Theories of quasi-linkage and "affinity": Some implications for population structure

(genetics/oncogenic virus/gametic phase disequilibrium/sustained meiotic affinity)

VALERIE MIKÉ

Biostatistics Laboratory, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

Communicated by Robert A. Good, June 8, 1977

ABSTRACT "Quasi-linkage" refers to nonrandom assortment of genes located on different chromosomes. Although this phenomenon has been widely observed in many organisms since the early part of this century, it is barely known. Interest in it was recently rekindled by the report of an association between quasi-linkage and the expression of genes belonging to a group of cancer viruses whose genomes are integrated in mouse chromosomes. This prompted an examination of the question whether "sustained meiotic affinity," which is one of the explanations proposed for quasi-linkage, can influence population structure in a manner unattainable by other known modes of heredity. It is shown for a two-locus two-allele system that equilibrium is attained with the gametic phase disequilibrium $D > 0$, leading to a permanent excess of the preferred genotypes. The possible relationship of these concepts to the inheritance of susceptibility to cancer and other diseases is discussed.

According to Mendel’s second law, "the relation of each pair of different characters in hybrid union is independent of the other differences in the two original parental stocks." In modern terms this law, modified to take account of chromosomal linkage, states that the alleles of genes located on different chromosomes (or more than 50 centimorgans apart in the same chromosome) assort independently of each other.

But exceptions to this rule were already being observed in a variety of organisms during the development of classical genetics in the early part of this century, and numerous others have been recorded since. They take the form of apparent associations between unlinked genes, giving rise to segregation data that (i) simulate linkage, or (ii) indicate paradoxically an excess of recombination, meaning a deficiency rather than an excess of parental types; they are characterized as weak associations. They have been called "quasi-linkage" and "super-recombination," respectively (1), although they are frequently referred to collectively as quasi-linkage.

Neither phenomenon has received much general notice; standard textbooks of genetics usually omit mention of them altogether. This may be due in part to the fact that the observed segregation distortions were weak. Also, the underlying mechanism is unknown, and its continuity through many generations is uncertain. Therefore, its influence on population structure was probably thought to be too insignificant to be worth attention.

1. Current interest in quasi-linkage

Renewed interest in the topic has now been sparked by the recognition of an association between quasi-linkage and a certain class of cancer virus in the mouse (2); in this context the phenomenon of quasi-linkage may have far-reaching implications (3). This work will be sketched briefly below.

Recombination. Segregation tests are normally carried out for the purpose of establishing the location of genes. Consider two loci with two alleles each, $Aa$ and $Bb$. Assume that the hybrid resulting from a cross between double homozygotes $AB/AB$ and $ab/ab$ is backcrossed to $AB/AB$, and that the parental haplotypes $AB$ and $ab$ are reversed by crossing over in 10% of meiotic divisions. Then, other things being equal, 90% of the progeny will be of parental type, and 10% of nonparental or recombinant type, and the map distance between the two loci will be estimated as 10 centimorgans. It is important to note, however, that these figures are obtained by ascertaining the frequencies of progeny types, and therefore do not indicate directly the reason for the deficiency of recombinants. In cases of loose linkage (high recombination fractions) this is an important reservation. For the deficiency of recombinants might be caused by preferential fertilization between gametes of parental type ($AB$ and $ab$) or by loss of embryos of nonparental type. These possibilities have to be taken into account in considering quasi-linkage.

Quasi-Linkage and Oncogenic Virus. Viruses of a group known variously as "oncornavirus," "leukemia-sarcoma virus," and "C-type and B-type" virus commonly cause cancers in mice. Their genetic material is RNA, but DNA copies of the viral genome are integrated in the chromosomes of the mouse. These DNA viral genes in the chromosomes do not usually cause the production of the virus itself, and are inherited like Mendelian genes. But one of these viral genes may be expressed, in which case one component of the virus appears in pedigrees as a Mendelian character without evident virus. $G_{IX}$-gp70 is one such molecule. It is a glycoprotein which is found in the plasma membrane of thymocytes and in certain other sites, and which forms the major component of the outer coat of C-type leukemia virus, if the cell is producing virus. There are strains of mice in which the $G_{IX}$-gp70 molecule is expressed, and other strains in which it is not. The crucial finding of Stockert et al. (2) is that the expression of $G_{IX}$-gp70 is associated with quasi-linkage between one of the genes for $G_{IX} (Gv-1)$ and two other genes, $Fe-1$ on chromosome 4, and the major histocompatibility region $H-2$ on chromosome 17. It is provocative that both these latter loci are known to affect the production of C-type leukemia virus.

Boyse (3) makes two points: (i) The biological role of these integrated viral genomes is a mystery. They may have basic physiological functions. The realization that one such viral gene is involved in the phenomenon of quasi-linkage may suggest ways to approach this problem. (ii) Expression of $G_{IX}$-gp70 may influence susceptibility to leukemia, and it also causes a severe autoimmune disease in mice of certain genotypes (4). Quasi-linkage, whatever its mechanism, may help to exclude disad-
Mechanisms of quasi-linkage

The association between quasi-linkage and the partial expression of a leukemia viral genome is obviously of potential importance. This is what prompted the present report, the aim of which is to consider the effect of quasi-linkage on population structure.

The whole question of quasi-linkage and super-recombination, and their possible mechanisms, now requires closer study. The literature reporting such associations has not yet been completely reviewed. For the present we need only consider two general theories whose implications for population structure are quite different. As stated above, quasi-linkage may have nothing to do with meiosis, and Boyse (3) describes segregation tests that would answer this question. If it is meiotic, two modes of inheritance can be envisaged.

Boyse (3) defines one possible mode as "variable affinity." [The term "affinity" was proposed by Malinowski (5) in 1927.] According to this model the association of one allele with another would follow the parental pattern. Thus, if the heterozygote in the cross referred to above received alleles A and B from one parent, and alleles a and b from the other (which we can arbitrarily call "coupling"), then the same associations would be seen in the progeny of this heterozygote (quasi-linkage). And if the heterozygote received A and b from one parent and a and B from the other (which we may call "repulsion"), then these parental associations would be preserved in the backcross (again quasi-linkage). The effects of variable affinity would then not be different from loose linkage. Moreover, super-recombination would have to be interpreted on a basis different from that of quasi-linkage.

In what Boyse calls "sustained affinity," however, A would tend to be associated with B, and a with b, regardless of coupling or repulsion. The result would be quasi-linkage in coupling crosses, and super-recombination in repulsion crosses. Sustained affinity has the virtue of simplicity in that (i) it is easier to envisage how different members of two pairs of non-homologous chromosomes might tend to "stick together" (affinity), and (ii) quasi-linkage and super-recombination would be explicable by the same mechanism.

In the following we will examine the implications for population structure of the model of sustained meiotic affinity in a two-locus, two-allele system. This will entail a review of known results for linkage, and these as mentioned apply also to the model of variable affinity.

3. The classical framework: Hardy–Weinberg law

In its basic version this law, formulated in 1908 independently by G. H. Hardy and W. Weinberg, states that for a gene with two alleles A and a, with frequencies p and q, under the assumption of random mating in an infinitely large population where there is no selection, migration, or mutation, the equilibrium frequencies of the three genotypes AA, Aa, and aa are p^2, 2pq, and q^2, respectively, and they are reached in one generation.

For two or more loci considered simultaneously the attainment of equilibrium is not immediate. Consider as before two genes with two alleles Aa and Bb, yielding the four types of gametes AB, Ab, aB, and ab. Let their respective relative frequencies in the population be denoted by x_1, x_2, x_3, and x_4, where

\[ \sum_{i=1}^{4} x_i = 1. \]

The gene frequencies may then be expressed as

\[ P(A) = p_1 = x_1 + x_2 \]
\[ P(a) = q_1 = x_3 + x_4 \]
\[ P(B) = p_2 = x_1 + x_3 \]
\[ P(b) = q_2 = x_2 + x_4. \]

Because random mating is equivalent to random union of gametes, the genotype frequencies are the products of the corresponding gamete frequencies.

Let \( r = \) recombination fraction between the two loci; \( r \leq 0.5 \), with the equality holding for the case of unlinked loci. Then under mating conditions as above, assuming discrete generations and identical behavior of male and female gametes, it is a well-known result (see, e.g., ref. 6) that gamete frequencies in two successive generations can be expressed in the form

\[ x_1' = x_1 - rD \]
\[ x_2' = x_2 + rD \]
\[ x_3' = x_3 + rD \]
\[ x_4' = x_4 - rD, \]

where

\[ D = x_1x_4 - x_2x_3, \]

a term referred to in the literature variously as "linkage disequilibrium," "gametic association," or "gametic phase disequilibrium."

Also, for any generation,

\[ x_1 = p_1p_2 + D \]
\[ x_2 = p_1q_2 - D \]
\[ x_3 = q_1p_2 - D \]
\[ x_4 = q_1q_2 + D \]

and, in the nth generation,

\[ D^{(n)} = (1 - r)^nD^{(0)}, \]

so that

\[ \lim_{n \to \infty} D^{(n)} = 0. \]

That is, at equilibrium each gamete frequency is the product of the corresponding gene frequencies and independent of the recombination fraction \( r \). Thus, the Hardy–Weinberg equilibrium is attained, but not in one generation. The rate of approach is a function of \( r \) and is highest for \( r = 0.5 \), the case of unlinked loci.

This result was obtained for an arbitrary number of generations by Jennings (7), and in its asymptotic form by Robbins (8). Geiringer (9) derived the analogous theorem for three loci; the general case of \( m \) loci was treated by Geiringer (10) and Bennett (11).

4. Quasi-linkage as sustained meiotic affinity

Gamete Frequencies. As above, we consider two loci with two alleles each, and the same notation for gene and gamete frequencies.

Let \( r = \) recombination/affinity fraction involving the two double heterozygotes. In the context of sustained meiotic affinity \( r \) is the probability that \( AB/ab \) (coupling phase) produces recombinant-type gametes \( Ab \) and \( aB \), and also the probability that \( Ab/aB \) (repulsion phase) produces parental-type gametes.
Table 1. Gametic output of 10 genotypes under model of sustained meiotic affinity

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
<th>Gametic output</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB/AB</td>
<td>x₁²</td>
<td>AB</td>
</tr>
<tr>
<td>Ab/Ab</td>
<td>x₂²</td>
<td>Ab</td>
</tr>
<tr>
<td>aB/aB</td>
<td>x₃²</td>
<td>aB</td>
</tr>
<tr>
<td>ab/ab</td>
<td>x₄²</td>
<td>ab</td>
</tr>
<tr>
<td>AB/Ab</td>
<td>½x₁; x₂</td>
<td>½AB + ½ab</td>
</tr>
<tr>
<td>Ab/AB</td>
<td>½x₂; x₃</td>
<td>½AB + ½aB</td>
</tr>
<tr>
<td>ab/Ab</td>
<td>½x₃; x₄</td>
<td>½Ab + ½aB</td>
</tr>
<tr>
<td>Ab/ab</td>
<td>½x₄; x₁</td>
<td>½Ab + ½ab</td>
</tr>
<tr>
<td>Ab/Ab</td>
<td>½x₁; x₂</td>
<td>½(1-r)AB + ½(1-r)ab + ½rAb + ½raB</td>
</tr>
<tr>
<td>Ab/Ab</td>
<td>½x₂; x₃</td>
<td>½(1-r)AB + ½(1-r)ab + ½rAb + ½raB</td>
</tr>
</tbody>
</table>

Ab and aB. In other words, either double heterozygote produces AB and ab gametes with probability (1-r). Thus, for repulsion phase hybrids (1-r) is the probability of what has been referred to as super-recombination, and r itself may be called the “affinity fraction.” We assume 0 < r < 0.5.

Under the mating assumptions stated in the previous section, the gametic output provided by the ten possible genotypes is shown in Table 1. Only the last of these is affected by the affinity model under consideration; the other genotypes have the same gametic output as in the case of conventional linkage. From this table we can obtain the gametic frequencies xᵢ in the succeeding generation:

\[ x'_i = x_i^2 + x_1x_2 + x_1x_3 + (1-r)x_1x_4 + (1-r)x_2x_3 \]

\[ = x_1 - [rx_1x_4 - (1-r)x_2x_3]; \]

similarly,

\[ x'_2 = x_2 + [rx_1x_4 - (1-r)x_2x_3] \]
\[ x'_3 = x_3 + [rx_1x_4 - (1-r)x_2x_3] \]
\[ x'_4 = x_4 - [rx_1x_4 - (1-r)x_2x_3]. \]

The expression resulting from the model of sustained meiotic affinity

\[ A = rx_1x_4 - (1-r)x_2x_3 \]

is analogous to the gametic phase disequilibrium term

\[ rD = r(x_1x_4 - x_2x_3) \]

obtained for linkage.

Convergence of “Affinity Component” A. Let

\[ A^{(n)} = rx_1^{(n)}x_4^{(n)} - (1-r)x_2^{(n)}x_3^{(n)}, \]

n = 0, 1, 2, ... be the “affinity component” of gamete frequencies in the (n + 1)st generation. It can be shown that A⁽ⁿ⁾ is a monotone sequence, and that

\[ \lim_{n \to \infty} A^{(n)} = 0. \]

Because

\[ x_i^{(n+1)} - x_i^{(n)} = \delta_i A^{(n)}, \]

where \( \delta_i = \begin{cases} +1, & \text{for } i = 2,3 \\ -1, & \text{for } i = 1,4 \end{cases} \]

this result is equivalent to the convergence of gamete frequencies. Without loss of generality let two consecutive terms be denoted by A and A'. Then

\[ A' = r(x_1 - A)(x_4 - A) - (1-r)(x_2 + A)(x_3 + A) \]
\[ = rx_1x_4 - (1-r)x_2x_3 - A(1-r)(x_2 + x_3) - r(A - x_1 - x_4) \]
\[ = A - A[(1-r)(A + x_2 + x_3) - r(A - 1 + x_2 + x_3)] \]
\[ = [(1-r) - (1-2r)(A + x_2 + x_3)]A \]
\[ = kA. \]

But

\[ r < k < 1 - r, \]

because

\[ 0 < A + x_2 + x_3 < 1. \]

We have

(i) A + x₂ = x₂ > 0 and x₃ ≥ 0, for A > 0
A + x₂ = x₂ > 0 and x₃ > 0, for A < 0,
and
(ii) A + x₂ + x₃ < x₂ + x₃ < 1, for A > 0
A + x₂ + x₃ < x₂ + x₃ ≤ 1, for A < 0,
so that the result follows. Thus the sequence is monotonic, and we have in addition shown that

\[ r|A| < |A'| < (1-r)|A|. \]

Now let

\[ k_j = (1-r) - (1-2r)(A^{(j)} + x_2^{(j)} + x_3^{(j)}), \quad j = 0,1,2,\ldots \]

Then

\[ A^{(1)} = k_0A^{(0)} \]
\[ A^{(2)} = k_1k_0A^{(0)} \]
\[ \vdots \]
\[ A^{(n)} = \left( \prod_{j=0}^{n-1} k_j \right)A^{(0)}, \]

and

\[ r^n|A^{(0)}| < |A^{(n)}| < (1-r)^n|A^{(0)}|, \]

so that A⁽ⁿ⁾ → 0. The rate of convergence is bounded by r and (1-r), and is a function also of the initial gamete frequencies.

Table 2. Pooled segregation data obtained from backcrosses involving several inbred strains of mice

<table>
<thead>
<tr>
<th>Genes</th>
<th>Genotypes</th>
<th>Recombinant</th>
<th>Non-recombinant</th>
<th>Total</th>
<th>r</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gv-1:H-2</td>
<td>sx₁=311</td>
<td>sx₂=580</td>
<td>sx₃=891</td>
<td>0.35</td>
<td>0.32-0.38</td>
<td></td>
</tr>
<tr>
<td>Gv-1:Gpd-1</td>
<td>sx₁=105</td>
<td>sx₂=219</td>
<td>sx₃=324</td>
<td>0.32</td>
<td>0.27-0.37</td>
<td></td>
</tr>
<tr>
<td>H-2:Gpd-1</td>
<td>sx₁=223</td>
<td>sx₂=221</td>
<td>sx₃=444</td>
<td>0.50</td>
<td>0.45-0.55</td>
<td></td>
</tr>
<tr>
<td>Gv-1:Tla (homozygous for H-2)</td>
<td>sx₁=110</td>
<td>sx₂=120</td>
<td>sx₃=230</td>
<td>0.48</td>
<td>0.42-0.54</td>
<td></td>
</tr>
<tr>
<td>Gv-1:Gpd-1 (homozygous for Fl-1)</td>
<td>sx₁=84</td>
<td>sx₂=76</td>
<td>sx₃=160</td>
<td>0.53</td>
<td>0.45-0.61</td>
<td></td>
</tr>
</tbody>
</table>
Convergence to equilibrium is thus always faster than for conventional linkage.

**Quasi-Linkage in Equilibrium.** Given the above, we can solve for the equilibrium gamete frequencies using the system of equations

\[
\begin{align*}
    p_1 &= x_1 + x_2 \\
    p_2 &= x_1 + x_3 \\
    x_1 + x_2 + x_3 + x_4 &= 1
\end{align*}
\]

and the asymptotic result

\[r x_1 x_4 - (1 - r) x_2 x_3 = 0.\]

Solution of a quadratic equation yields

\[x_1 = \frac{(p_1 + p_2)(1 - 2r) + r \pm \sqrt{(p_1 + p_2)(1 - 2r) + r^2 - 4(1 - 2r)(1 - r)p_1 p_2}}{2(1 - 2r)}\]

It can be shown that the smaller of the two roots is always the unique equilibrium value, regardless of initial state. The other frequencies can readily be obtained from the first three equations above.

We see here that unlike in the case with linkage, the equilibrium gamete (and hence genotype) frequencies in sustained meiotic affinity depend not only on the (constant) gene frequencies \(p_1\) and \(p_2\), but also on the recombination/affinity fraction \(r\).

5. Example

The recombination data reported by Stockert et al. (2) for a series of backcrosses derived from several inbred strains of mice are summarized in Table 2. The five genes considered in these studies are represented in Fig. 1. All observed associations in these experiments represented quasi-linkage; i.e., the deficiency of segregants was of parental, not nonparental, types.

A recombination fraction of 0.35 was observed in the pooled segregation data for Go-1:H-2. If this were true linkage, then Go-1 should also be linked to Tla, which is located near the H-2 region on chromosome 17. But the Go-1:Tla results indicate no association \((r = 0.48)\). The recombination fraction between Go-1 and Gpd-1 was 0.32; this association disappeared in crosses homozygous for Fe-1, a gene closely linked to Gpd-1 on chromosome 4. This indicates that the observed association was due to quasi-linkage between Go-1 and Fe-1. Thus, the actual location of Go-1 is not known. In a three-point backcross involving heterozygosity for Go-1, H-2, and Gpd-1, both associations were seen in the same population. There was no association between H-2 and Gpd-1.

6. Discussion

The three genes for which quasi-linkage was observed in the studies reviewed above all have a connection with leukemia virus. Each of these loci, Go-1, H-2, and Fe-1, can be suspected of affecting the susceptibility of individual mice to leukemia and other cancers of similar viral etiology; this has already been confirmed for spontaneous leukemia and the H-2 locus (12).

The GIX antigen is also involved in the expression of an autoimmune disease that occurs in all hybrids positive for this antigen in a cross in which neither inbred parent strain is affected (4). It has been proposed that this autoimmune response depends on two dominant Ir genes, one being contributed by each parent.

If groups of unlinked alleles predispose to disease syndromes, then a mechanism for preserving certain constellations of alleles and eliminating others has selective advantage.

Very closely linked genes preserve their association for many generations, which can give rise to linkage disequilibrium as a mechanism for lowering the incidence of individuals susceptible to disease. In loose linkage this effect is transient. Of the two modes of meiotic affinity, variable affinity would have the same effect as weak linkage. As can be seen in Fig. 2, for the model provided by two inbred strains, gametic phase equilibrium is soon attained. Sustained affinity, on the other hand, guarantees a permanent excess of the favorable genotypes. The effect may be small in any one instance; but judging from the

*Fig. 2. Gametic phase disequilibrium \(D = x_1 x_4 - x_2 x_3\) for linkage (or variable affinity) and quasi-linkage (sustained affinity). Recombination/affinity fraction \(r = 0.35\). For linkage \(r\) denotes the conventional recombination fraction; for quasi-linkage it is the probability of recombination for coupling phase heterozygotes and of nonrecombination (affinity) for repulsion phase heterozygotes.*
literature quasi-linkage is common, so that multiple effects may generally be operative.

The mathematical structure of quasi-linkage has been examined here for the case of a single pair of loci. The next step in analyzing this model will entail estimating the combined effect of quasi-linkage involving several loci, with due consideration of selection pressures against disadvantageous genotypes.

The author is grateful to Dr. Edward A. Boyse for advice on biological matters, and to Drs. Mark Brown and H. Tzvi Thaler for stimulating discussions concerning the equilibrium properties of the model. This work was supported by National Cancer Institute Grant CA 08748.

7. Jennings, H. S. (1917) "The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage," *Genetics* 2, 97–154.