro (1992).

If directly selected sites exist, neutral sites linked to them are likely to be under selection indirectly due to linkage. At a site (or locus) of interest (whether it is directly or indirectly under selection), if the heterozygote fitness is between or higher than the two corresponding homozygotes, we will refer to it as dominance or overdominance selection, respectively. By definition, dominance selection includes both selective sweeps (hitchhiking, Maynard Smith & Haigh, 1974; Thomson, 1977; Kaplan, Hudson & Langley, 1989; Stephan, Wiehe & Lenz, 1992; Wiehe & Stephan, 1993) and background selection (Charlesworth, Morgan & Charlesworth, 1993; Charlesworth, Charlesworth & Morgan 1995; Hudson & Kaplan, 1994, 1996). Given the widely observed phenomena of heterosis, dominance selection generally implies dominance ($0 \leq h \leq 1.0$, $h \neq 0.5$) and excludes additivity ($h = 0.5$) as the within locus gene action for most of the loci underlying the traits exhibiting heterosis.

2. Under dominance selection, whether it is a selective sweep due to advantageous mutations or background selection due to slightly deleterious mutations, in chromosomal regions of low recombination rate, the level of neutral variation is low, and in chromosomal regions of high recombination rate, the level of neutral variation is high. This is because, under selective sweeps, the variation at a neutral site will be reduced due to the relatively quick fixation of advantageous mutations at a linked site, pulling the neutral variant linked with it to fixation and eliminating unlinked variants (Maynard Smith & Haigh, 1974; Thomson, 1977; Kaplan, Hudson & Langley, 1989; Stephan, Wiehe & Lenz, 1992; Wiehe & Stephan, 1993). Under background selection, the effective population size is reduced for neutral sites embedded in a region or close to a site subject to continuous deleterious mutation and selection (Charlesworth, Morgan & Charlesworth, 1993; Charlesworth, Charlesworth & Morgan, 1995; Hudson & Kaplan, 1994, 1995). the effects of both kinds of selection decrease with an increasing recombination rate. Therefore, the level of neutral variation and recombination rate

are expected to be positively correlated under dominance selection (Figure 2).

3. For neutral sites closely linked to a site under overdominance selection, the coalescent time can be much longer, resulting in a detectably higher level of neutral variation than neutral expectation (Hudson & Kaplan, 1988; Kaplan, Darden & Hudson, 1988; Hudson, 1990). In this case, the level of neutral variation is higher with a smaller recombination rate (Hudson & Kaplan, 1988; Kaplan, Darden & Hudson, 1988). Therefore, a negative relationship between the level of neutral variation and recombination rate is expected under overdominance selection (Figure 2).

The exact relationships between the level of neutral variation and the recombination rate under the two selection scenarios may not be simple straight lines as depicted in Figure 2. The exact relationship in reality depends on the strength of selection, among other things etc; In regions of free recombination, selection is unlikely to affect the level of neutral variation; thus, the lines under selective and neutral scenarios are expected to intersect at one point (Figure 2).

The distinct relationships between the level of neutral variation and the recombination rate under neutral, dominant, and overdominant selection provide a logical basis for distinguishing dominance and overdominance for fitness as the major mode of gene action at polymorphic sites. This inference approach avoids all the potential problems mentioned earlier. The level of neutral variation in different genomic regions is shaped by different evolutionary forces on the *total fitness over all the natural environments* that populations experience during their long-term evolutionary history. Thus, the conclusions obtained with this approach should be more directly relevant to inferring the mechanisms maintaining the standing genetic variability in most genomic regions. Additionally, by studying the level of neutral variation, the dependence of the apparent features of dominance and overdominance on the definition of a locus is avoided. A locus could be a nucleotide site or a nonrecombining genomic region that does not need to be explicitly identified.

Dominance is the major genomic mode in *Drosophila* and possibly in mice and humans

In *Drosophila*, inference for the major genomic mode of dominance and overdominance has been made by estimating the coefficient of the mean within-locus nonadditive genetic effects \bar{h} (Crow, 1993). Because \bar{h} estimated is almost always greater than zero, overdominance (where $h < 0$) is excluded. However, this approach is valid only when *either* dominance *or* overdominance exists (Deng, 1998). With more a plausible mixture of dominance and overdominance at different loci in the genome, this approach is misleading and always favors a biased interpretation of dominance (Deng, 1998). In molecular population genetics, it has already been demonstrated that overdominance may exist in the *Adh* region in *Drosophila* (Oakeshott et al., 1982; Kreitman & Aguade, 1986; Hudson, Kreitman & Aguade, 1987). Recently, Houle, Morikawa and Lynch (1996) concluded that data on the mutational variabilities are consistent with mutation-selection balance, which is compatible with the dominance hypothesis. However, their analyses are implicitly based on the assumption that the populations' standing genetic variability is due *entirely* to mutation-selection balance. For example, mutation-selection balance must be assumed in order to approximate the mean persistence time of mutations by the ratio of the standing genetic variability to the mutational variability. The mean persistence time of mutations is then, in return, employed by them to support the mutation-selection balance hypothesis. Therefore, there is a problem of logical circulation there.

A remarkable pattern has emerged from molecular population genetic studies in recent years. In *Drosophila*, there is a significantly positive relationship between the level of neutral variation and the recombination rate in the whole genome (Begun & Aquadro, 1992), and on the third and x chromosomes (Aquadro, Begun & Kindahl, 1994; Stephan, 1994; Aguade & Langley, 1994). More recent data summarized by Hudson and Kaplan (1995) and Charlesworth, Charlesworth and Morgan (1995) also confirm or are consistent with the above pattern. Because the data have already been admirably analyzed and summarized, the details are not presented here. Interested readers are referred to the above-cited references for the data.

Two mechanisms are invoked to explain the observed data: selective sweeps (Maynard Smith & Haigh, 1974; Thomson, 1977; Kaplan, Hudson & Langley, 1989; Stephan, Wiehe & Lenz, 1992;

Wiehe & Stephan, 1993) and background selection (Charlesworth, Morgan & Charlesworth, 1993; Charlesworth, Charlesworth & Morgan 1995; Hudson & Kaplan, 1994, 1995). Researchers in the field have been trying to discern the relative importance of the selective sweep and background selection (e.g., Aquadro, Begun & Kindahl, 1994). However, the precise relevance of the data to the prominent and longstanding question concerning dominance and overdominance has been largely ignored. As pointed out earlier, both selective sweeps and background selection are sources of dominance selection, since in neither case does the heterozygote have higher fitness than the corresponding homozygotes. The data cannot be explained if overdominance is the major genomic mode. We therefore conclude that dominance is indeed the major genomic mode in *Drosophila*. This does not exclude the possibility that overdominance exists in some genomic regions. However, it does exclude overdominance as the general mode of gene action in most genomic regions in *Drosophila*.

Data recently emerging from other organisms conform to the same pattern. For example, the level of neutral variation is positively correlated with the recombination rate for four X-linked loci in mice *Mus domesticus* (Nachman, submitted to Genetics), and for seven X-linked loci in humans (Nachman et al., pers. comm.). More efforts may be needed to test these patterns in mice, humans, and other organisms to see if they are as general and robust as in *Drosophila*. However, the currently available data do suggest that dominance may be the major genomic mode in mice and humans.

Robustness of the inference approach

The discussion of the maintenance of genetic variability under dominance (mutation-selection balance) and overdominance (balancing selection) is generally based on a one-locus-two-allele model. It has been shown (Mandel, 1959; Li, 1967; Crow & Kimura, 1970) that with multiple alleles, even if some heterozygotes are inferior to some homozygotes, stable genetic variability can be maintained without continuous mutation to supplement the less-fit alleles. On the other hand, it is not sufficient for all heterozygotes to be superior to maintain stable genetic variability. However, for within-population DNA sequence data, except at some mutation hot spots such as those in the human mitochondrial control region (Vigilant et al., 1991; Tamura & Nei, 1993), multiple hits seldom

occur and very rarely does any site have more than two nucleotides. With DNA sequence data, a locus can be as small as a nucleotide site. This reduces the complexity of discussing maintenance of genetic variability introduced by the problem of multiple alleles.

When a mixture of dominance and overdominance exists at different genomic regions, the sign of the correlation coefficient between the level of neutral variation and recombination rate obtained by regression analysis should reflect the prevailing mode. This is the standard 'averaging effect' of regression analysis. When one mode dominates, the 'outliers' in the regression analysis may signal genomic regions with the other genetic mode. This may then serve as a basis for further genetic analysis concentrating on these regions. When the correlation coefficient is zero, the distribution of data points may bear some valuable information. Uniform scattering of data points around the flat line (Figure 3) indicates that few sites in the genomic regions examined are under selection. This is because the level of neutral variation is about the same across the genome, except by sampling error. If data points scatter more widely in low than high recombination regions (Figure 3), selected dominant and overdominant sites are common and are likely to be distributed evenly in the genome. This is because the level of neutral variation in low recombination regions would be either higher (under overdominance selection) or lower (under dominance selection) than the neutral expectation, thus having larger variation. In high recombination regions, the level of neutral variation would be about that expected under neutrality regardless of selection, thus having less variation.

For a locus with two alleles, overdominance unconditionally leads to a stable equilibrium in an infinite population, thus maintaining genetic variability. In a finite population, overdominance is a factor retarding fixation if the equilibrium allele frequencies lie between 0.2 - 0.8 (Robertson, 1962). Outside this range, there are some values of $N_e(s_1 + s_2)$, where for which s_1 and s_2 are selection coefficients against both homozygotes, overdominance actually accelerates fixation (Robertson, 1962; Crow & Kimura, 1970, p. 413), thus generating the patterns of neutral variation at linked sites similar to that under dominance selection. Thus, while our inference approach applies nicely to populations of large sizes such as *Drosophila*, in populations of small sizes, the conclusions may occasionally, depending on values of $N_e(s_1 + s_2)$ when equilibrium allele frequencies lie outside 0.2 - 0.8, tend to favor dominance in a biased fashion.

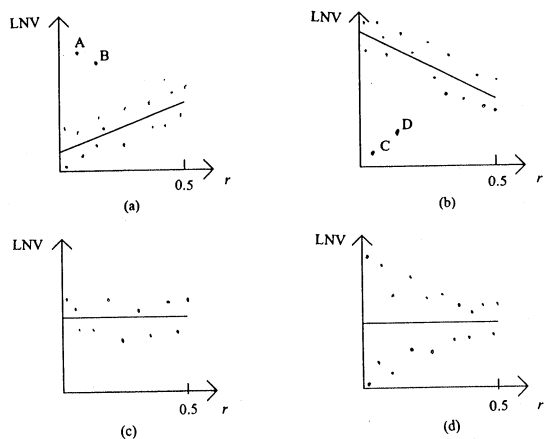


Figure 3. Outliers in the regression analysis may help identify the minor mode of genetic effects: a) outliers (points A and B) indicate possible overdominant genomic regions while most genomic regions are under dominance; b) outliers (points C and D) indicate potential dominant genomic regions while most genomic regions are under overdominance. Zero slope in the regression analysis may indicate: c) few selected sites across genomic regions examined, if points are scattered evenly around the flat line; or d) dominance and overdominance distribute about equally across genomic regions examined, if points scatter more widely in genomic regions of lower r .

Associative overdominance (Houle, 1989) is a perplexing problem in almost all short-term marker associated analyses of dominance and overdominance concerning a chromosomal region. However, for our approach it may at most elevate the level of neutral variation in some particular chromosomal regions of restricted recombination for an evolutionarily short period of time. Over time, the effects of associative overdominance will decay and eventually be eliminated due to recombination. As our conclusion and approach are based on the level of the neutral variation patterns across the genome that are shaped by the long-term evolutionary history of populations, associative overdominance is unlikely to be a problem. In any case, the available data reveal a major genomic mode of dominance in *Drosophila* and possibly in mice and humans, so the associative overdominance is not a concern.

Other genetic processes such as gene conversion (Hilliker et al., 1994) may potentially have some impact on the proposed patterns between the level of neutral variation and recombination rate. However, little is known about the frequency of gene conversion and the effects it may have on the proposed patterns between the level of neutral variation and recombination rate. Clearly, more information is needed. Even if

gene conversion is frequent, there is little reason for it to be responsible for the positive relationship between the level of the neutral variation and recombination rate observed in *Drosophila*. Thus, our inference about dominance as the major genomic mode in *Drosophila*, mice, and humans should be robust.

Discussion

We present a new method of inference approach to the mode of dominance and overdominance based on existing theoretical results and employ well-corroborated data to show that dominance is unambiguously the major genomic mode in *Drosophila*. Recently emerging data from mice and humans also suggest that dominance is the major genomic mode in these organisms. This inference approach is useful for discriminating dominance and overdominance as the major genomic mode in organisms where controlled breeding is difficult. It is directly related to the issues concerned with discriminating the mechanisms responsible for maintaining genetic variability for most of the genomic regions. It is also highly relevant to discriminating: 1) dominance and overdominance as the major cause for heterosis, and 2) mutation-selection balance and balancing selection as the major mechanism for maintaining genetic variability of fitness. It should be noted that the above two issues and inferring the major genomic mode of dominance and overdominance are not entirely the same but highly related questions. Unless there are only a few major genomic regions responsible for the most of genetic variation and heterosis of fitness (which is unlikely), the major genomic mode of dominance and overdominance should be the major mechanism for heterosis and for maintaining genetic variability of fitness.

During decades of studies on testing the neutral theory of molecular evolution (Kimura, 1968, 1983; King & Jukes, 1969), methodologies have been developed and data are being accumulated that may help answer some long-standing questions in evolutionary and quantitative genetics. In principle, neutrality tests using DNA sequence data, such as Tajima's test (Tajima, 1983), Fu and Li's test (Fu & Li, 1993), the HKA test (Hudson, Kreitman & Aquade, 1987), and the McDonald-Kreitman test (McDonald & Kreitman, 1991) may also be employed to discriminate dominance and overdominance concerning a specific genomic region.

Most previous efforts at distinguishing the dominance and overdominance hypotheses depend on controlled breeding, which is not feasible for many organisms, such as humans. However, controlled breeding is not a pre-requisite in the proposed approach here, which can essentially be applied to any organism of interest. Much effort is being spent on assembling physical and genetic maps in several species. Examining whether a correlation exists between the level of neutral variation and recombination rate not only sheds light on the long-standing problems concerning neutrality vs. selection (Begun & Aquadro, 1992), but also provides an excellent opportunity to test the even longer-standing questions concerning the mechanisms for the maintenance of genetic variability and heterosis for fitness. With the rapid advent of new molecular techniques, more data are expected to be available in the near future.

Acknowledgements

We thank Drs. D. Charlesworth, R. Adkins, and D. Stivers and an anonymous reviewer for comments on the manuscript. H.-W. Deng would like to thank Drs. A. Clark, C. Langley, D. Houle, A. Kondrashov, and especially Dr. D. Charlesworth for discussions, and thank Dr. D. Hedgecock for providing a support to attend the conference 'The Genetic and Physiological Bases of Heterosis', which greatly benefited this work. The work was supported by a Health Future Foundation grant from Creighton University to Dr. R. Recker and a FIRST AWARD from NIH to Y.-X. Fu.

References

- Aguade, M. & C. Langley, 1994. Polymorphism and divergence in regions of low recombination in *Drosophila*, pp. 67–76 in Non-neutral Evolution, edited by B. Golding, Chapman & Hall, New York.
- Aquadro, C.F., D.J. Begun & E.C. Kindahl, 1994. Selection, recombination, and DNA polymorphism in *Drosophila*, pp. 47–70 in Non-neutral Evolution, edited by B. Golding, Chapman & Hall, New York.
- Barrett, S.C.H. & D. Charlesworth, 1991. Effects of a change in the level of inbreeding on the genetic load. *Nature* 352: 522–524.
- Begun, D. & C.F. Aquadro, 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356: 519–520.
- Cavalli-Sforza, L. L. & W. F. Bodmer, 1971. *The Genetics of Human Populations*. W.H. Freeman, New York.
- Charlesworth, B., M. T. Morgan & D. Charlesworth, 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134: 1289–1303.
- Charlesworth, D. & B. Charlesworth, 1987. Inbreeding depression and its evolutionary consequences. *Ann. Rev. Ecol. Syst.* 18: 237–268.
- Charlesworth, D., B. Charlesworth & M. T. Morgan, 1995. The pattern of neutral variation under the background selection model. *Genetics* 141: 1618–1632.
- Charlesworth, B. & K. A. Hughes, 1997. The maintenance of genetic variation in life-history traits, in *Evolutionary Genetics from Molecules to Morphology*, edited by R.S. Singh & C.B. Krimbas. Cambridge University Press, Cambridge, UK, in press.
- Crow, J. F., 1952. Dominance and Overdominance, pp. 282–297 in *Heterosis*, edited by J. W. Gowen. Iowa State College Press, Ames, Iowa.
- Crow, J. F., 1993. Mutation, mean fitness, and genetic load, in *Oxford Surveys in Evolutionary Biology*, Vol. 9., Oxford.
- Crow, J. & M. Kimura, 1970. *An Introduction to Population Genetics Theory*. Happer & Row, New York.
- Davenport, C. B., 1908. Degeneration, albinism and inbreeding. *Science* 28: 454–455.
- Deng, H.-W., 1998. Estimating (over)dominance coefficient and discriminating dominance vs. overdominance as the genetic cause of heterosis. *Genetics* 148: in press.
- East, E. M., 1908. Inbreeding in corn. *Rept. Conn. Agr. Exp. Stn.* (1907) 419–428.
- Falconer, D. S., 1989. *Introduction to Quantitative Genetics*. Longman, New York.
- Fu, Y.-B. & K. Ritland 1996. Marker-based inference about epistasis for gene influencing inbreeding depression. *Genetics* 144: 339–348.
- Fu, Y.-X. & W.-H. Li 1993. Statistical test of neutrality of mutations. *Genetics* 133: 693–709.
- Gregory, W. C., 1965. Mutation frequency, magnitude of change and the probability of improvement in adaptation. *Radiation Botany* 5 (Suppl.): 429–441.
- Hilliker, A. J., G. Harauz, A. G. Reaume, M. Gray, S. H. Clark & A. Chovnick, 1994. Meiotic gene conversion tract length distribution within the rosy locus of *Drosophila melanogaster*. *Genetics* 137: 1019–1026.
- Houle, D., 1989. Allozyme-associated heterosis in *Drosophila melanogaster*. *Genetics* 123: 789–801.
- Houle, D., 1994. Adaptive distance and the genetic basis of heterosis. *Evolution* 48: 1410–1417.
- Houle, D., B. Morikawa & M. Lynch, 1996. Comparing mutational variabilities. *Genetics* 143: 1467–1483.
- Hudson, R.R., 1983. Properties of a neutral allele model with intra-genic recombination. *Theor. Popul. Biol.* 23: 1183–201.
- Hudson, R.R., 1990. Gene genealogies and coalescent process. pp 1–44 in *Oxford Surveys in Evolutionary Biology*, edited by D. Futuyma & J. Antonovics. Oxford University Press, Oxford.
- Hudson, R.R. & N.L. Kaplan, 1988. The coalescent process in models with selection and recombination. *Genetics* 120: 831–40.
- Hudson, R.R. & N.L. Kaplan, 1994. Gene trees with background selection. pp. 140–153 in *Non-Neutral Evolution*, edited by B. Golding. Chapman & Hall, New York.
- Hudson, R.R. & N.L. Kaplan, 1995. Deleterious background selection with recombination. *Genetics* 141: 1605–1617.
- Hudson, R.R., M. Kreitman & M. Aquade, 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116: 153–159.
- Kaplan, N.L., T. Darden & R.R. Hudson, 1988. The coalescent process in models with selection. *Genetics* 120: 819–29.

- Kaplan, N.L., R.R. Hudson & C.H. Langley, 1989. The 'hitchhiking effect' revisited. *Genetics* 123: 887–899.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217: 624–626.
- Kimura, M., 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- King, J.L. & T.H. Jukes, 1969. Non-Darwinian evolution. *Science* 164: 788–798.
- Kreitman, M. & M. Aquade, 1986. Excess polymorphism at the ADH locus in *Drosophila melanogaster*. *Genetics* 114: 93–110.
- Lande, R. & D.W. Schemske, 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* 39: 24–40.
- Lewontin, R.C., 1974. *The Genetic Basis of York*.
- Li, C.C., 1967. Genetic equilibrium under selection. *Biometrics* 23: 397–484.
- Lynch, M. & B. Walsh. 1997. *Principles of Evolutionary Quantitative Genetics*. in press. Sinauer, Sunderland, MA.
- Mandel, S.P.H. 1959. The stability of a multiallelic system. *Heredity* 13: 289–302.
- Maynard Smith, J. & J. Haigh, 1974. The hitch-hiking effect of a favorable gene. *Genet. Res.* 231: 1114–1116.
- McDonald, J. & M. Kreitman, 1991. Adaptive protein evolution at Adh locus in *Drosophila*. *Nature* 351: 652–654.
- Minvielle, F., 1987. Dominance is not necessary for heterosis: a two-locus model. *Genetical Research* 49: 245–247.
- Mitton, J.B., 1993. Theory & data pertinent to the relationship between heterozygosity and fitness, pp. 17–41 in *The History of Inbreeding and Outbreeding*, edited by N.D. Thornhill. The University of Chicago Press, Chicago.
- Nachman, M.W., in review for *Genetics*. DNA variability, selection, and recombination at four X-linked loci in *Mus domesticus*.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia Univ. Press. New York.
- Oakeshott, J.G., J.B. Gibson, P.R. Anderson, W.R. Knibb, D.G. Anderson & G.K. Chambers, 1982. Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on three continents. *Evolution* 36: 86–96.
- Richey, F.D., 1942. Mock-dominance and hybrid vigor. *Science* 96: 280–281.
- Robertson, A., 1962. Selection for heterozygotes in small populations. *Genetics* 47: 1291–1230.
- Schemske, D.W. & R. Lande, 1985. The evolution of self-fertilization and inbreeding depression in plants. II Empirical observations. *Evolution* 39: 41–52.
- Schnell, F.W. & C.C. Cockerham, 1992. Multiplicative vs. arbitrary gene action in heterosis. *Genetics* 131: 461–469.
- Shields, W.M., 1982. *Phylotary, Inbreeding, and the Evolution of Sex*. SUNY press, Albany, New York.
- Shull, G.H., 1908. The composition of a field of maize. *Rpt. Am. Breed. Assoc.* 4: 296–301.
- Simmons, M.J. & J.F. Crow, 1977. Mutations affecting fitness in *Drosophila* populations. *Ann. Rev. Genet.* 11: 49–78.
- Soule, M., 1986. *Conservation Biology*. Sinauer Association, Sunderland, Mass.
- Sprague, G.F., 1983. Heterosis in maize: Theory and practice, pp. 47–70 in *Heterosis: Reappraisal of Theory and Practice*, edited by R. Frankel, Springer-Verlag, Berlin.
- Stearns, S.C., 1992. *The Evolution of Life-Histories*. Oxford Univ. Press. New York.
- Stephan, W., 1994. Effects of genetic recombination and population subdivision on nucleotide sequence variation in *Drosophila ananassae*, pp. 57–66 in *Non-neutral Evolution*, edited by B. Golding, Chapman & Hall, New York.
- Stephan, W., T.H.E. Wiehe & M.W. Lenz, 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Popul. Biol.* 41: 1039–1045.
- Stuber, C.W., S.E. Lincoln, D.W. Wolff, T. Helentjaris & E.S. Lander, 1992. Identification of genetic factors in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132: 823–839.
- Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Tajima, F., 1990. Relationship between DNA polymorphism and fixation time. *Genetics* 125: 447–454.
- Tamura, K. & M. Nei, 1993. Estimating the number of nucleotide substitutions in the control region of mitochondrial DNA in human and chimpanzees. *Mol. Biol. Evol.* 10: 512–526.
- Thomson, G., 1977. The effect of a selected locus on a linked neutral locus. *Genetics* 85: 752–778.
- Vigilant L., M. Stoneking, H. Harpending, K. Hawkes, & A. C. Wilson, 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503–1507.
- Wallace, B., 1989. One selectionist's perspective. *Quart. Rev. Biol.* 64: 127–145.
- Wiehe, T.H.E. & W. Stephan, 1993. Analysis of a genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol. Biol. Evol.* 10: 824–854.
- Wright, S. 1977. *Evolution and the Genetics of Populations*. Vol. 3. Experimental results and evolutionary deductions. The University of Chicago Press, Chicago.
- Zukerkandl, E., 1976. Gene control in eukaryotes and the C-value paradox: 'Excess' DNA as an impediment to transcription of coding sequences. *J. Mol. Evol.* 9: 73–104.