CIS-TRANS EFFECTS INDUCED BY LINKAGE DISEQUILIBRIUM

MICHAEL TURELLI

Department of Genetics University of California, Davis, California 95616

Manuscript received January 20, 1982
Revised copy accepted August 26, 1982

ABSTRACT

This note concerns theoretical and experimental studies of multifactorial traits, especially fitness and its components, in which (1) the loci studied are only a subset of those relevant to the character of interest and (2) the genotypes at the loci studied are in nonrandom association (linkage disequilibrium) with genotypes at the loci ignored. In these cases, phenotypic differences between cis and trans double heterozygotes can occur even though no linkage phase effects are inherent in the genetic determination of the trait. Examples are drawn from both theoretical and experimental work, and implications in both areas are discussed.

THEORETICIANS and experimentalists routinely assume that the linkage phase of alleles governing a multifactorial trait is irrelevant to the determination of phenotype. For example, a widespread assumption in two-locus selection theory is that coupling and repulsion double heterozygotes have the same fitness (see KARLIN 1975; EWENS 1979, p. 57). The purpose of this note is to point out that this assumption will usually be violated whenever linkage disequilibrium exists and the number of loci studied is smaller than the total number of loci governing the trait of interest. The magnitude of the discrepancy depends on both the intergenotypic differences in phenotype and the degree of nonrandom association between the loci being studied and those ignored. This observation has implications for theoretical studies of general two-locus systems. Moreover, it can explain some experimentally obtained results and provides a tool for empirical studies.

Consider a multifactorial character, for instance, viability, governed by \( n \) polymorphic loci with \( n > 2 \). Denote by \( i_m \) and \( j_m \) two arbitrarily chosen alleles at locus \( m \). Let \( m(i_1j_2 \ldots i_n/j_1j_2 \ldots j_n) = m(i/j) \) be the mean phenotype (viability) of individuals derived from \( i_1i_2 \ldots i_n \) maternal gametes and \( j_1j_2 \ldots j_n \) paternal gametes, and let \( F_n(i_1j_2 \ldots i_n/j_1j_2 \ldots j_n) = F_n(i/j) \) denote their frequency in a given population. No assumptions are needed concerning the mechanisms producing this genotype frequency distribution. For this complete set of polymorphic controlling loci, assume that there are no maternal effects and no linkage phase effects, i.e., assume that all \( 2^n \) permutations of the \( i \) and \( j \) alleles between the two uniting gametes will produce the same mean phenotype. In particular, this will imply that \( m(i/j) = m(j/i) \) and

\[
m(i_1j_3 \ldots i_n/j_1i_3 \ldots j_n) = m(i/j). \tag{1}
\]

Suppose that one is able to observe, or able to do the algebra concerning, only
the first two of these loci. (The extension to \( k \) loci with \( 2 < k < n \) is trivial.) Interest then centers on the mean phenotypes and associated frequencies of two-locus genotypes such as \( i_1i_2/j_1j_2 \) and \( i_1j_2/j_1i_2 \). The frequency of \( i_1i_2/j_1j_2 \) individuals in the population is

\[
F_1(i_1i_2/j_1j_2) = \sum_{i_3} \ldots \sum_{i_n} \sum_{j_3} \ldots \sum_{j_n} F_n(i_1i_2i_3 \ldots i_n/j_1j_2j_3 \ldots j_n)
\]

(2)

with the summations running over indices corresponding to all alleles at loci 3 through \( n \). The associated mean phenotype is

\[
\bar{m}(i_1i_2/j_1j_2) = \sum_{i_3} \ldots \sum_{i_n} \sum_{j_3} \ldots \sum_{j_n} m(i/j) F_n(i/j) / F_2(i_1i_2/j_1j_2)
\]

(3)

(cf. EWENS and THOMSON 1977).

To see how linkage disequilibrium produces cis-trans effects, it suffices to compare \( \bar{m}(i_1i_2/j_1j_2) \) and \( \bar{m}(i_1j_2/j_1i_2) \), the phenotypes of the cis and trans double heterozygotes averaged over their respective genetic backgrounds. According to equation 3, these are identical if and only if

\[
\sum_{i_3} \ldots \sum_{j_n} [m(i_1i_2i_3 \ldots i_n/j_1j_2j_3 \ldots j_n) F_n(i_1i_2i_3 \ldots i_n/j_1j_2j_3 \ldots j_n) / F_2(i_1i_2/j_1j_2)] - m(i_1j_2i_3 \ldots i_n/j_1i_2j_3 \ldots j_n) F_n(i_1j_2i_3 \ldots i_n/j_1i_2j_3 \ldots j_n) / F_2(i_1j_2/j_1i_2) = 0.
\]

Applying the assumption of no intrinsic linkage phase effects, in particular equation 1, this reduces to

\[
\sum_{i_3} \ldots \sum_{j_n} m(i/j) [F_n(i/j) / F_2(i_1i_2/j_1j_2)] - F_n(i_1j_2i_3 \ldots i_n/j_1i_2j_3 \ldots j_n) / F_2(i_1j_2/j_1i_2) = 0.
\]

(4)

There are two obvious sufficient conditions for equation 4 to hold. The first is that the \( n \)-locus genotype frequency distribution can be factored as the product of the two-locus marginal frequency distribution for the loci under study and the \((n - 2)\)-locus marginal distribution for the “hidden” loci, i.e., that

\[
F_n(i_1 \ldots i_n/j_1 \ldots j_n) = F_2(i_1i_2/j_1j_2) F_{n-2}(i_3 \ldots i_n/j_3 \ldots j_n).
\]

(5)

This means that the two-locus genotypes under study are randomly associated with the \((n - 2)\)-locus genotypes at the hidden loci. When equation 5 holds, the terms on the left hand side of equation 4 enclosed in square brackets all vanish and equation 4 is satisfied. If mating is random and newly formed zygotes are censused, equation 5 is replaced by an equivalent factorization of the \( n \)-locus gamete frequencies. In that case, the factorization extends the two-locus concept of linkage equilibrium to blocks of loci. Whenever equation 5 is violated, the left hand side of equation 4 will consist of both positive and negative terms that will not usually cancel. When they do not, the disequilibrium between the studied loci and the hidden ones will have created a cis-trans effect. Its cause is intuitively clear. When equation 5 does not hold, different genotypes at the loci observed will experience statistically different genetic backgrounds. The cis-trans effect arises from differences between the average effects of these distinct backgrounds. Analogously, apparent maternal effects, e.g., \( \bar{m}(i_1i_2/j_1i_2) \)
ASSOCIATIVE CIS-TRANS EFFECTS

\( \neq \tilde{m}(i_1j_2/i_1i_2) \), can be produced if genotype frequencies violate \( F_n(i/j) = F_n(j/i) \) for some reason.

A second sufficient condition for equation 4 to be satisfied is that

\[ m(i_1 \ldots i_n/j_1 \ldots j_n) = m(i_1i_2/j_1j_2), \]

i.e., that only the two loci under consideration contribute to the phenotype. When equation 6 holds, \( m(i/j) \) can be factored out of equation 4 and the remaining positive and negative terms sum to 1 and -1, respectively. Not surprisingly, the opposite situation in which the loci studied contribute nothing to the phenotype, i.e.,

\[ m(i_1 \ldots i_n/j_1 \ldots j_n) = m(i_3 \ldots i_n/j_3 \ldots j_n), \]

does not ensure that equation 4 is satisfied. Here too, background differences will be reflected in mean phenotypes. In this case, the creation by linkage disequilibrium of a cis-trans effect at inherently "neutral" loci is analogous to the phenomenon of associative overdominance (Frydenberg 1963; Kimura and Ohta 1971). This suggests the name associative cis-trans effects for this phenomenon.

EXAMPLES AND DISCUSSION

To illustrate these calculations, I will first consider the equilibria of a specific three-locus selection model studied by Feldman, Franklin and Thomson (1974). The model describes the dynamics of gamete frequencies at three diallelic loci subject to viability selection, recombination and random mating. Following their notation, the alleles at the three loci will be denoted \( A, a, B, b, \) and \( C, c \). The specific model of interest is their example 5. It assumes recombination rates of 0.0099 between the \( A \) and \( B \) loci, 0.0103 between \( B \) and \( C \) and no interference. Fitnesses are assigned to genotypes solely on the basis of the number of heterozygous loci. Let \( \gamma_i \) denote the fitness of individuals with \( i \) heterozygous loci and \( 3-i \) homozygous loci. Feldman, Franklin and Thomson (1974) assume \( \gamma_i = 0.8^{3-i} \) for \( i = 0, 1, 2, 3 \). This reflects multiplicative fitness effects across loci with identical symmetric overdominance at each locus. Two classes of stable symmetric equilibria were found. Representatives of each are shown in Table 1. Of interest are the induced fitnesses of the coupling and repulsion double heterozygotes at loci \( A \) and \( B \). Formula 3 yields

\[ \tilde{m}(AB/ab) = (\gamma_2[F_3(ABC)F_3(abC) + F_3(ABc)F_3(abc)]) + \gamma_3[F_3(ABC)F_3(abC)]+ F_3(ABc)F_3(abc))/F_2(AB)F_2(ab), \]

and an analogous formula for \( \tilde{m}(Ab/aB) \). Equilibrium I in Table 1 satisfies equation 5 and both double heterozygotes have fitness 0.9. However, for equilibrium II, \( \tilde{m}(AB/ab) = 0.9207831 \), whereas \( \tilde{m}(Ab/aB) = 0.9117230 \). For equilibrium III, these numerical values are reversed. To interpret the magnitude of this effect, it is useful to normalize the cis-trans difference by the maximum difference between the induced fitnesses. The normalized difference is approximately 0.025. Although small, it must be remembered that only one of three relevant loci is being ignored. More extreme numerical examples are provided in Table 3 below.
This example illustrates two points: (1) even extremely symmetrical models can produce associative cis-trans effects, and (2) even if fitness is inherently multiplicative and/or based on heterozygosity, unless all relevant loci are studied or there is global linkage equilibrium, induced fitnesses can follow irregular patterns. In particular, if viability selection is indeed multiplicative but sufficiently intense to generate stable linkage disequilibrium, then the induced fitnesses observed in subsystems will usually not be multiplicative. Lewontin (1964) first pointed out that two-locus theory cannot explain the patterns of two-locus gamete frequencies generated by multilocus selection. The associative cis-trans effect illustrates the complementary fact that patterns of induced two-locus fitnesses cannot be understood without multilocus theory. Hence, to understand the consequences of multiplicative or heterozygosity-based viability selection, one must study models that treat arbitrary numbers of loci (e.g., Roux 1974, 1978; Karlin 1979; Karlin and Avni 1981). This also demonstrates that empirical evidence for two-locus multiplicative epistasis does not in itself rule out multilocus multiplicative selection involving the loci observed.

Ewens and Thomson (1977) proved that at equilibria for n-locus selection models, the marginal frequencies of all lower order subsystems are equilibria for the corresponding induced fitness structures. Hence, one justification for the study of equilibrium properties of general two-locus selection schemes is that they reflect the induced properties of two-locus systems embedded in larger systems. This rationale has recently been used in a numerical study by Hastings (1981) relating the intensity of selection to the amount of linkage disequilibrium associated with equilibria of two-locus selection models. As demonstrated, cis-trans effects must be allowed if the Ewens and Thomson result is to be invoked correctly. Thus, there is no a priori justification for ignoring cis-trans differences in studies of general two-locus selection (cf. Hastings 1981; Feldman, Christiansen and Brooks 1980; Karlin and Carmelli 1975). With the exception of a specialized analysis by Parsons (1963a, b), no theoretical studies have explicitly explored the consequences of such effects. However, two classes of analysis that use general “small parameter” perturbation arguments have allowed them. The results of Karlin (1975, 1978) concerning the number of possible stable equilibria and the “induced overdominance” principle for multilocus selection with “small” recombination rates depend only on general multiple allele selection results and the perturbation results of Karlin and McGregor (1972). Similarly, general results of Nagylaki (1976, 1977) concerning “weak” selection are derived without assumptions on the assignment of fitnesses. Further analysis is required to determine whether the incorporation

---

**TABLE 1**

Equilibria of a three-locus model described in the text

<table>
<thead>
<tr>
<th>Equilibrium</th>
<th>( P_2 (ABC) = \frac{F_2 (abc)}{P_2 (abc)} )</th>
<th>( P_1 (AbC) = \frac{F_1 (AbC)}{P_2 (abc)} )</th>
<th>( P_3 (aBC) = \frac{F_3 (aBC)}{P_2 (abc)} )</th>
<th>( P_4 (AB) = \frac{F_4 (AB)}{P_2 (abc)} )</th>
<th>( P_5 (Ab) = \frac{F_5 (Ab)}{P_2 (abc)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.1375</td>
<td>0.1375</td>
<td>0.1125</td>
<td>0.1125</td>
<td>0.275</td>
</tr>
<tr>
<td>II</td>
<td>0.2628815</td>
<td>0.0982481</td>
<td>0.0456614</td>
<td>0.0932090</td>
<td>0.3611296</td>
</tr>
<tr>
<td>III</td>
<td>0.0932090</td>
<td>0.0456614</td>
<td>0.0982481</td>
<td>0.2628815</td>
<td>0.1388704</td>
</tr>
</tbody>
</table>
The existence of these effects may be more relevant to experimenters than theoreticians. Empirical demonstrations of cis-trans effects in eukaryotes have generally been restricted to very closely linked genetic elements (e.g., the analysis of the bithorax complex by Lewis 1955 and the analysis of the rosy locus by Chovnick et al. 1976). Typical interpretations involve cis-acting regulation of gene expression dependent on physical proximity. Few reports of cis-trans effects spanning several map units have appeared. Their rarity may be an artifact of the assumption that such effects do not exist, so that appropriate tests are rarely performed. However, two studies by Clark and Feldman (1981a, b) report such effects for recombination rates and various fitness components associated with nonoverlapping second chromosome inversions in Drosophila melanogaster. The inversions used contained recessive lethals so that only five adult inversion genotypes occurred, two of them doubly heterozygous. Some of the statistically significant, and fairly large, cis-trans effects that they reported are shown in Table 2. A slight reinterpretation of the calculations reveals that these effects may be attributable to nonrandom (but incomplete) association of the marked inversions with alleles, both within and outside the inverted segments, that govern recombination and fitness. The possible importance of such "background effects" was indicated by the dependence of the cis-trans difference on the genetic background used (see p. 463 of Clark and Feldman 1981b). As suggested by studies of whole chromosome mutation rates for viability (e.g., Mukai et al. 1972; Mukai and Cockerham 1977), the number of loci that contribute to fitness differences is probably quite large. The existence of loci governing region-specific recombination rates is demonstrated by selection experiments (e.g., Kidwell 1972) and assays of recombination rates in genetically distinct lines (e.g., Carson 1953 using D. robusta, Levine and Levine 1955 using D. pseudoobscura, and Marks and Brooks 1982 using D. melanogaster). Disequilibrium involving such loci could arise from two sources to generate the cis-trans differences observed by Clark and Feldman. The existence of linkage disequilibrium between inversions and allozyme loci in D. melanogaster is well documented (e.g., Langley, Tobari and Kojima 1974; Mukai, Watanabe and Yamaguchi 1974). Because founder effects are a likely cause (Ishii and Charlesworth 1977; Nei and Li 1980), disequilibrium probably also exists between inversions and loci affecting fitness and recombination. In addition to this "intrinsic" disequilibrium, more was probably generated in the laboratory population sample used by Clark and Feldman.

The hypothesis that linkage disequilibrium contributes to cis-trans effects, such as those reported by Clark and Feldman, is testable. Crosses that decrease linkage disequilibrium should result in progressively smaller cis-trans differences. (The homozygous and heterozygous lines of Clark and Feldman do not constitute such a test because they differ with respect to more than the occurrence of backcrossing; see Clark, Feldman and Christiansen 1981.) Although phenomena other than linkage disequilibrium may contribute to the cis-trans differences observed by Clark and Feldman, they are difficult to identify. Cis-trans effects spanning several map units are classically associated
### Table 2
Experimental estimates of cis-trans effects (± one standard error) from CLARK AND FELDMAN (1981a, b)

<table>
<thead>
<tr>
<th>Trait</th>
<th>cis</th>
<th>trans</th>
<th>Relative difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recombination</td>
<td>0.357 ± 0.012</td>
<td>0.414 ± 0.010</td>
<td>0.057</td>
</tr>
<tr>
<td>Viability</td>
<td>0.875 ± 0.038</td>
<td>0.774 ± 0.031</td>
<td>0.010</td>
</tr>
<tr>
<td>♂ Fecundity</td>
<td>0.753 ± 0.026</td>
<td>0.682 ± 0.024</td>
<td>0.071</td>
</tr>
<tr>
<td>♀ Fecundity</td>
<td>1.034 ± 0.033</td>
<td>0.925 ± 0.031</td>
<td>0.109</td>
</tr>
</tbody>
</table>

*a* Estimates of viability and fecundity compare double heterozygotes to wild type-double homozygotes.

*b* Defined by |cis-trans|/maximum difference observed between nonlethal genotypes.

*c* Estimated with a multiplicative model (see CLARK AND FELDMAN 1981b for an explanation).

With chromosomal rearrangements via position-effect variegation mediated by heterochromatin (see SPOFFORD 1976 for a review). Other effects, such as transvection phenomena (see LEWIS 1955 and JACK and JUDD 1979), attributable to the reduced efficiency of somatic pairing of homologs caused by the inversions would be expected to act equivalently in the cis and trans karyotypes. Explanation of the CLARK-FELDMAN results requires interaction of position-dependent effects produced by each inversion. This seems unlikely because the breakpoints of the inversions used (Cy: In(2L)22D1-2;33F5-34A1 and Pm²: In(2LR)40F;59E) are separated by more than 250 bands, whereas SPOFFORD (1976) reports a maximum known “effective distance” for variegation of 80 bands in D. melanogaster. Moreover, only the Pm² inversion displaces centric heterochromatin.

GIBSON and THODAY (1962) reported a linkage phase-dependent lethal effect which motivated PARSONS’ (1963a, b) analyses. Their results cannot be explained by linkage disequilibrium because, despite a distance of 20 map units, they were unable to produce the presumed lethal cis double heterozygote by recombination from marked stocks of the viable trans double heterozygote. This establishes that the lethality must be position dependent because only position distinguishes the cis recombinants from the parental trans types.

To illustrate the potential magnitude of associative cis-trans effects, computer simulations were performed. Viability selection at three diallelic loci was simulated with fitnesses randomly assigned to the 27 distinct, position-independent genotypes. The fitnesses were independent and uniformly distributed between zero and one. For each set of fitnesses, a single initial gamete frequency vector was chosen at random. The standard deterministic three-locus recursions were then iterated until a gamete was lost (defined as frequency below 10⁻³) or a stable three-locus polymorphism found. (The details of the simulation procedure and additional results will be reported elsewhere.) For the polymorphic cases, induced fitnesses, denoted \( \tilde{w}_i \), were computed for two adjacent loci. The recombination rate between these “observed” loci is denoted \( r_o \); the recombination rate between the middle locus and the “hidden” locus is \( r_h \). No interference was assumed. The relative magnitude of the induced cis-trans effects was
Statistics describing the relative magnitude of associative cis-trans effects observed in 100 simulations of three-locus selection*

<table>
<thead>
<tr>
<th>( r_0 )</th>
<th>0.01</th>
<th>0.1</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_h )</td>
<td>0.01</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Mean</td>
<td>0.1123</td>
<td>0.0892</td>
<td>0.0137</td>
<td>0.0213</td>
<td>0.0056</td>
</tr>
<tr>
<td>Median</td>
<td>0.0383</td>
<td>0.0384</td>
<td>0.0061</td>
<td>0.0159</td>
<td>0.0042</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.1656</td>
<td>0.1593</td>
<td>0.0197</td>
<td>0.0204</td>
<td>0.0048</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.7548</td>
<td>0.9174</td>
<td>0.0992</td>
<td>0.0878</td>
<td>0.0243</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.0002</td>
<td>0.0011</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

* See text for explanation.

Quantified by the index

\[
d = |\tilde{w}_{14} - \tilde{w}_{23}| / [\max(\tilde{w}_j) - \min(\tilde{w}_j)].
\]

For five separate linkage arrangements, 100 polymorphisms were generated. Statistics describing the results, in terms of \( d \), appear in Table 3. The most surprising result is that \( d \) can exceed 0.9. Thus, if only one of three relevant loci is ignored, the associative cis-trans effect can be quite large. In general, however, as shown by the large standard deviations and small medians, ignoring one of three relevant loci produces a small effect. As expected, the average magnitude generally decreases with looser linkage because of the lower levels of disequilibrium maintained (but compare \( r_0 = 0.5, r_h = 0.1 \) with \( r_0 = 0.5, r_h = 0.01 \) for an exception). However, even for three unlinked loci fairly large cis-trans differences can appear in this admittedly artificial system.

In addition to explaining seemingly anomalous results, associative cis-trans effects can be usefully exploited by experimenters. If multilocus linkage phase differences spanning several map units are found, a simple explanation is that they are consequences of disequilibrium as described. In organisms such as Drosophila in which very little disequilibrium between loci has been found in nature, the protocol of the experimenter is the most likely source. As demonstrated by Jones and Yamazaki (1974), such artificially generated disequilibrium can critically bias fitness estimation. Two-locus cis-trans tests provide an indirect one-way test for disequilibrium among fitness-determining loci. They also provide a check for studies aimed at determining the major loci governing a trait. The existence of associative cis-trans effects, experimentally supported by suitable crosses, implies that additional relevant loci remain to be located.

I am most grateful to Professor Jerry Coyne for several discussions and critical comments on earlier drafts. Suggestions from Ian Boussy, Andrew Clark, John Gillespie, Mel Green, Alan Hastings, John Kiger, Richard Lewontin, Bill Marks, Timothy Prout and an anonymous reviewer were extremely helpful. I thank Robert Barker and Richard Kennedy for the numerical calculations. This research was supported by National Institutes of Health grant GM22221.

**LITERATURE CITED**


ASSOCIATIVE CIS-TRANS EFFECTS


Corresponding editor: B. WEIR