Digenic genotypes: the interface of inbreeding, linkage, and linkage disequilibrium

Reginald D. Smith

Ronin Institute 127 Haddon Pl, Montclair, New Jersey 07043, US; and Supreme Vinegar LLC, 3430 Progress Dr. Suite D, Bensalem, PA 19020

Abstract

Many traits in populations are well understood as being Mendelian effects at single loci or additive polygenic effects across numerous loci. However, there are important phenomena and traits that are intermediate between these two extremes and are known as oligogenic traits. Here we investigate digenic, or two-locus, traits and how their frequencies in populations are affected by non-random mating, specifically inbreeding, linkage disequilibrium, and selection. These effects are examined both separately and in combination to demonstrate how many digenic traits, especially double homozygous ones, can show significant, sometimes unexpected, changes in population frequency with inbreeding, linkage, and linkage disequilibrium. The effects of selection on deleterious digenic traits are also detailed. These results are applied to both digenic traits of medical significance as well as measuring inbreeding in natural populations.

Keywords: digenic traits, inbreeding, selection, linkage disequilibrium, medical genetics, conservation genetics

1. Introduction

Starting in the early 1960s, population geneticists began building on the advances made understanding the dynamics of single loci subject to evolutionary forces to analyze the dynamics of multilocus systems with two or more loci. Early work on multilocus genetics included (Hogben, 1932) and (Li, 1953) on double homozygous recessive traits and (Haldane, 1949) in the

Email address: rsmith@supremevinegar.com (Reginald D. Smith)

first steps of understanding the effects of inbreeding at two loci. Digenic models were also essential in theories by Fisher and others on the evolution of dominance (Fisher, 1928a,b, 1929).

The dynamics of two loci subject to selection was a subject of early work by Kimura (Kimura, 1956) but was first developed into a general theory of the effect of selection on digenic traits by Lewontin & Kojima (Lewontin & Kojima, 1960). Their paper described the general equations for the evolution of haplotype frequencies and the effects of linkage and linkage disequilibrium on changes in haplotype frequency and stable states. Key to the work by both groups of authors was the analysis of the dynamics of two loci where the relative fitness matrix is symmetric allowing the problem to be analytically tractable and guaranteeing the existence of stable solutions under certain conditions of relative fitness.

Lewontin and Kojima's paper stimulated much interest in the topic, especially in light of the increasing availability of polymorphism data, which include (Bodmer & Felsenstein, 1967; Karlin & Feldman, 1970; Karlin, 1975, 1979; Karlin & Avni, 1981; Bürger, 2020). A good review of some of the historically important results can be found in (Karlin, 1975) and (Bürger, 2020).

1.1. Medical genetics and digenic traits

One of the first practical applications of the population genetics of two loci was in medical genetics. From the late 1970s to the 1990s, digenic traits, especially where they could possibly be of relevance to explaining medical conditions, were increasingly investigated. A disease model based of the symmetric fitness models of Lewontin, Kojima, and Karlin was given by (Merry et. al., 1979). In addition, the enumeration of medically relevant two locus genotypes and the covariance of relatives where the loci are unlinked is given by (Neuman & Chakravarti, 1992) building on earlier work by (Hartl, 1968). Full enumeration of all possible digenic trait genotypes were given by (Li & Reich, 2000; Hallgrímsdóttir & Yuster, 2008).

Two locus disease models continued to evolve throughout the 1980s and 1990s with important contributions by (Hodge, 1981a; Hodge & Spence, 1981b; Goldin & Weeks, 1993). These were complimented by the first confirmed discovery of a digenic disease trait: a double heterozygous genotype that causes a variation of retinitis pigmentosa (Kajiwara et. al., 1994). However, much interest in digenic traits temporarily waned with the advent of advanced technologies such as next generation sequencing and methods such

as GWAS which allowed analysis of large numbers of loci compared to previously limited polymorphism data.

However, there has been a resurgence of interest in so-called oligogenic disorders, defined as disorders whose aetiology is described by the genotypes at two or more loci but whose primary mechanisms are epistatic and not additive. Oligogenic disorders occupy a midpoint between single locus Mendelian disorders and complex diseases defined as quantitative traits. Several reviews (Badano & Katsanis, 2002; Cooper et. al., 2013; Schäffer, 2013; Deltas, 2018) have discussed the increasing number of discoveries of digenic or oligogenic diseases. In (Schäffer, 2013) ninety-five reports of oligogenic disorders were listed along with the method of inheritance for each locus. In the Online Mendelian Inheritance in Man (OMIM) database (OMIM, 2022), there are currently 33 entries for disorders that have an either digenic dominant or digenic recessive aetiology. About half are inherited in an autosomal dominant manner with double heterozygote genotypes most commonly described. This however, is still minuscule compared to the over 2,800 disorders with autosomal recessive aetiology and over 2,000 disorders with autosomal dominant aetiology in OMIM. While still relatively small in confirmed numbers, they are subject of renewed search and interest especially in investigations of disorders with incomplete penetrance across a known disease genotype.

1.2. Measuring inbreeding using digenic genotypes

One of the most well-known effects of inbreeding is the increase of homozygosity and the decrease of heterozygosity at individual loci. However, when two locus dynamics are considered, some additional and unexpected effects emerge. As will be described in detail later, one effect of inbreeding is to cause correlations between genotypes at different loci, particularly when they are linked. This correlation is positive for two loci with homozygous genotypes or two loci with heterozygous genotypes. This has been exploited in fields such as genetics of wild populations or conservation genetics (David, 1998; Balloux et. al., 2004; David et. al., 2007) as a way to measure the extent of inbreeding in populations using genetic data where pedigrees are unavailable. This provides an alternate measure of inbreeding besides other measures such as runs of homozygosity.

1.3. Applications of the dynamics of digenic traits

While much rich theoretical work has been done on two locus systems over the past 50 years, the bulk of work has been restricted to the theoretical

realm, much more so than either single locus or polygenic population genetics. One of the most interesting aspects has been the lack of attention to the effects of selection on favorable or deleterious digenic traits which are driven to fixation or extinction. The majority of theoretical studies on selection and digenic traits have been the studies of stable equilibria. Similarly, while the effects of linkage disequilibrium and non-random mating are known, they have only occasionally entered the applied genetics realm, possibly due to a relative lack of expositions that clearly apply theoretical results to applications of interest in medical or conservation genetics. The purpose of this paper is to not only review past results but allow them to be put into a context where they are more accessible to the applied genetics field and can be used to understand or interpret results in the burgeoning study of oligogenic traits, particularly deleterious ones. This will enable a greater appreciation for the effects of inbreeding, linkage disequilibrium, and selection for digenic traits which is common for Mendelian or complex traits.

2. The effects of inbreeding on two loci in linkage equilibrium

It has been known for centuries that breeding of relatives can lead to an increasing frequency of relatively rare disorders. The modern analysis of inbreeding was formalized by Wright (Wright, 1922) and improved by Malécot's probabilistic definition of the probability of two alleles at a locus being identical by descent (IBD) (Malécot, 1970). The inbreeding coefficient of descent F is defined as the IBD probability for two alleles in a locus and is equal to the coefficient of coancestry, θ, of the parents, also called kinship (e.g. (Jacquard, 1975)). In other words, the inbreeding of the progeny reflects the consanguineous relationship of the parents. Therefore the probability, F, that two alleles in the progeny are IBD is equal to the probability that two alleles from the same locus in each parent are IBD as well.

The genotypic results of inbreeding are widely known, most markedly increased homozygosity by an amount $p(1-p)F$ and a decline in heterozygosity. This increase in homozygosity allows the appearance of rare homozygous genotypes at significantly higher frequencies than if the population was outbred. This increases the frequencies of recessively inherited disorders in inbred populations.

The story for two loci is a bit more complex. First investigated by JBS Haldane (Haldane, 1949) and most thoroughly expounded in the collaboration of C. Clark Cockerham and Bruce Weir (Weir & Cockerham, 1968;

Cockerham & Weir, 1968, 1973, 1977; Weir & Cockerham, 1974), the effects of inbreeding across two loci involves not just the effects of increased homozygosity on the joint occurrence of genotypes.

Where two loci are concerned, we want to first investigate the conditions under which alleles at each locus in different individuals are identical by descent. We will designate the two alleles at the first locus in individuals 1 and 2 as a_1 and a_2 and at the second locus as b_1 and b_2. Note that these only represent the *identity* of the alleles and not the specific allele present. Per the notations of Weir and Cockerham we will define four different types of two-locus inbreeding coefficients:

$$\begin{aligned} F_{11} &= P(a_1 \equiv a_2) \text{ and } P(b_1 \equiv b_2) \\ F_{10} &= P(a_1 \equiv a_2) \text{ and } P(b_1 \not\equiv b_2) \\ F_{01} &= P(a_1 \not\equiv a_2) \text{ and } P(b_1 \equiv b_2) \\ F_{00} &= P(a_1 \not\equiv a_2) \text{ and } P(b_1 \not\equiv b_2) \end{aligned} \tag{1}$$

The $\equiv$ symbol indicates identity by descent. Note the single locus inbreeding coefficient for locus a is $F_{1.} = F_{11} + F_{10}$ and the coefficient for locus b is $F_{.1} = F_{11} + F_{01}$. The average of these two is often used to represent the average probability of IBD across both loci and is designated as $F_1 = (F_{1.} + F_{.1})/2$. For our analysis, we will assume the IBD probabilities at each locus are identical so $F_{.1} = F_{1.} = F_1 = F$. In evaluating the effects of inbreeding on two loci, the coefficient F_{11} is most useful in describing the impact of inbreeding at both loci. A graphical representation of F_{11} is given in Figure 3.

In particular for the determination of genotype frequencies is the importance of the identity disequilibrium (ID), η_{11}. The ID is a measurement of the increased frequency of joint IBD alleles at both loci above the IBD expected based on the product of the inbreeding coefficients at each locus. Where the inbreeding coefficient is the same for both loci, ID is defined as

$$\eta_{11} = F_{11} - F^2 \tag{2}$$

The coefficient F_{11}, however is not a fixed value across all loci pairs and all individuals in the population. In particular, F_{11} depends on the level of linkage between the two loci as well as the pedigree of ancestors back to the original gamete to account for all possibilities of transmission. Its boundary values are $F_{11} = F^2$ if $c = 1/2$, where c is the recombination frequency, and there is no linkage. Thus for $c = 1/2$, $\eta_{11} = 0$. For complete linkage where

$c = 0$, $F_{11} = F$ and thus $\eta_{11} = F(1-F)$. Intermediate values of F_{11} are calculated with algorithms given the pedigree of the individual. These will be explained in Appendix A but are introduced in (Haldane, 1949; Weir & Cockerham, 1968).

The main effect of identity disequilibrium generated by inbreeding is to increase the frequency of double homozygous and double heterozygous genotypes while reducing the frequency of genotypes that are pairs of homozygous and heterozygous genotypes. The increase in double homozygotes under inbreeding is largely expected but a positive counterbalance to the decrease of double heterozygous genotypes under inbreeding is often not appreciated. The effects of η_{11} by genotype are shown in Table 1. One way to interpret identity disequilibrium is as the correlation between genotypes at different loci. In other words, two loci are more likely to show double homozygosity or double heterozygosity than expected by chance.

	AA	Aa	aa
BB	$P_{AA}P_{BB} + \sigma_A^2\sigma_B^2\eta_{11}$	$P_{Aa}P_{BB} - 2\sigma_A^2\sigma_B^2\eta_{11}$	$P_{aa}P_{BB} + \sigma_A^2\sigma_B^2\eta_{11}$
Bb	$P_{AA}P_{Bb} - 2\sigma_A^2\sigma_B^2\eta_{11}$	$P_{Aa}P_{Bb} + 4\sigma_A^2\sigma_B^2\eta_{11}$	$P_{aa}P_{Bb} - 2\sigma_A^2\sigma_B^2\eta_{11}$
bb	$P_{AA}P_{bb} + \sigma_A^2\sigma_B^2\eta_{11}$	$P_{Aa}P_{bb} - 2\sigma_A^2\sigma_B^2\eta_{11}$	$P_{aa}P_{bb} + \sigma_A^2\sigma_B^2\eta_{11}$

Table 1: Two loci genotype frequencies under inbreeding at linkage equilibrium taking identity disequilibrium into account. For original explanations and derivations see (Haldane, 1949), though he uses ϕ for η_{11}. The variance of each locus is represented by $\sigma_A^2 = p_A(1-p_A)$ and $\sigma_B^2 = p_B(1-p_B)$. The variable η_{11} is the identity disequilibrium. Per the regular results of inbreeding, $P_{AA} = p_A^2 + p_A(1-p_A)F$, $P_{Aa} = 2p_A(1-p_A)(1-F)$, $P_{aa} = (1-p_A)^2 + p_A(1-p_A)F$, $P_{BB} = p_B^2 + p_B(1-p_B)F$, $P_{Bb} = 2p_B(1-p_B)(1-F)$, $P_{bb} = (1-p_B)^2 + p_B(1-p_B)F$. If the loci are unlinked $\eta_{11} = 0$ and if they are completely linked $\eta_{11} = F(1-F)$. Note double heterozygotes have a factor of four for identity disequilibrium due to the two types of combining gametes (AB and ab or Ab and aB) that create double heterozygotes and the possibility of each gamete being on the paternal or maternal contributed chromosome. Similarly homozygotes-heterozygotes have a factor of two in relation to identity disequilibrium since each genotype can appear two different ways depending on its contribution from maternal or paternal chromosomes.

Where inbreeding is present, the expected single-locus homozygous and heterozygote frequencies are given in the text of Table 1. Measuring covariance similar to that in linkage disequilibrium, we can show the covariances for double heterozygosity and double homozygosity at loci are only based on the identity disequilibria and the variances of heterozygosity frequency at each locus.

$$
\begin{aligned}
\mathrm{cov}_{Aa,Bb} =& P_{AaBb} - P_{Aa}P_{Bb} = & P_{Aa}P_{Bb} + 4\sigma_A^2\sigma_B^2\eta_{11} - P_{Aa}P_{Bb} =& 4\sigma_A^2\sigma_B^2\eta_{11} \\
\mathrm{cov}_{AA,BB} =& P_{AABB} - P_{AA}P_{BB} = & P_{AA}P_{BB} + \sigma_A^2\sigma_B^2\eta_{11} - P_{AA}P_{BB} =& \sigma_A^2\sigma_B^2\eta_{11} \\
\mathrm{cov}_{aa,bb} =& P_{aabb} - P_{aa}P_{bb} = & P_{aa}P_{bb} + \sigma_A^2\sigma_B^2\eta_{11} - P_{aa}P_{bb} =& \sigma_A^2\sigma_B^2\eta_{11} \\
\mathrm{cov}_{AA,bb} =& P_{AAbb} - P_{AA}P_{bb} = & P_{AA}P_{bb} + \sigma_A^2\sigma_B^2\eta_{11} - P_{AA}P_{bb} =& \sigma_A^2\sigma_B^2\eta_{11} \\
\mathrm{cov}_{aa,BB} =& P_{aaBB} - P_{aa}P_{BB} = & P_{aa}P_{BB} + \sigma_A^2\sigma_B^2\eta_{11} - P_{aa}P_{BB} =& \sigma_A^2\sigma_B^2\eta_{11}
\end{aligned}
\tag{3}
$$

The correlations, however, are different though positive in these cases as shown below. A key finding is that the strength of the correlations varies not only with the identity disequilibrium but also directly with the degree of polymorphism of the loci. The larger the variance of each locus, the larger the correlation. As F increases towards 1, however, F_{11} and F both approach one (see (Cockerham & Weir, 1973), page 316) since the widespread homozygosity reduces the effect of recombination to create different varieties of genotypes in offspring and so the correlation approaches zero equaling zero when $F = 1$.

$$
\begin{aligned}
r_{Aa,Bb} =& \frac{4\sigma_A^2\sigma_B^2\eta_{11}}{\sqrt{P_{Aa}P_{Bb}(1-P_{Aa})(1-P_{Bb})}} \\
r_{AA,BB} =& \frac{\sigma_A^2\sigma_B^2\eta_{11}}{\sqrt{P_{AA}P_{BB}(1-P_{AA})(1-P_{BB})}} \\
r_{aa,bb} =& \frac{\sigma_A^2\sigma_B^2\eta_{11}}{\sqrt{P_{aa}P_{bb}(1-P_{aa})(1-P_{bb})}} \\
r_{AA,bb} =& \frac{\sigma_A^2\sigma_B^2\eta_{11}}{\sqrt{P_{AA}P_{bb}(1-P_{AA})(1-P_{bb})}} \\
r_{aa,BB} =& \frac{\sigma_A^2\sigma_B^2\eta_{11}}{\sqrt{P_{aa}P_{BB}(1-P_{aa})(1-P_{BB})}}
\end{aligned}
\tag{4}
$$

2.1. Using identity disequilibrium to estimate selfing and measure inbreeding

One of the first practical uses for identity disequilibrium was given by (David et. al., 2007). Using previous results from (Cockerham & Weir, 1973b) that showed that identity disequilibrium could exist at unlinked loci when the organism was capable of self-fertilization, the authors of (David et. al., 2007) devised a metric g_2 that used the double heterozygosity frequency across many pairs of loci to approximate η_{11}. Of importance was the fact

that this was applicable only to those organisms who could self-fertilize: mostly plants but also some animals such as the freshwater snail *Bulinus truncatus*. Double heterozygosity frequencies could then be compared to heterozygosity frequencies at individual loci to estimate s, the percent of the population that was selfing. While g_2 is both useful and well-characterized statistically, subsequent studies often departed from the selfing requirement and estimated g_2 to determine if identity disequilibrium, and thus inbreeding, existed in populations where no selfing occurred. A meta-analysis of many of these studies (Miller & Coltman, 2014) found that g_2 estimates were often not significantly different from zero. This is an issue not due to the number of markers, as some suspected, but rather the assumptions that unlinked loci can easily estimate identity disequilibrium in populations with no selfing.

As stated earlier, for unlinked loci $\eta_{11} = 0$ for pedigree mating in non-selfing populations. It is possible though for average values of $F_{11} - F^2$ to exceed zero for unlinked loci where there is random mating in finite inbred populations. This was demonstrated in (Cockerham & Weir, 1969) (see discussion) where they showed that in finite populations where the population averages of the descent coefficient and inbreeding coefficient are given by $\overline{F_{11}}$ and $\overline{F}$

$$\eta_{11} = \overline{F_{11}} - \overline{F}^2 = \eta'_{11}(\lambda) + \sigma_F^2 \tag{5}$$

The variable $\eta'_{11}(\lambda)$ is the former definition of identity disequilibrium based on pedigree mating and linkage. The dependent parameter λ is the Schnell measure of linkage $\lambda = 1 - 2c$ (Schnell, 1961) that ranges from $[0, 1]$ as c ranges from $[1/2, 0]$. The variance σ_F^2 represents the variance in the measure of inbreeding across the same locus in the finite population that exists due to finite sampling effects. When the two loci are unlinked, $\lambda = 0$ and $\eta'_{11}(\lambda) = 0$ but the component σ_F^2 remains. However, in (Cockerham & Weir, 1969) it is shown through numerical calculation and plotting that σ_F^2 is normally of a near imperceptible magnitude and thus, η_{11} is still not likely to be much different from zero when the loci are unlinked.

Therefore it is important to state that g_2, calculated amongst mostly unlinked loci, is not useful in determining inbreeding in populations where no selfing occurs. In (Cockerham & Weir, 1973b) (first equation on page 259), it is indicated that for unlinked loci, $\eta_{11} = 0$ when $s = 0$. There are, however, alternate ways that double heterozygosity can be used to infer the presence of inbreeding.

2.1.1. g_2 vs. linkage regression

Given that $\eta_{11} \approx 0$ for unlinked loci and $\eta_{11} = F(1-F)$ for complete linkage, the measurement of inbreeding using identity disequilibrium measures sure as g_2 must account for linkage in non-selfing populations. This requires the use of loci that have linkage and will likely require a more dense set of markers in the genomes under analysis, with many markers occupying the same chromosomes required to analyze the changes with linkage. There should be a positive dependence of g_2 on linkage, measured for example by regression, if inbreeding is present. This is complicated though by the effects of linkage disequilibrium that can be larger at linked loci. Correcting this will be addressed later in the paper.

2.2. Double homozygosity and double heterozygosity from first cousin matings

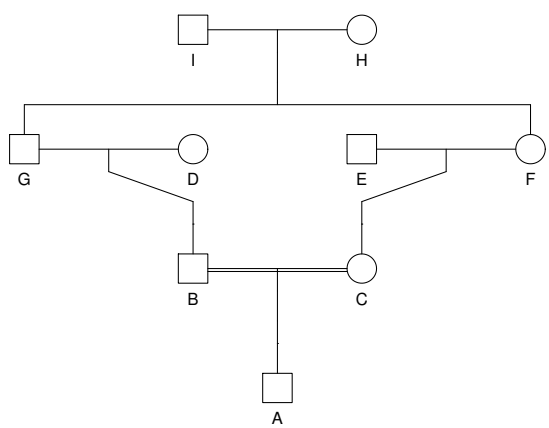


Figure 1: Pedigree of first cousin inbreeding.

As discussed earlier, rare genetic disorders will cluster more in families or populations where consanguineous matings are more common. Consanguineous matings can be a significant portion of the population, even in countries where culture and legal environments mitigate against it. For example, using UK Biobank data (Yengo et. al., 2019) showed that about 1 out of every 3,652 individuals showed extreme inbreeding as measured by

runs of homozygosity. Here we will show the effect of first cousin matings on the expected occurrence of double homozygous and double heterozygous genotypes. This is an important angle where the medical genetics of rare disorders is concerned.

First derived in (Haldane, 1949), but also derived in Appendix A, the two locus descent coefficient F_{11} for the progeny of first cousins is

$$F_{11} = \frac{1}{16}\Big((1-c)^6 + c^2(1-c)^4 + \frac{1}{2}c^2(1-c)^2\Big) \tag{6}$$

Given that $F = 1/16$ for first cousins as well, the identity disequilibrium is given by

$$\eta_{11} = \frac{1}{16}\Big((1-c)^6 + c^2(1-c)^4 + \frac{1}{2}c^2(1-c)^2\Big) - \frac{1}{256} \tag{7}$$

It is clear when $c = 1/2$, $\eta_{11} = 0$ and likewise when $c = 0$, $\eta_{11} = F(1-F) = 15/256$. Per Table 1 when the loci are unlinked the first cousin digenic genotypes are the products of the genotypes at each locus under inbreeding. With linkage, however, the frequencies of double homozygous and double heterozygous genotypes begin to increase. This is particularly important for double homozygous genotypes since both the single and dual locus genotypes increase under inbreeding. Double heterozygote genotypes increase with linkage, however, the effect of identity disequilibrium is to reduce the rate of double heterozygosity decline due to the decline of heterozygosity at single loci and thus is more limited.

In Figure 2, the increase in double homozygotes and double heterozygotes are shown for increasing linkage where the minor allele frequencies are 0.01 or 0.1 in each case. The x-axis uses λ in order to show increasing linkage using an increasingly positive variable.

It is clear that the double heterozygous genotypes show little increase with linkage while double homozygous genotypes increase by over an order of magnitude with increasing linkage when the minor allele as a frequency of 0.01. Without inbreeding, the double homozygous frequency would only be 1×10^{-8} or 1 in 100 million. For first cousins this increases to 5.2×10^{-7} (1 in 1.9 million) when the loci are unlinked. For full linkage, however,this increases to 6.3×10^{-6} about 0.63 cases per 100,000 first cousin births. Granted these probabilities are calculated across the population of progeny of full cousins only but are still a remarkable increase in likelihood for such rare disorders.

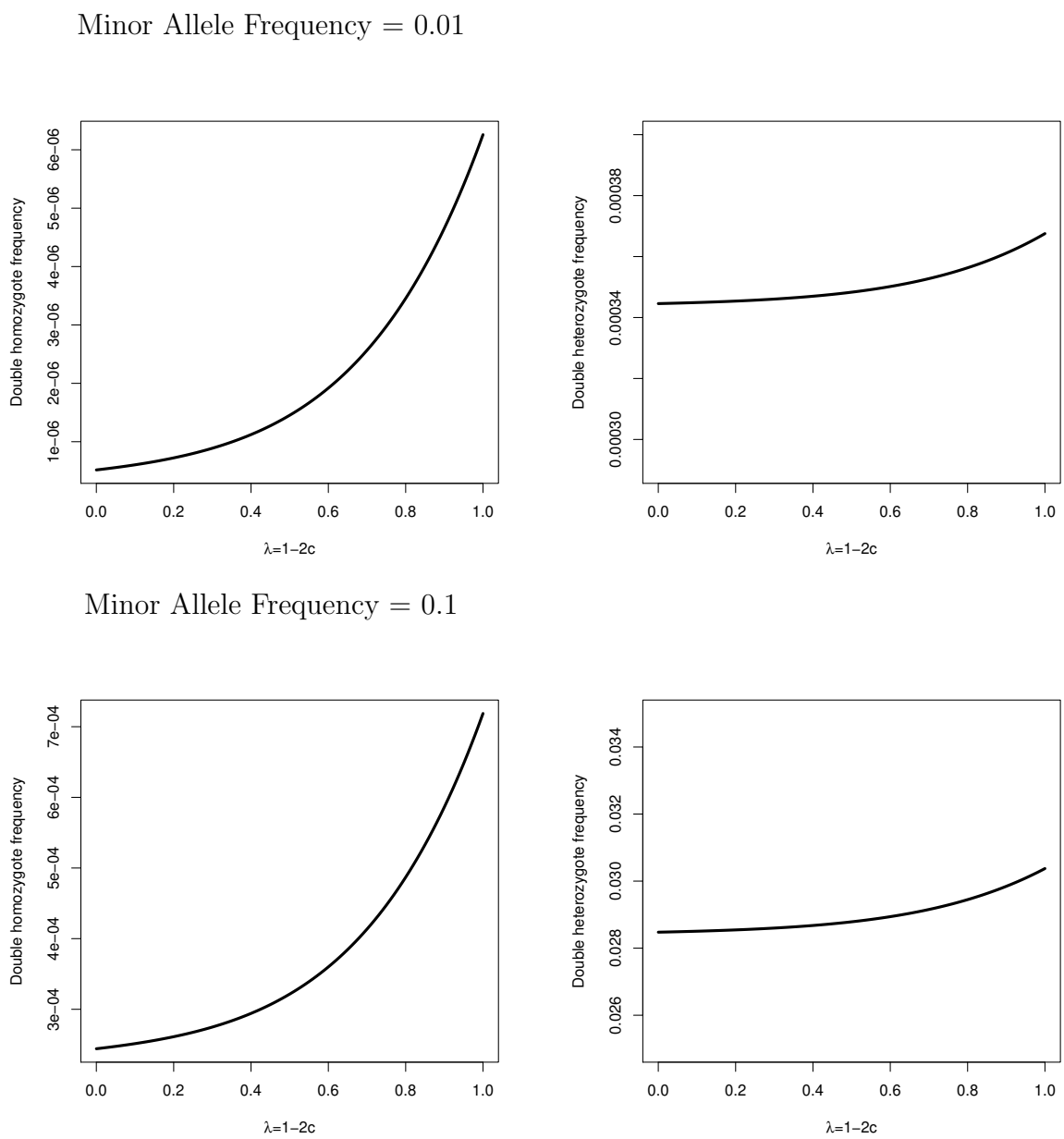


Figure 2: Plots of the frequencies of double homozygous and double heterozygous genotypes where the minor allele q at both loci is 0.01 (top graphs) and 0.1 (bottom graphs) versus linkage designated by the Schnell linkage value $\lambda = 1 - 2c$.

This has implications for the populations in which digenic genotypes, especially those of medical interest occur. Double heterozygous genotypes do not see a significant increase frequency with linkage and the heterozygosity at the individual loci is also reduced by inbreeding. Therefore, autosomal dominant transmitted disorders of digenic character are likely more prevalent in the general population than consanguineous families, regardless of linkage. In the review of then known oligogenic disorders (2+ loci) (Schäffer, 2013), 48 of the 95 then published digenic maladies were transmitted in an autosomal dominant manner at both loci. In almost all of the individual reports, the genotypes reported are double heterozygous and most occurred in unrelated families and in only 13 of the disorders where the loci linked.

In contrast, double homozygous genotypes see large increases in frequency due to linkage and consanguineous matings. Of the 8 disorders recognized as having an autosomal recessive transmission at both loci in (Schäffer, 2013), 3 of 8 had some linkage and 5 of 8 looked at cases where the parents were consanguineous. This emphasizes the numerical results above that for rare variants at two loci, double homozygous genotypes are extremely rare in the absence of inbreeding and made more frequent by linkage.

3. The effects of inbreeding on two loci in linkage disequilibrium

Having discussed the frequencies of digenic genotypes when both loci are in linkage equilibrium, we can now address the more complex situation when the loci are in linkage disequilibrium and the genotype distributions reflect the effects of inbreeding. While linkage disequilibrium is ubiquitous, certain demographic processes make it more widespread in populations. In particular, population admixture and population bottlenecks significantly increase the linkage disequilibrium amongst loci over large sections of the genome.

3.1. Demographic process that generate linkage disequilibrium

One of the key aspects of digenic traits separating them from monogenic ones is the importance of both linkage and linkage disequilibrium. The frequencies of digenic genotypes and the effects of mating systems on those frequencies are heavily dependent on these variables. Given this fact, populations that have large linkage disequilibrium should be expected to have impacts on the frequencies of digenic traits, especially when mating between relatives occurs.

It has been long established that many demographic processes can generate or reduce linkage disequilibrium. Population admixture and population bottlenecks can generate linkage disequilibrium amongst loci while rapid population growth can reduce it (Slatkin, 1994). Linkage disequilibrium creation will be briefly reviewed for these two processes in the next section. Here linkage disequilibrium D and it's normalized version, D' are defined according to the standard notations for alleles A and B and different loci.

$$D = P_{AB} - P_A P_B = P_{AB} P_{ab} - P_{aB} P_{Ab} \tag{8}$$

For D' where $D > 0$

$$D' = \frac{D}{\min\left[P_A(1-P_B), P_B(1-P_A)\right]} \tag{9}$$

and where $D < 0$

$$D' = \frac{D}{\max\left[-P_A P_B, -(1-P_B)(1-P_A)\right]} \tag{10}$$

3.1.1. Linkage disequilibrium from population admixture

Admixture between populations with different allele frequencies at a pair of loci will generate linkage disequilibrium between these loci, even if it is absent in the original populations. Though this linkage disequilibrium decays over time unless maintained by evolutionary forces, it can last for a substantial number of generations in cases where the loci in question are linked and encounter recombination with much lesser frequency.

The amount of linkage disequilibrium generated by admixture was first worked out by (Nei & Li, 1973). For the case of bi-allelic loci where population one contributes a proportion m_1 to the final population this leads to

$$D = m_1 D_1 + (1-m_1) D_2 + m_1(1-m_1)(p_{A1} - p_{A2})(p_{B1} - p_{B2}) \tag{11}$$

The D_1 and D_2 terms represent initial linkage disequilibrium in the original populations and the probabilities are the frequencies of allele A and B in population 1 or 2. This linkage disequilibrium affects haplotype frequencies which in turn help determine digenic trait frequencies.

3.1.2. Linkage disequilibrium from population bottlenecks

A bottleneck is generally defined as a drastic reduction in effective population size. This reduction in population size allows drift to become a much stronger force leading to a reduction in genetic variation, increase in homozygosity, and linkage disequilibrium that is generated from the effects of drift and balanced by recombination (Hill & Robertson, 1968). As long as the population remains small, the linkage disequilibrium can be maintained since it is continuously generated by drift. This is in contrast to linkage disequilibrium by admixture which is a one-time even unless the admixture is continuous. However, as the population grows, the linkage disequilibrium will decline as the effects of drift weaken and recombination becomes dominant (Slatkin, 1994).

3.2. The effects of linkage disequilibrium on digenic genotype frequencies

Linkage disequilibrium has long been known to generate genotypes with an excess of homozygosity for two locus genotypes. Some measures of linkage disequilibrium explicitly take this homozygosity into account to measure linkage disequilibrium across groups of multiple loci (Sabatti & Risch, 2002; Rosenberg & Blum, 2007). In the next sections, we will outline the separate effects on digenic genotype frequencies due to linkage disequilibrium and inbreeding. First we will look at the basic effects of linkage disequilibrium on two locus genotype frequencies in outbred populations. Second we will introduce new descent coefficients required to understand the effect of linkage disequilibrium at two loci when inbreeding occurs followed by detailed calculations of genotype frequencies. Finally, we will revisit the first cousin mating example and discuss the measurement of linkage disequilibrium when Hardy-Weinberg equilibrium is not present.

3.2.1. Digenic genotypes in outbred populations

When the population is outbred, the effect of linkage disequilibrium on genotype frequencies is clear as shown in Table 2. For positive linkage disequilibrium between loci, the frequencies of some genotypes can increase markedly. This alone can increase the frequency of digenic genotypes in recently admixed populations.

For example, for two loci with minor allele frequencies of 0.01, the expected double homozygous genotype frequency for the minor alleles is 1×10^{-8}. With linkage disequilibrium equivalent to $D' = 0.25$, however, this

	AA	Aa	aa
BB	P_{AB}^2	$2P_{AB}P_{aB}$	P_{aB}^2
Bb	$2P_{AB}P_{Ab}$	$2P_{AB}P_{ab} + 2P_{Ab}P_{aB}$	$2P_{aB}P_{ab}$
bb	P_{Ab}^2	$2P_{Ab}P_{ab}$	P_{ab}^2

Table 2: Two loci genotype frequencies as a function of linkage disequilibrium absent inbreeding. See also (Lewontin & Kojima, 1960; Weir, 2008). Chart adopted from (Weir, 2008). $P_{AB} = p_A p_B + D$, $P_{ab} = (1 - p_A)(1 - p_B) + D$, $P_{Ab} = p_A(1 - p_B) - D$, $P_{aB} = (1 - p_A)p_B - D$. Note the double heterozygote ($AaBb$) genotype has two terms due to the genotype arising from either the dual gametes AB and ab or Ab and aB.

can increase to 6.6×10^{-6}, over a 600-fold increase though still relatively uncommon.

3.2.2. The combined impacts of inbreeding and linkage disequilibrium

The sole addition of identity disequilibrium to the genotype frequencies described in Table 1 assumes that the loci are in linkage equilibrium. This greatly simplifies the analysis. The presence of linkage disequilibrium changes frequencies due to the fact initial linkage disequilibrium interacts with linkage to change the expected frequencies of alleles being IBD within loci. It also means we have to account for several additional descent coefficients.

Linkage disequilibrium greatly increases the complexity of the analysis to an extent that cannot be covered fully in this paper. For detailed descriptions on the coefficients and their relationships in pedigree analysis, the author refers readers to (Cockerham & Weir, 1973, 1977). Their techniques for calculating coefficients with known pedigrees is also given in Appendix A. A key result though is linkage disequilibrium combined with inbreeding increases the frequency of double homozygous and double heterozygous genotypes in an additive fashion. In other words, there is an additive increase in the covariance between genotypes under inbreeding when linkage disequilibrium is present.

The first new coefficient, the parental descent coefficient F^{11}, measures the probability that both of the gametes are identical to the gamete in the common ancestor both alleles descended from. In other words, F^{11} measures the probability that both gametes in each individual are passed down from the common ancestor without recombination.

Assuming $\equiv$ means the alleles are from the same ancestral gamete, we can define the two locus parental coefficients similar to the inbreeding coefficients.

$$
\begin{aligned}
F^{11} &= P(a_1 \equiv b_1) \text{ and } P(a_2 \equiv b_2) \\
F^{10} &= P(a_1 \equiv b_1) \text{ and } P(a_2 \not\equiv b_2) \\
F^{01} &= P(a_1 \not\equiv b_1) \text{ and } P(a_2 \equiv b_2) \\
F^{00} &= P(a_1 \not\equiv b_1) \text{ and } P(a_2 \not\equiv b_2)
\end{aligned}
\tag{12}
$$

The average parental coefficient across both individuals can be defined similarly to the inbreeding coefficient case and is designated as F^1. The coefficient F^{11} is graphically represented in Figure 3.

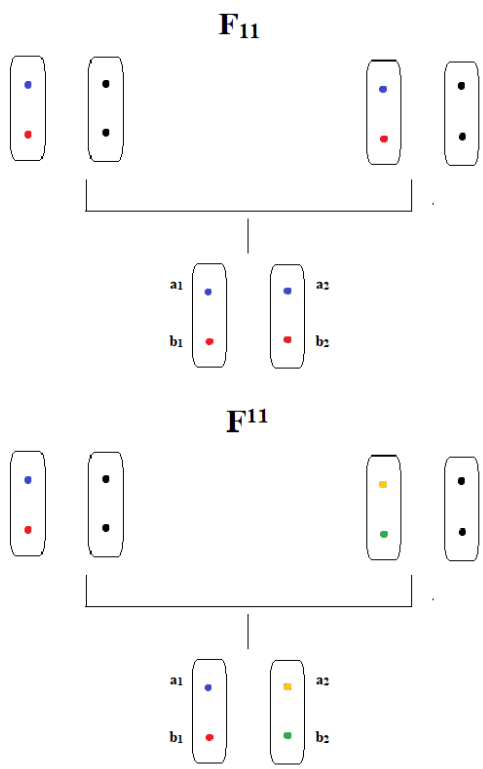


Figure 3: Images representing F_{11} and F^{11} based on ancestral chromosome haplotypes and the haplotypes of descendants. All identically colored non-black colored alleles are identical by descent. For all diagrams, the two pairs of chromosomes are arbitrary paternal and maternal ancestors (which can be one individual sharing both chromosomes) and the bottom chromosome pair is a single descendant.

The second new coefficient, the recombination descent coefficient ${}_{11}F$ is a measure of the probability that an allele at the first locus on one gamete in one

individual and an allele at the second locus on the other gamete in the same individual were originally part of the same gamete in the common ancestor. This measures the probability that there was a recombination event in one or both of the common parents and the alleles in the final generation are on different gametes than they started out. For more details see (Cockerham & Weir, 1973, 1977).

Similar to the definitions above for the parental coefficients

$$\begin{aligned} {}_{11}F &= P(a_1 \equiv b_2) \text{ and } P(a_2 \equiv b_1) \\ {}_{10}F &= P(a_1 \equiv b_2) \text{ and } P(a_2 \not\equiv b_1) \\ {}_{01}F &= P(a_1 \not\equiv b_2) \text{ and } P(a_2 \equiv b_1) \\ {}_{00}F &= P(a_1 \not\equiv b_2) \text{ and } P(a_2 \not\equiv b_1) \end{aligned} \tag{13}$$

The $\equiv$ symbol here indicates the two alleles at different loci on different gametes were originally together on the common ancestral gamete. These descent coefficients are necessary since descent identity no longer relies on alleles at a single locus but the shared gametic ancestry (or lack thereof) of the alleles on the gametes inherited by descendants. The coefficient 11_F is graphically represented in Figure 4.

The differential effects of linkage combined with linkage disequilibrium on digenic genotype frequencies creates a condition, not present in analyses of single loci, where populations with linkage disequilibrium between loci can have greater or lesser frequencies of digenic genotypes than similarly inbred populations where loci are at linkage equilibrium.

3.2.3. Calculations of the linkage disequilibrium impact under inbreeding

We will analyze the effect of linkage disequilibrium on digenic double homozygous and double heterozygous genotypes. The resultant expressions for the effect of inbreeding with linkage disequilibrium will be described below with the reader referred to (Cockerham & Weir, 1973) for detailed derivation and explanation of the source equation which involves two, three, and four gamete linkage disequilibrium effects. The value of D is the value of linkage disequilibrium in the nearest common ancestor of the related parents of the individual.

In contrast to the previous examples of linkage disequilibrium or inbreeding alone where two locus genotypes are described by the two locus genotypes notation (i.e. $AaBb$) we will describe genotypes for the combined case based on the two gametes which generate the genotype since this will allow us

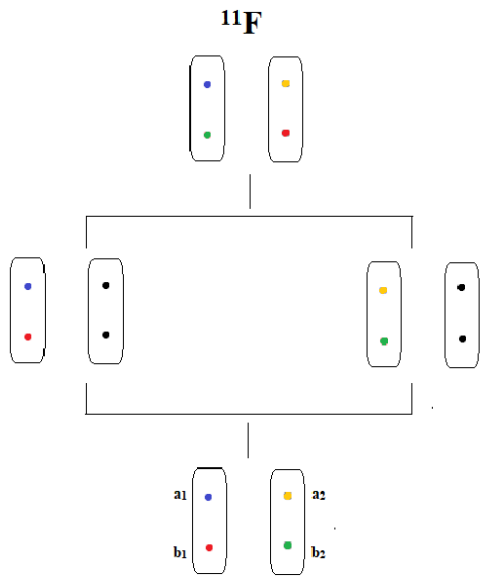


Figure 4: Image representing $_{11}F$ based on ancestral chromosome haplotypes and the haplotypes of descendants. All identically colored non-black colored alleles are identical by descent. For all diagrams, the two pairs of chromosomes are arbitrary paternal and maternal ancestors (which can be one individual sharing both chromosomes) and the bottom chromosome pair is a single descendant. For $_{11}F$, the top ancestor is a common ancestor of all descended individuals and experiences at least two recombination events to pass a chromosome to a paternal and maternal ancestor.

to show equivalency between all situations for a pair of gametes and will also allow the reader to follow equivalent examples in (Cockerham & Weir, 1973, 1977). The double homozygous genotypes thus will be represented as $P(ab|ab)$, where a and b represent the (minor) alleles at their respective loci and the vertical line separates the two gametes. The double heterozygous genotypes are of four types depending on the composition of the gametes and which of the two chromosomes they are located on. $P(AB|ab) = P(ab|AB)$ and $P(Ab|aB) = P(aB|Ab)$. Here $p_A = 1 - p_a$ and $p_B = 1 - p_b$. Where linkage disequilibrium is not present, the genotypes for each will be described with an asterisk and are equal to the below

$$
\begin{aligned}
P(ab|ab)^* =& (p_a^2 + p_A(1-p_A)F)(p_b^2 + p_B(1-p_B)F) + p_A(1-p_A)p_B(1-p_B)\eta_{11} \\
P(AB|ab)^* =& P(ab|AB)^* = P(Ab|aB)^* = P(aB|Ab)^* \\
=& p_A(1-p_A)p_B(1-p_B)(1-F)^2 + p_A(1-p_A)p_B(1-p_B)\eta_{11}
\end{aligned} \tag{14}
$$

The additional changes in digenic genotype frequency caused by linkage disequilibrium involve F^1, ${}_1F$ and related coefficients. Fortunately, the additive adjustments to the double homozygote and double heterozygote genotype frequencies are very similar (see page 311 in (Cockerham & Weir, 1973) for details).

$$
\begin{aligned}
P(ab|ab) =& P(ab|ab)^* + D^2\Big[F^{11} + {}_{11}F - 2F_{11}^{11}\Big] \\
&+ D\Big[2(F^1 + {}_1F)p_a p_b \\
&+ F_{11}^{11}(1-2p_a)(1-2p_b) + 2{}_1F_1^1(p_a + p_b - 4p_a p_b)\Big] \\
P(AB|ab) =& P(ab|AB) = P(AB|ab)^* + D^2\Big[F^{11} + {}_{11}F - 2F_{11}^{11}\Big] \\
&+ D\Big[F^1(p_A p_B + (1-p_A)(1-p_B)) - {}_1F(p_A(1-p_B) + \\
&(1-p_A)p_B) - 2{}_1F_1^1(1-2p_A)(1-2p_B) \\
&+ F_{11}^{11}(p_A p_B + (1-p_A)(1-p_B) - p_A(1-p_B) - (1-p_A)p_B)\Big] \\
P(Ab|aB) = P(aB|Ab) =& P(Ab|aB)^* + D^2\Big[F^{11} + {}_{11}F - 2F_{11}^{11}\Big] + D\Big[{}_1F(p_A p_B + \\
&(1-p_A)(1-p_B)) - F^1(p_A(1-p_B) + p_A(1-p_B)) \\
&- 2{}_1F_1^1(1-2p_A)(1-2p_B) \\
&+ F_{11}^{11}(p_A p_B + (1-p_A)(1-p_B) - p_A(1-p_B) - (1-p_A)p_B)\Big]
\end{aligned} \tag{15}
$$

The two new coefficients above are F_{11}^{11} and ${}_1F_1^1$. The first represents the probability that alleles at both loci are IBD as well as that alleles on both gametes came from the original ancestral gamete. The second is the average of the four probabilities of the joining occurrence of: one locus having IBD alleles, one gamete being inherited intact from an ancestor, and two alleles from different gametes having been on the same gamete in an ancestral generation. Their expressions can be defined as

$$
\begin{aligned}
F^{11}_{11} =& P(a_1 \equiv a_2 \equiv b_1 \equiv b_2) \\
{}_1F^1_1 =& \frac{1}{4}[P(a_1 \equiv a_2 \equiv b_1) + P(a_1 \equiv a_2 \equiv b_2) + P(a_1 \equiv b_1 \equiv b_2) + P(a_2 \equiv b_1 \equiv b_2)]
\end{aligned} \tag{16}
$$

Again $\equiv$ designates alleles from the same initial gamete (Weir & Cockerham, 1974)

The total double heterozygote frequency is the sum of all four separate double heterozygote variations

$$
\begin{aligned}
P(AaBb) =& 2P(AB|ab) + 2P(Ab|aB) = 4P(AB|ab)^* + 4D^2\Big[F^{11} + {}_{11}F - 2F^{11}_{11}\Big] \\
&+ 2D\Big[(F^1 + {}_1F)(p_Ap_B + (1-p_A)(1-p_B) - p_A(1-p_B) \\
&- (1-p_A)p_B) - 4{}_1F^1_1(1-2p_A)(1-2p_B) \\
&+ 2F^{11}_{11}(p_Ap_B + (1-p_A)(1-p_B) - p_A(1-p_B) - (1-p_A)p_B)\Big]
\end{aligned} \tag{17}
$$

Equations 15 and 17 can take a variety of values based on the value of D, allele frequencies, as well as the relative size of the identity coefficients. It does demonstrate that linkage disequilibrium can alter the genotype frequencies where inbreeding is involved. Two main questions however are, first how does inbreeding affect genotypic expectations in populations with linkage disequilibrium versus the case where there is linkage disequilibrium with no inbreeding and second, how does linkage disequilibrium affect genotype expectations compared to inbreeding at linkage equilibrium.

The first question is addressed by the fact that when there is no inbreeding, the genotype frequencies above reduce to those expected in Table 2. Under these conditions $F^1 = F^{11} = 1$ and ${}_1F^1_1 = {}_1F = {}_{11}F = F^{11}_{11} = 0$ (Cockerham & Weir, 1973, 1977). These then give the genotype frequencies as expected from Table 2.

To address the second question, we look at the overall change from the first term with the product of the genotypes at each locus and the term with the identity disequilibrium multiplied the variances of both loci. What is clear is that all terms multiplied by D are positive for double homozygotes and there is only one relatively minor negative term for double heterozygotes. Thus these terms change the genotype frequency proportional to the magnitude and sign of the linkage disequilibrium. The term multiplied by

D^2 is also almost always positive given F_{11}^{11} is usually very small and negligible except when the loci are tightly linked. Thus, compared to the case with inbreeding but no linkage disequilibrium, positive and negative linkage disequilibrium in the ancestor increase or decrease the genotype frequencies of double homozygotes and double heterozygotes, respectively.

The increases in genotype frequencies are obviously also associated with increases in the correlations between single loci genotypes as well. First, we will revisit the first cousin mating example explored previously to incorporate the effects of linkage disequilibrium and second we will introduce the idea of composite linkage disequilibrium which applies to linkage disequilibrium measures when Hardy-Weinberg equilibrium is not present.

3.3. First cousin inbreeding example

Here we will revisit the first cousin pedigree example and look at the combined effects of linkage disequilibrium and inbreeding at two loci with rare minor allele frequencies of 0.01. As will be demonstrated, linkage disequilibrium significantly increases the frequencies of both double homozygous and double heterozygous genotypes though, as before, the double homozygous genotypes receive the most substantial effect.

The process to derive the descent coefficients in the case of first cousin mating will be outlined in Appendix A. The results for our purposes are summarized below:

$$
\begin{aligned}
F_1 &= \frac{1}{16} \\
{}_1F &= \frac{1}{16} \\
F^1 &= (1-c)^3 \\
F_{11} &= \frac{1}{16}\Big((1-c)^6 + c^2(1-c)^4 + \frac{1}{2}c^2(1-c)^2\Big) \\
F^{11} &= (1-c)^6 \\
{}_{11}F &= \frac{1}{16}((1-c)^4 - 2c(1-c)^4 + 2c^2(1-c)^4) \\
F_{11}^{11} &= \frac{1}{16}(1-c)^6 \\
{}_1F_1^1 &= \frac{1}{32}(2(1-c)^5 + c^2(1-c)^2)
\end{aligned}
\tag{18}
$$

These descent coefficients can then be used to calculate the double homozygous and double heterozygous genotype frequencies per equations 15 and 17. The two key variables affecting the genotype frequency are both the degree of linkage between the loci and the magnitude of the linkage disequilibrium. For the two rare alleles analyzed, D has a range of $[-0.0001, 0.0099]$ and c ranges between zero and one-half.

In Figure 5 are graphs of the frequency of double homozygotes and double heterozygotes for various values of D and c. The plot gives linkage on the x-axis as Schnell's linkage value.

Given a minor allele of interest has a population frequency of 0.01 we once again consider the double homozygous and double heterozygous frequencies. As stated in the previous cousin mating example, the double homozygous genotype would occur one case in 100 million persons with no inbreeding or linkage disequilibrium. Under normal Hardy-Weinberg equilibrium with no inbreeding and linkage equilibrium between the two loci, the double heterozygote, the largest frequency genotype for digenic autosomal dominant disorders, would have a prevalence of 39 cases per 100,000.

These incidence rates change markedly with the introduction of inbreeding, linkage disequilibrium, or both. For first cousin inbreeding where $F = 1/16$, at two loci which are unlinked and not in linkage disequilibrium, double homozygosity rising to 1 in 1.92 million, a 50-fold increase, and double heterozygosity actually reducing to 34 cases per 100,000. These effects are solely due to the combined effects of single locus inbreeding genotype changes at each locus. For complete linkage, this increases to 0.63 per 100,000 for double homozygotes as stated previously and 37 per 100,000 for double heterozygotes.

The introduction of linkage disequilibrium, however, removes the ambiguous effect of inbreeding on double heterozygotes. For linkage disequilibrium of $D' = 0.25$ without inbreeding for unlinked loci, double homozygotes have an incidence of 1.7 per 10 million and double heterozygotes a rate of 99 per 100,000. For $c = 0$, double homozygotes increase to 0.66 per 100,000 with double heterozygotes rising to 520 per 100,000. Under inbreeding at $F = 1/16$ and $D' = 0.25$, the rates for unlinked loci are 0.34 per 100,000 for double homozygotes (a 20X increase over the identical outbred scenario) and 117 per 100,000 for double heterozygotes. The rate for double heterozygotes is now higher than the outbred case for the same amount of linkage disequilibrium. This rises to 17 per 100,000 and 484 per 100,000 (about 1 per 206) for double homozygotes and double heterozygotes respectively for completely

linked loci where double heterozygotes under inbreeding see their frequency gap with the outbred counterparts disappear and even slightly reverse.

Therefore inbreeding and linkage disequilibrium can combine to significantly increase the digenic genotypes for consanguineous matings. Again, the increase is more marked for double homozygous genotypes than for double heterozygotes except for near complete linkage disequilibrium. Therefore, due to linkage disequilibrium, population admixture and recent bottlenecks in populations that have not yet undergone exponential growth will possibly show a significantly higher frequency of rare digenic genotypes than allele frequencies at individual loci would initially suggest. Consanguineous matings in these populations would be more likely to experience the deleterious effects of digenic disorders as well.

While some of the double homozygous disorders from (Schäffer, 2013) occur in non-European populations such as those of Pakistan (Basit et. al., 2011; Riazuddin et. al., 2000) and Morocco (Ebermann et. al., 2007), there are no reports from admixed populations such as those from the Americas about new digenic disorders. This may be due to coverage and attention more than an actual reflection on the distribution of such disorders. More research on rare and unsolved disorders in admixed or recent bottleneck populations may reveal that some of these are oligogenic.

4. Composite linkage disequilibrium when inbreeding is present

One finding mentioned in (Weir & Cockerham, 1974) (see page 312 for more details) and expanded later by (Weir, 1979; Weir & Cockerham, 1989) is the idea of composite linkage disequilibrium. Namely, traditional linkage disequilibrium calculations like Equation 8 imply the population is in Hardy-Weinberg equilibrium. Departures from this, such as in mating systems, guarantees that the frequencies of the two types of double heterozygotes, namely attractive $P(AB|ab)$ and repulsive $P(Ab|aB)$ reflect not only correlations between alleles on the same gamete but also correlations between alleles on different gametes. For loci with only two alleles, the frequency difference between the two types of double heterozygosity when there is inbreeding can be seen from Equation 15 and is given by

Inbreeding	Ancestral Linkage Disequilibrium	Linkage	P(aabb)	P(AaBb)
$F = 0$	$D' = 0$	Any	1.0×10^{-8}	3.9×10^{-4}
	$D' = 0.25$	$c = 1/2$	1.7×10^{-7}	9.9×10^{-4}
		$c = 1/4$	1.3×10^{-6}	2.4×10^{-3}
		$c = 0$	6.6×10^{-6}	5.2×10^{-3}
	$D' = 1$	$c = 1/2$	1.8×10^{-6}	2.8×10^{-3}
		$c = 1/4$	1.8×10^{-5}	8.3×10^{-3}
		$c = 0$	1.0×10^{-4}	0.02
$F = \frac{1}{16}$	$D' = 0$	$c = 1/2$	5.2×10^{-7}	3.4×10^{-4}
		$c = 1/4$	1.4×10^{-6}	3.5×10^{-4}
		$c = 0$	6.3×10^{-6}	3.7×10^{-4}
	$D' = 0.25$	$c = 1/2$	3.4×10^{-6}	1.2×10^{-3}
		$c = 1/4$	3.1×10^{-5}	2.5×10^{-3}
		$c = 0$	1.7×10^{-4}	4.8×10^{-3}
	$D' = 1$	$c = 1/2$	1.3×10^{-5}	3.7×10^{-3}
		$c = 1/4$	1.3×10^{-4}	8.8×10^{-3}
		$c = 0$	7.2×10^{-4}	0.018

Table 3: Expected double homozygote and double heterozygote frequencies under various combinations of ancestral linkage disequilibrium and inbreeding for $p_a = p_b = 0.01$ and consanguineous mating between first cousins determining F and the descent coefficients for the inbred case. Note that the final linkage disequilibrium in the affected population has undergone three generations of recombination from the ancestral linkage disequilibrium and in each case is $D'(1-c)^3$.

$$
\begin{aligned}
2P(AB|ab) - 2P(Ab|aB) =& 2D\Big[(F^1 - {}_1F)(p_A p_B + (1-p_A)(1-p_B) \\
& + p_A(1-p_B) + (1-p_A)p_B)\Big] \\
P(AB|ab) - P(Ab|aB) =& D(F^1 - {}_1F)
\end{aligned} \tag{19}
$$

This is another demonstration on how composite linkage disequilibrium is useful, and even essential, to understand where inbreeding is known or suspected. In inbreeding the ratio amongst the two types of gamete pairs that form double heterozygotes is affected by the additional probability of inheritance of a gamete from a recent common ancestor in addition to the probability of inheriting the gamete as it randomly segregates in the population. Therefore, using their frequency difference to calculate linkage disequilibrium gives biased estimates.

5. Simplified expressions where linkage disequilibrium is present

For similar situations to this example where minor allele frequencies are small and maximum linkage disequilibrium is not large, the genotype frequency expressions can be greatly simplified to only the most familiar variables. While the genotype frequency expressions can be complicated and involve many new variables in the case of linkage disequilibrium, the relatively small contributions of most of the descent coefficients as well as small effect of D^2, especially in the double heterozygote case, can give us simpler approximations that can be useful for estimation. In short, for the double homozygous and double heterozygous genotype frequencies, the approximations can be given by

$$
\begin{aligned}
P(ab|ab) =& P(ab|ab)^* + 2D(F^1 + {}_1F)p_a p_b \\
P(AaBb) =& 4P(AB|ab)^* + 2D(F^1 + {}_1F)(p_A p_B + (1-p_A)(1-p_B) - + p_A(1-p_B) - (1-p_A)p_B)
\end{aligned} \tag{20}
$$

In this simplification, only F_{11} and F need to be known to calculate η_{11} for the base case and for linkage disequilibrium, only the averages F^1 and ${}_1F$. If only the parents of the final progeny are consanguineous (all other ancestors are outbred), and the parents belong to the same generation this can be simplified further to

$$
\begin{aligned}
P(ab|ab) =& P(ab|ab)^* + 2D((1-c)^n + F)p_a p_b \\
P(AaBb) =& 4P(AB|ab)^* + 2D((1-c)^n + F)(p_A p_B + (1-p_A)(1-p_B) \\
& - p_A(1-p_B) - (1-p_A)p_B)
\end{aligned} \tag{21}
$$

The variable n is the number of generations between the progeny and the nearest common ancestor of the consanguineous parents. So $n = 2$ for the progeny of full or half-sibs, $n = 3$ for progeny of first cousins, etc.

For unlinked loci $(1 - c) = 1/2$ so

$$
\begin{aligned}
P(ab|ab) =& P(ab|ab)^* + 2D\left[\left(\frac{1}{2}\right)^n + F\right]p_a p_b \\
P(AaBb) =& 4P(AB|ab)^* + 2D\left[\left(\frac{1}{2}\right)^n + F\right](p_A p_B + (1-p_A)(1-p_B) \\
& - p_A(1-p_B) - (1-p_A)p_B)
\end{aligned} \tag{22}
$$

Finally, for completely linked loci, the D^2 term may become substantial and can be simplified with the below taking into account for completely linked loci $F^{11} = 1$ and $F_{11} = F_{11}^{11} = F$.

$$
\begin{aligned}
P(ab|ab) =& P(ab|ab)^* + D^2(1-F) + 2D(1+F)p_a p_b \\
P(AaBb) =& 4P(AB|ab)^* + 4D^2(1-F) + 2D(1+F)(p_A p_B + (1-p_A)(1-p_B) \\
& - p_A(1-p_B) - (1-p_A)p_B)
\end{aligned} \tag{23}
$$

6. Purifying selection against deleterious digenic genotypes

As was discussed in the introduction, the general theory for the calculation of the effects of selection on two-locus genotypes was established by Lewontin & Kojima (Lewontin & Kojima, 1960). Using the methods from that paper, we will demonstrate how rapidly selection removes deleterious alleles given selection against double homozygous or double heterozygous genotypes. The methods of their paper will be briefly discussed here with only initial and final results shown. For details on derivations the readers are

directed to (Lewontin & Kojima, 1960), particularly pages 462 and 463. To model the effects of selection each two-locus genotype has a relative fitness value shown in Figure 6. In addition, the relative fitness values for the double homozygous and double heterozygous genotypes are shown in Figure 7.

For the two loci, A/a and B/b which epistatically interact, the basic haplotype evolution equations are:

$$\begin{aligned}
\Delta P_{AB} &= \frac{P_{AB}(W_{AB} - \overline{W}) - RDW_{11}}{\overline{W}} \\
\Delta P_{Ab} &= \frac{P_{Ab}(W_{Ab} - \overline{W}) + RDW_{11}}{\overline{W}} \\
\Delta P_{aB} &= \frac{P_{aB}(W_{aB} - \overline{W}) + RDW_{11}}{\overline{W}} \\
\Delta P_{ab} &= \frac{P_{ab}(W_{ab} - \overline{W}) - RDW_{11}}{\overline{W}}
\end{aligned} \tag{24}$$

The average fitness across the population is given by $\overline{W}$. Using the methods from (Lewontin & Kojima, 1960) and relative fitness values of Figure 7 we will now address the effects of selection on first double homozygotes and then double heterozygotes.

6.1. Deleterious double homozygotes

In Figure 7, two selection coefficients s_1 and s_2 are used where s_1 represents the deleterious effects of aa and $s_1 + s_2$ represents the deleterious effects of $aabb$. The coefficient s_1 is included so we can later show how the deleterious digenic genotype not only reduces the frequency of the ab haplotype but also the a allele. The relative fitness of all genotypes not including aa is reduced to W which will be set at 1. The relative fitness of genotypes with aa but not bb, W_{02} and W_{01}, is reduced to $W_1 = 1 - s_1$. The relative fitness of $aabb$ is given by $W_{00} = 1 - s_1 - s_2$.

Following derivations based on these parameters, we find the final haplotype evolution equations. The variable c is the recombination rate and D is the linkage disequilibrium.

$$\begin{aligned}
\Delta P_{aB} &= \frac{P_{aB}^2(W_1 - W) + P_{aB}(W - \overline{W}) + P_{ab}P_{aB}(W_1 - W) + cDW}{\overline{W}} \\
\Delta P_{ab} &= \frac{P_{ab}^2(W_{00} - W) + P_{ab}(W - \overline{W}) + P_{ab}P_{aB}(W_1 - W) - cDW}{\overline{W}}
\end{aligned} \tag{25}$$

Given the case of a deleterious double homozygous genotype where $s_1 = 0$, and thus $W_1 = W = 1$ and $W_{00} = 1 - s_2$ these can be revised as

$$\begin{aligned} \Delta P_{aB} &= \frac{P_{aB}(1-\overline{W}) + cD}{\overline{W}} \\ \Delta P_{ab} &= \frac{-s_2 P_{ab}^2 + P_{ab}(1-\overline{W}) - cD}{\overline{W}} \end{aligned} \tag{26}$$

The average fitness can be calculated as

$$\overline{W} = 1 \times (1 - P_{ab}^2) + (1 - s_2) \times P_{ab}^2 = 1 - s_2 P_{ab}^2 \tag{27}$$

The rate of decrease of the haplotypes given s_2 can thus be shown as

$$\begin{aligned} \Delta P_{aB} &= \frac{s_2 P_{aB} P_{ab}^2 + cD}{1 - s_2 P_{ab}^2} \\ \Delta P_{ab} &= \frac{-s_2 P_{ab}^2 + s_2 P_{ab}^3 - cD}{1 - s_2 P_{ab}^2} \end{aligned} \tag{28}$$

If the haplotype P_{ab} is rare in the population, we can approximate the changes due to selection as

$$\begin{aligned} \Delta P_{aB} &\approx cD \\ \Delta P_{ab} &\approx -s_2 P_{ab}^2 - cD \end{aligned} \tag{29}$$

The equation above shows that the effects of selection are felt differently on each haplotype. Haplotype aB increases in frequency, however, based on the small value of the product $P_{aB}P_{ab}^2$, the effects due to selection are small and the increase is primarily due to the effects of recombination where unlinked loci will provide the fastest rate of increase. While selection has a positive effect, it is minimal and only significant for linked loci or at linkage equilibrium.

The decrease in frequency of ab, however, is due to both the effects of selection and recombination. Similar to the expression for a deleterious recessive allele at a single locus, the impact of selection is proportional to the frequency of $aabb$. However, the two-locus case also includes the effect of recombination and linkage disequilibrium. if the loci are not linked and there is linkage disequilibrium present, the change in haplotype frequency is dominated by recombination until near linkage equilibrium. When linkage

equilibrium is reached or if the loci are tightly linked, the forces of selection gradually continue to reduce the haplotype frequency in the population. Given the small value of P_{ab}^2, the haplotype frequency rate of decrease is slow and the haplotype will be maintained in the population for many generations. Therefore, deleterious double homozygote haplotypes which are rare are likely most commonly existing in the population as tightly linked loci or unlinked loci in linkage equilibrium and can maintain themselves for many generations even if s_2 is large.

6.1.1. The decline in frequency for alleles linked with deleterious double homozygotes

Given equation 25, the rate of decrease of frequency for a is $\Delta P_a = \Delta P_{aB} + \Delta P_{ab}$. Using $W_1 = 1 - s_1$ the evolution equation is

$$\begin{aligned}
\Delta P_a &= \frac{P_a(W - \overline{W}) + 2P_{aB}P_{ab}(W_1 - W) + P_{aB}^2(W_1 - W) + P_{ab}^2(W_{00} - W)}{\overline{W}} \\
\Delta P_a &= \frac{P_a(1 - \overline{W}) - 2s_1P_{aB}P_{ab} - s_1P_{aB}^2 - (s_1 + s_2)P_{ab}^2}{\overline{W}} \\
\Delta P_a &= \frac{P_a(1 - \overline{W}) - s_1P_a^2 - s_2P_{ab}^2}{\overline{W}}
\end{aligned} \tag{30}$$

The average fitness works out to be $\overline{W} = 1 - s_1P_a^2 - s_2P_{ab}^2$. This gives a final form of

$$\Delta P_a = \frac{-(1 - P_a)(s_1P_a^2 + s_2P_{ab}^2)}{1 - s_1P_a^2 - s_2P_{ab}^2} \tag{31}$$

Again using the example from the paper where $s_1 = 0$ this reduces to

$$\Delta P_a = \frac{-(1 - P_a)s_2P_{ab}^2}{1 - s_2P_{ab}^2} \approx (\Delta P_{ab} + cD)(1 - P_a) \tag{32}$$

The decrease in a depends on both the frequency of the haplotype ab as well as the frequency of a. The likely rare frequency of ab demonstrates that the allele a can also be maintained in the population for many generations since the effects of selection are dependent on the infrequent appearance of the $aabb$ genotype.

6.1.2. The decline in frequency for alleles linked with deleterious double heterozygotes

In order to understand the extent of a deleterious double heterozygote, the expressions for all four haplotypes are necessary

$$\begin{aligned}
\Delta P_{AB} &= \frac{P_{AB}P_{ab}(W_{11}-W)+P_{AB}(W-\overline{W})-cDW_{11}}{\overline{W}} \\
\Delta P_{ab} &= \frac{P_{AB}P_{ab}(W_{11}-W)+P_{ab}(W-\overline{W})-cDW_{11}}{\overline{W}} \\
\Delta P_{Ab} &= \frac{P_{Ab}P_{aB}(W_{11}-W)+P_{Ab}(W-\overline{W})+cDW_{11}}{\overline{W}} \\
\Delta P_{aB} &= \frac{P_{Ab}P_{aB}(W_{11}-W)+P_{aB}(W-\overline{W})+cDW_{11}}{\overline{W}}
\end{aligned} \tag{33}$$

Given $W_{11} = 1 - s_1$ and $W = 1$ per Figure 7 the average fitness is $\overline{W} = 1 - s_1 P(AaBb)$ and the equations are modified to

$$\begin{aligned}
\Delta P_{AB} &= \frac{-s_1(P_{AB}P_{ab}(1-c)+P_{Ab}P_{aB}c)+s_1P_{AB}P(AaBb)-cD}{1-s_1P(AaBb)} \\
\Delta P_{ab} &= \frac{-s_1(P_{AB}P_{ab}(1-c)+P_{Ab}P_{aB}c)+s_1P_{ab}P(AaBb)-cD}{1-s_1P(AaBb)} \\
\Delta P_{Ab} &= \frac{-s_1(P_{AB}P_{ab}c+P_{Ab}P_{aB}(1-c))+s_1P_{Ab}P(AaBb)+cD}{1-s_1P(AaBb)} \\
\Delta P_{aB} &= \frac{-s_1(P_{AB}P_{ab}c+P_{Ab}P_{aB}(1-c))+s_1P_{aB}P(AaBb)+cD}{1-s_1P(AaBb)}
\end{aligned} \tag{34}$$

So the decline of the AB and ab haplotype frequencies depends on the frequency of the attractive AB/ab double heterozygotes frequencies while the decline in frequency for the Ab and aB haplotype frequencies depends on the repulsive Ab/aB double heterozygote frequencies. There is also a term to increase the haplotype frequency based on the product of the haplotype frequency and the double heterozygote (all types) frequency. What is interesting is if we look at the fate of any given allele, in this case a. Where again $\Delta P_a = \Delta P_{aB} + \Delta P_{ab}$

$$\Delta P_a = \frac{s_1P(AaBb)(P_a-1/2)}{1-s_1P(AaBb)} \tag{35}$$

The change in the frequency of allele a is expressed as a form of underdominance, similar to the case of underdominance with a deleterious single locus heterozygote, but also depending on the frequency of the double heterozygotes in the population. If $P_a > 1/2$ then a gradually increases with frequency until it becomes fixed. In the more likely case for a deleterious allele, $a < 1/2$ and the frequency of a declines until it is removed from the population.

6.1.3. Comparison of the effects of selection on deleterious double homozygotes and double heterozygotes

Given the derivations above, a key question is whether a deleterious allele decreases in frequency more rapidly when the deleterious genotype is a double homozygote or double heterozygote. This is simulated in Figure 8 where given $a = 0.01$ and $b = 0.01$ the reduction in allele a is shown for the cases of a deleterious double homozygote, double heterozygote, and the ordinary case of a deleterious recessive allele at one locus. This analysis assumes the absence of inbreeding.

In each case, the value of the relevant selection coefficient is the same with $s = 0.3$. The frequency change of a for the single locus homozygous case is given by $\Delta P_a = -(s_1(1-P_a)P_a^2)/(1-s_1P_a^2)$. In the first two cases, one with neither linkage nor linkage disequilibrium and one with complete linkage and initial maximum positive linkage disequilibrium, the deleterious double heterozygote causes a much faster decay in the allele frequency, going quickly to extinction in the case of complete linkage and maximum positive linkage disequilibrium for ab. In the third case of complete linkage and maximum negative linkage disequilibrium, the double heterozygote still decays much faster than the double homozygote but the single locus comparison decays similarly.

Therefore, in general terms one can say when there is little recombination and linkage disequilibrium is strong and positive, the decay of an allele associated with a deleterious double homozygote decays similarly to the single locus homozygous case and when linkage disequilibrium is strong but negative, the double heterozygote decays similarly to the single locus homozygous case. Both cases do not decay exactly like the single locus case since two locus genotype frequencies include the effects of linkage disequilibrium.

The dual implications of these simulations seem clear. First, although double heterozygotes may usually have a relatively higher frequency of appearance in a population, the causative alleles are removed much more rapidly

so that in general, their overall population frequency will still be relatively low. Second, although double homozygous recessive cases are rare outside of inbreeding or tightly linked loci, they can be maintained for many generations in a population since their low frequency of occurrence makes the effects of selection very gradual.

The vast majority of digenic disorders identified are usually autosomal dominant at each locus and are mostly seen as double heterozygotes at loci that are not linked nor in linkage disequilibrium. This raises the question of how these digenic disorders emerge and then subsequently segregate in populations. Their relatively quick decay seems to imply that recent genetic events, such as founder effects or drift in small populations, can establish these alleles at a low frequency and then as they segregate through the population and the deleterious genotype occurs, the causative alleles are relatively rapidly removed from the population. The smaller numbers of double homozygous disorders known is likely due to their low frequency of occurrence that typically requires consanguineous pairings or linkage disequilibrium. However, these can maintain themselves for long periods in populations and may not necessarily have a recent origin.

7. Discussion

The foregoing has primarily been a discussion on traits which show full penetrance with certain genotypes at two loci. These digenic traits can show substantial changes in their frequency in a population given not only the expected effects of linkage disequilibrium or inbreeding but also the complimentary actions of the two along with the unique property of identity disequilibrium.

7.1. Medical genetics implications

In medical genetics there are many disorders where known alleles or chromosome regions discovered by linkage studies can only account for a proportion of cases. This incomplete penetrance can be an argument for necessary environmental factors but it is also often the cause for a search for oligogenic or polygenic inheritance. With the advent of techniques such as whole exome sequencing that allow for a fuller view of possibly pathological variants, more oligogenic traits will possibly be discovered.

Part of the reason for the paucity of known digenic traits may be not only their frequency of occurrence but that past techniques made only unlinked

digenic traits most obvious. In particular, linkage analysis with microsatellite marker loci can obscure the fact that multiple loci in linkage are causing the disorder since only one chromosomal region shows linkage disequilibria between marker loci in affected individuals. Most studies involving markers find digenic traits when they are due to different loci on different chromosomes so two regions are highlighted. Of the ninety-five disorders described in (Schäffer, 2013), only 13 are due to linked loci and three of these are described as weakly linked.

Not only do these older methods likely obscure linked digenic loci but they may also be unintentionally classifying Mendelian disorders that should be oligogenic as suggested earlier. One possible example is North American Indian Childhood Cirrhosis (NAIC)(OMIM, 2016), a serious liver disorder requiring a liver transplant that is limited to the Ojibwe-Cree First Nations population of Quebec. Pedigree analysis and linkage work (Bétard et. al., 2000; Chagnon et. al., 2002) led to a putative locus that was homozygous recessive in only affected children, heterozygous in all parents of affected children, and heterozygous or homozygous dominant in unaffected siblings. It was thus very likely that this locus, whose prevalence was likely a founder effect, was the cause of the disease. However, further work on genomes from other populations showed this homozygous recessive genotype was found in multiple individuals in other populations who had no history of liver disorders (Lek et. al., 2016). It is unclear what the final answer is but another locus, likely in close linkage, is possible given linkage studies highlighted only the chromosome region 16q22.

Additionally, many digenic or oligogenic genotypes that show incomplete penetrance for a disease may themselves just be large effect variants for a disorder that is ultimately polygenic. Deleterious variants with large effect in complex disorders tend to cluster in genes expressed in cells closely related to the system most affected by the disorder, even if they only explain a fraction of genetic variance (Boyle et. al., 2017). Therefore while oligogenic traits may help explain 'Mendelian' disorders with incomplete penetrance, they may not be the end of the story but additional loci of large effect for a complex disease.

For digenic disorders, given the higher genotype frequency of double heterozygotes when alleles are rare, the most common digenic disorders should be those that have autosomal dominant inheritance at each locus and occur most frequently as double heterozygotes. This is indeed what the data bears out. However, given the demonstrations of the effects of selection, the associ-

ated allele frequencies should be low and linked loci for these disorders should be uncommon. This is the case in the data from (Schäffer, 2013) where only 8 of 48 of the disorders that are autosomal dominant at both loci are linked. Granted as discussed before, some linked digenic traits may be undercounted but this low number of deleterious disorders which most commonly appear as double heterozygotes is expected.

Also in line with the results from the simulation on selection, the frequencies of alleles associated with deleterious digenic disorders is low. Using the MIM numbers for all digenic dominant and digenic recessive disorders from OMIM and referencing these in Clinvar for loci that are designated as 'pathogenic' and mention 'digenic' in their description, 96 SNPs with global minor allele frequency data were found. Their average global minor allele frequency was about 0.0002 with many one to two orders of magnitude less frequent. The highest frequency alleles were associated with a digenic form of inherited deafness involving a double heterozygote across the unlinked genes *GJB2* and *GJB3* which had minor allele frequencies of 0.002. Both those disorders caused by double homozygotes and double heterozygotes had alleles that were at low frequencies.

There is one other factor that may reduce the frequency of many digenic trait alleles. For a large number of digenic disorders, the epistasis occurs between loci affected by two different types of variants. Rather than just two SNPs representing minor alleles, often one allele is a SNP while the other is another variant type, usually a deletion, sometimes many kb long. Whether this is a result of the fact that epistasis between two SNPs is often of smaller effect or another reason is beyond the scope of this paper.

The foregoing also gives us insight into the type of epistasis that is most common in medical digenic traits. Epistasis amongst two loci in linkage equilibrium was well categorized by (Kempthorne, 1954; Cockerham, 1954) and comes in three types: additive-additive ($A \times A$), additive-dominant ($A \times D$) (or dominant-additive ($D \times A$)) , and dominant-dominant ($D \times D$). As shown in Figure 9, the type of epistasis depends on the homozygosity or heterozygosity of the two loci involved. Homozygous loci contribute additive epistatic effects while heterozygous loci contribute dominance effects. Therefore the double heterozygous, most common, are essentially $D \times D$ epistasis while double homozygotes are $A \times A$. $A \times D$ can also occur and have been found roughly at the same rate as $A \times A$ disorders in (Schäffer, 2013). Given the higher frequency heterozygous locus in $A \times D$ epistasis one would expect these to occur at a much higher frequency than $A \times A$ epistasis, however, the

effects of linkage disequilibrium on homozygous-heterozygous genotypes can be ambiguous while the effects of identity disequilibria are clearly negative so their frequency is likely reduced due to these effects.

The correlations between relatives with $A \times A$ epistasis of unlinked double homozygotes in linkage equilibrium and homozygotes in Mendelian disorders is also both similar and different. Single locus disorders whose effects are due to dominance have a 1/4 correlation between full siblings as does $A \times A$ epistasis for unlinked loci in linkage equilibrium (Cockerham, 1954). Therefore, these disorders can often be confused with Mendelian autosomal recessive disorders with methods such as intraclass correlation. However, single locus dominance shows no correlation between parents and children while $A \times A$ epistasis demonstrates a 1/4 correlation. This value increases when the loci are linked, maxing out at 3/8 for full linked loci (Cockerham, 1956), not including the effects of linkage disequilibrium.

Double heterozygous digenic traits with $D \times D$ epistasis have only correlation between full sibs similar to single locus disorders but the correlation is much lower at 1/16 and should be easier to differentiate. However, if the loci are fully linked, the correlation increases to 1/4 and again could be confounded with a single locus homozygous disorder.

While the discussion here has mostly focused on deleterious digenic disorders, because they are more well-known and characterized, there is also an active effort to discover gene-gene interactions in association studies (Van Steen, 2012; Moore & Williams, 2016) which could shed even more light on which kinds of epistasis contribute to the genetic variance of complex traits.

In summary the amount of digenic traits of medical interest is possibly much larger than currently known and may significantly overlap with currently known traits of purportedly Mendelian character. This could require new techniques and analysis that use digenic inheritance as an alternative hypothesis before settling on the common one-locus aetiology for disease.

7.2. Measuring inbreeding in populations

The correlation between heterozygosity amongst pairs of loci can be used in a manner similar to the frequency of homozygosity at single loci to infer the presence of inbreeding. Using the heterozygosity correlation allows estimates of identity disequilibrium instead of the common inbreeding coefficient, F, but can definitely be a primary or confirmatory test for inbreeding in a population given the appropriate number of loci. One important point that seems

to have been incorrectly deemphasized in many studies is that linkage significantly magnifies the identity disequilibrium measurements in non-selfing populations. Estimating the identity disequilibrium using unlinked loci, even in large numbers, can often return a result not significantly different from zero even when the population is significantly inbred. Therefore, it is more appropriate to measure heterozygosity correlations amongst linked loci.

However, there is a possible confounding in this process due to linkage disequilibrium especially when the loci are tightly linked or if the population is admixed, small, or recently experienced a bottleneck. While the best case scenario for clear results is linked loci at linkage equilibrium, it may be necessary to adjust calculations removing the effects of linkage disequilibrium to accurately measure identity disequilibrium.

A simplified procedure to adjust the heterozygote frequencies before calculating a metric like g_2 can be suggested by equation 17. Since the value of D^2 is usually low as are the values of the coefficients F^{11}_{11} and ${}_1F^1_1$ the frequency of the heterozygotes can be shown as

$$\begin{aligned} P(AaBb) =& 4P(AB|ab)^* + 2D\Big[(F^1 + {}_1F)(p_A p_B + (1-p_A)(1-p_B) - p_A(1-p_B) \\ & - (1-p_A)p_B)\Big] \end{aligned} \tag{36}$$

While the value of $(F^1 + {}_1F)$ is unknown for a population where a pedigree is not available, it has a maximum of one so a lower bound of the corrected heterozygote frequency to remove the effects of linkage disequilibrium can be given by

$$P(AaBb) - 2D\Big[(p_A p_B + (1-p_A)(1-p_B) - p_A(1-p_B) - (1-p_A)p_B)\Big] \tag{37}$$

These corrected frequencies can be thus used to estimate g_2 which will not give an exact value but still show the presence of inbreeding without overestimating the heterozygosity correlation due to linkage disequilibrium. If only linkage disequilibrium and not inbreeding is present g_2 should be approximately zero or negative among all loci pairs.

8. Conclusion

Far from being a staid and unimportant area of concern, the research of digenic and other oligogenic traits shows many promises to provide new

insights and benefits in fields as diverse as genetic counseling to conservation genetics. The impact of such traits is complicated by both inbreeding and linkage disequilibrium but this paper has attempted to detail these influences as well as provide simplified relations that can be used for analysis. Future research should hopefully be directed at finding oligogenic traits of interest in the studies of evolving populations such as digenic traits under selection as predicted by (Lewontin & Kojima, 1960) and others. This will help understand the variation in populations in both allele frequencies and linkage disequilibria as well as further elucidate the relative importance of epistasis in evolution and adaptation.

Appendix A. Descent coefficient derivation

This section will give a brief overview of the derivation of the descent coefficients in simple pedigrees. For more detailed derivation and analysis, see (Weir & Cockerham, 1968; Cockerham & Weir, 1968, 1973, 1977, 1973; Weir & Cockerham, 1974). For the average descent coefficients, F_1, F^1, and ${}_1F$, we will use the lettering of the individuals in Figure 1 and show results in Table A.4.

Also, we will use Schnell's recombination values to make calculations more simple and easy to derive. Using Schnell's terminology, the probability of recombination is $\frac{1-\lambda}{2}$ and the probability of no recombination is $\frac{1+\lambda}{2}$.

For the initial ancestors in the pedigree, which we assume are outbred, F_1 and ${}_1F$ are both zero and F^1 is originally defined as one. Each subsequent generation of descent has a value of F^1 multiplied by the previous generation's value times the probability of no recombination. Note outbred avunculars of A: D and E, have zero probability of allele or gamete identity so have all descent coefficients as zero. Only the final generation where inbreeding is present has values for F_1 and ${}_1F$. The value of F_1 is the inbreeding coefficient based on the coancestry of the parents. The value of ${}_1F$ is given by an algorithm from (Cockerham & Weir, 1977). Where n_1 and n_2 are the number of generations between the parents of A (B and C respectively) and the common ancestors (H and I) and M is the number of distinct paths between B and C and H/I, we can state

$$ {}_1F_A = \left(\frac{1}{2}\right)^{n_1+n_2} \left(\frac{{}_1F_I + F_I^1}{2}\right) M \tag{A.1} $$

ID	F_1	${}_1F$	F^1
I	0	0	1
H	0	0	1
G	0	0	$\frac{1+\lambda}{2}$
F	0	0	$\frac{1+\lambda}{2}$
E	0	0	0
D	0	0	0
C	0	0	$\left(\frac{1+\lambda}{2}\right)^2$
B	0	0	$\left(\frac{1+\lambda}{2}\right)^2$
A	$\frac{1}{16}$	$\frac{1}{16}$	$\left(\frac{1+\lambda}{2}\right)^3$

Table A.4: Average descent coefficients for the pedigree in Figure 1. See also (Cockerham & Weir, 1977).

The first calculation of F^{11} is straightforward for simple pedigrees. For F^1, by definition

$$\begin{aligned} F^1 =& \frac{F^{1\cdot} + F^{\cdot 1}}{2} = \frac{F^{11} + F^{10} + F^{11} + F^{01}}{2} = F^{11} + \frac{F^{10} + F^{01}}{2} = \left(\frac{1+\lambda}{2}\right)^3 \\ F^{10} =& F^{01} = \left(\frac{1+\lambda}{2}\right)^3 \left[1 - \left(\frac{1+\lambda}{2}\right)^3\right] \end{aligned} \tag{A.2}$$

The values of F^{10} and F^{01} are the probability of one gamete descending without recombination while another descends with any numbers of recombinations except zero. Given equation A.2 and our previously derived $F^1 = \left(\frac{1+\lambda}{2}\right)^3$ we can derive

$$F^{11} = \left(\frac{1+\lambda}{2}\right)^6 = (1-c)^6 \tag{A.3}$$

To calculate F_{11}, and ${}_{11}F$, F^{11}_{11} and ${}_1F^1_1$, one expands the descent coefficients backwards through the pedigree from the affected individual to the common ancestors of the parents. For example, F_{11A} for A is the same as the two locus coancestry coefficient for parents B and C, θ_{11BC}

$$\theta_{11BC} = \frac{1+\lambda}{2}\frac{\theta_{11BE}+\theta_{11BF}}{2} + \frac{1-\lambda}{2}\frac{\gamma_{11B,EF}+\gamma_{11B,FE}}{2} \tag{A.4}$$

The fractions $\frac{1+\lambda}{2}$ and $\frac{1-\lambda}{2}$ are the probabilities of no recombination and recombination respectively and the term $(\theta_{11BE} + \theta_{11BF})/2$ is the average of the coancestries of B and the parents of C. The average of $\gamma_{11B,EF}$ and $\gamma_{11B,FE}$ is the average of the three gamete probabilities which are the probability that one gamete comes from B and the other alleles which eventually form a gamete in C come from E and F separately after a recombination event.

Since E is unrelated to F or B, $\theta_{11BE} = 0$ and the terms $\gamma_{11B,EF}$ and $\gamma_{11B,FE}$ are zero as well since different parents cannot contribute to the same gamete in C. So we then have

$$\theta_{11BC} = \frac{1+\lambda}{4}\theta_{11BF} \tag{A.5}$$

Next we expand out θ_{11BF}

$$\theta_{11BF} = \frac{1+\lambda}{2}\frac{\theta_{11DF}+\theta_{11GF}}{2} + \frac{1-\lambda}{2}\frac{\gamma_{11F,DG}+\gamma_{11F,GD}}{2} \tag{A.6}$$

Using similar arguments to before the γ terms are zero and $\theta_{11DF} = 0$ so θ_{11BC} is now

$$\theta_{11BC} = \left(\frac{1+\lambda}{4}\right)^2 \theta_{11GF} \tag{A.7}$$

Now we approach the final expansions to the ancestors H and I.

$$\begin{aligned}\theta_{11GF} =& \frac{1+\lambda}{2}\frac{\theta_{11GH}+\theta_{11GI}}{2} + \frac{1-\lambda}{2}\frac{\gamma_{11G,HI}+\gamma_{11G,IH}+\gamma_{11G,HH}+\gamma_{11G,II}}{4} \\ \theta_{11GF} =& \frac{1+\lambda}{2}\frac{\theta_{11GH}+\theta_{11GI}}{2} + \frac{1-\lambda}{2}\frac{\gamma_{11G,HH}+\gamma_{11G,II}}{2}\end{aligned} \tag{A.8}$$

The second expression follows for γ since H and I are outbred so $\gamma_{11G,HI} = \gamma_{11G,IH} = 0$. Now we approach the final expansions to the ancestors H and I.Following

$$
\begin{aligned}
\theta_{11GH} =& \frac{1+\lambda}{2}\frac{\theta_{11HH}+\theta_{11HI}}{2} + \frac{1-\lambda}{2}\frac{\gamma_{11H,HI}+\gamma_{11H,IH}}{2} \\
\theta_{11GI} =& \frac{1+\lambda}{2}\frac{\theta_{11II}+\theta_{11IH}}{2} + \frac{1-\lambda}{2}\frac{\gamma_{11I,HI}+\gamma_{11I,IH}}{2} \\
\gamma_{11G,HH} =& \gamma_{11G,II} = \frac{1-\lambda}{2}\gamma_{11H,HH} = \frac{1-\lambda}{2}\gamma_{11I,II} = \frac{1-\lambda}{2}\frac{1}{4}
\end{aligned}
\tag{A.9}
$$

Note that since H and I are outbred, $\gamma_{11H,HI} = \gamma_{11H,IH} = \gamma_{11I,HI} = \gamma_{11I,IH} = 0$. The only three gamete probabilities are for one un-recombined gamete and two alleles from a recombined gamete to come from the same individual thus $\gamma_{11H,HH} = \gamma_{11I,II} = 1/4$. Since the ancestors are outbred, $\theta_{11HI} = \theta_{11IH} = 0$ and per the definition of two locus IBD with ones self (Cockerham & Weir, 1973), $\theta_{11HH} = \theta_{11II} = \frac{1+\lambda^2}{4}$. So finally we have

$$
\begin{aligned}
F_{11A} = \theta_{11BC} =& \left(\frac{1+\lambda}{4}\right)^4 (\theta_{11HH}+\theta_{11II}) + \left(\frac{1+\lambda}{4}\right)^2\left(\frac{1-\lambda}{4}\right)^2 (\gamma_{11H,HH}+\gamma_{11I,II}) \\
F_{11A} = \theta_{11BC} =& \left(\frac{1+\lambda}{4}\right)^4 \frac{1+\lambda^2}{2} + \frac{1}{2}\left(\frac{1+\lambda}{4}\right)^2\left(\frac{1-\lambda}{4}\right)^2 \\
F_{11A} =& \frac{1}{16}\left((1-c)^6 + c^2(1-c)^4 + \frac{1}{2}c^2(1-c)^2\right)
\end{aligned}
\tag{A.10}
$$

The final line is the result in terms of c identical to the result in (Haldane, 1949). To calculate ${}_{11}F$, F^{11}_{11} and ${}_1F^1_1$ we use identical expansions to the first line in Equation A.10 where the subscripts are changed and different values for ultimate common ancestor probabilities (identical subscripts) are used.

Descent Coefficient	Two Gamete Probabilities	Three Gamete Probabilities
${}_{11}F$	${}_{11}\theta_{HH} = {}_{11}\theta_{II} = \frac{1+\lambda^2}{4}$	${}_{11}\gamma^1_{1H,HH} = {}_{11}\gamma^1_{1I,II} = 0$
${}_1F^1_1$	${}_1\theta^1_{1HH} = {}_1\theta^1_{1II} = \frac{1+\lambda}{4}$	${}_1\gamma^1_{1H,HH} = {}_1\gamma^1_{1I,II} = 1/4$
F^{11}_{11}	$\theta^{11}_{11HH} = \theta^{11}_{11II} = \frac{(1+\lambda)^2}{8}$	$\gamma^{11}_{11H,HH} = \gamma^{11}_{11I,II} = 0$

Table A.5: Two and three gamete descent coefficients for the pedigree in Figure 1. For most derivations, see (Cockerham & Weir, 1973; Weir & Cockerham, 1974).

This gives

$$
\begin{aligned}
F^{11} &= \Big(\frac{1+\lambda}{2}\Big)^6 = (1-c)^6 \\
{}_{11}F &= \Big(\frac{1+\lambda}{4}\Big)^4 \frac{1+\lambda^2}{2} = \frac{1}{16}((1-c)^4 - 2c(1-c)^4 + 2c^2(1-c)^4) \\
F^{11}_{11} &= 4\Big(\frac{1+\lambda}{4}\Big)^6 = \frac{1}{16}(1-c)^6 \\
{}_1F^1_1 &= \frac{1}{16}\Big(\frac{1+\lambda}{4}\Big)^5 + \frac{1}{2}\left(\frac{1+\lambda}{4}\right)^2\left(\frac{1-\lambda}{4}\right)^2 \\
&= \frac{1}{32}(2(1-c)^5 + c^2(1-c)^2)
\end{aligned}
\tag{A.11}
$$

Appendix B. Data Availability

N/A

Declarations of interest: none. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Badano, J.L. & Katsanis, N. 2002. Beyond Mendel: an evolving view of human genetic disease transmission. Nat. Rev. Genet., 3:779-789. doi: 10.1038/nrg910

Balloux F., Amos W., and Coulson T., 2004. Does heterozygosity estimate inbreeding in real populations? Mol. Ecol., 13:3021–3031. doi: 10.1111/j.1365-294X.2004.02318.x

Basit, S., Wali, A., Aziz, A., Muhammad, N., Jelani, M. & Ahmad, W., 2011. Digenic inheritance of an autosomal recessive hypotrichosis in two consanguineous pedigrees. Clinical genetics, 79:273-281. doi: 10.1111/j.1399-0004.2010.01455.x

Bétard, C., Rasquin-Weber, A., Brewer, C., Drouin, E., Clark, S., Verner, A., Darmond-Zwaig, C., Fortin, J., Mercier, J., Chagnon, P. & Fujiwara, T.M., 2000. Localization of a recessive gene for North American Indian childhood cirrhosis to chromosome region 16q22—and identification of a shared haplotype. Am. J. Hum. Genet., 67:222-228. doi: 10.1086/302993

Bodmer, W. F., & Felsenstein, J. 1967. Linkage and selection: theoretical analysis of the deterministic two locus random mating model. Genetics 57:237-265. doi: 10.1093/genetics/57.2.237

Bürger, R., 2020. Multilocus population-genetic theory. Theor. Popul. Biol., 133: 40-48. doi: 10.1016/j.tpb.2019.09.004

Chagnon, P., Michaud, J., Mitchell, G., Mercier, J., Marion, J.F., Drouin, E., Rasquin-Weber, A., Hudson, T.J. & Richter, A., 2002. A missense mutation (R565W) in cirhin (FLJ14728) in North American Indian childhood cirrhosis. Am. J. Hum. Genet., 71:1443-1449. doi: 10.1086/344580

Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics, 39:859-882 doi: 10.1093/genetics/39.6.859

Cockerham, C.C., 1956. Effects of linkage on the covariances between relatives. Genetics, 41:138-143. doi: 10.1093/genetics/41.1.138

Cockerham, C.C. & Weir, B.S., 1968. Sib mating with two linked loci. Genetics, 60:629-640. doi: 10.1093/genetics/60.3.629

Weir, B.S. & Cockerham, C.C., 1969. Group inbreeding with two linked loci. Genetics, 63:711-742. doi: 10.1093/genetics/63.3.711

Cockerham, C.C. and Weir, B.S., 1973. Descent measures for two loci with some applications. Theor. Popul. Biol., 4:300-330. doi: 10.1016/0040-5809(73)90013-0

Weir, B.S. and Cockerham, C.C., 1973. Mixed self and random mating at two loci. Genet. Res., 21:247-262. doi: 10.1017/S0016672300013446

Cockerham, C.C. & Weir, B.S. 1977. Digenic descent measures for finite populations. Genet. Res., 30:121-147. doi: 10.1017/S0016672300017547

Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. 2013. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum. Genet., 132:1077-1130. doi: 10.1007/s00439-013-1331-2

David, P., 1998. Heterozygosity–fitness correlations: new perspectives on old problems. Heredity, 80:531-537. doi: 10.1046/j.1365-2540.1998.00393.x

David, P., Pujol, B., Viard, F., Castella, V. and Goudet, J., 2007. Reliable selfing rate estimates from imperfect population genetic data. Mol. Ecol., 16:2474-2487. doi: 10.1111/j.1365-294X.2007.03330.x

Deltas, C., 2018. Digenic inheritance and genetic modifiers. Clin. Genet., 93:429-438. doi: 10.1111/cge.13150

Ebermann, I., Walger, M., Scholl, H.P., Issa, P.C., Lüke, C., Nürnberg, G., Lang-Roth, R., Becker, C., Nürnberg, P. & Bolz, H.J., 2007. Truncating mutation of the DFNB59 gene causes cochlear hearing impairment and central vestibular dysfunction. Human mutation, 28:571-577. doi: 10.1002/humu.20478

Fisher, R.A., 1928. The possible modification of the response of the wild type to recurrent mutations. Am. Nat., 62:115-126. doi: 10.1086/280193

Fisher, R.A., 1928. Two further notes on the origin of dominance. Am. Nat., 62:571-574. doi: 10.1086/280234

Fisher, R.A., 1929. The evolution of dominance; reply to Professor Sewall Wright. Am. Nat., 63:553-556. doi:10.1086/280289

Goldin, L.R. & Weeks, D.E., 1993. Two-locus models of disease: comparison of likelihood and nonparametric linkage methods. Am. J. Hum. Genet., 53:908-915.

Haldane, J.B.S., 1949. The association of characters as a result of inbreeding and linkage. Ann. Eugen., 15:15-23. doi: 10.1111/j.1469-1809.1949.tb02418.x

Hallgrímsdóttir, I. B., & Yuster, D. S. 2008. A complete classification of epistatic two-locus models. BMC Genet., 9:1-15. doi: 10.1186/1471-2156-9-17

Hartl, D.L. and Maruyama, T., 1968. Phenogram enumeration: the number of regular genotype-phenotype correspondences in genetic systems. J. Theor. Biol., 20:129-163. doi: 10.1016/0022-5193(68)90186-0

Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. Theoretical and Applied Genetics, 38(6), 226-231. doi:10.1007/BF01245622

Hodge, S.E., 1981. Some epistatic two-locus models of disease. I. Relative risks and identity-by-descent distributions in affected sib pairs. Am. J. Hum. Genet., 33:381-395.

Hodge, S.E. and Spence, M.A., 1981. Some epistatic two-locus models of disease. II. The confounding of linkage and association. Am. J. Hum. Genet., 33:396-406.

Hogben, L., 1932. The genetic analysis of familial traits - II. Double gene substitutions, with special reference to hereditary dwarfism. J. Genet., 25:211-240. doi: 10.1007/BF02983255

Jacquard, A., 1975. Inbreeding: One word, several meanings. Theor. Popul. Biol., 7:338-363. doi: 10.1016/0040-5809(75)90024-6

Kajiwara, K., Berson, E.L. and Dryja, T.P., 1994. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science, 264: 1604-1608. doi: 10.1126/science.8202715

Karlin, S. & Feldman, M.W., 1970. Linkage and selection: two locus symmetric viability model. Theor. Popul. Biol., 1:39-71. doi: 10.1016/0040-5809(70)90041-9.

Karlin, S. 1975. General two-locus selection models: some objectives, results and interpretations. Theor. Popul. Biol., 7:364-398. doi: 10.1016/0040-5809(75)90025-8

Karlin, S. 1979. Principles of polymorphism and epistasis for multilocus systems. Proc. Natl. Acad. Sci. U.S.A., 76:541-545. doi: 10.1073/pnas.76.1.541

Karlin, S. & Avni, H. 1981. Analysis of central equilibria in multilocus systems: A generalized symmetric viability regime. Theor. Popul. Biol., 20: 241-280. doi: 10.1016/0040-5809(81)90012-5

Kempthorne, O. 1954. The correlation between relatives in a random mating population. Proc. Royal Soc. B, 143:103-113. doi: 10.1098/rspb.1954.0056

Kimura, M. (1956). A model of a genetic system which leads to closer linkage by natural selection. Evolution, 10:278-287. doi:10.1111/j.1558-5646.1956.tb02852.x

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B. & Tukiainen, T., 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536:285-291. doi: 10.1038/nature19057

Lewontin, R.C. & Kojima, K.I. 1960. The evolutionary dynamics of complex polymorphisms. Evolution, 14:458-472. doi: 10.2307/2405995

Li, C.C., 1953. Some general properties of recessive inheritance. Am. J. Hum. Genet., 5:269-279.

Li, W. & Reich, J., 2000. A complete enumeration and classification of two-locus disease models. Hum. Hered., 50:334-349. doi: 10.1159/000022939

Malécot, G. 1970. The Mathematics of Heredity. W.H. Freeman: San Francisco.

Merry, A., Roger, J.H. & Curnow, R.N., 1979. A two-locus model for the inheritance of a familial disease. Ann. Hum. Genet., 43:71-80. doi: 10.1111/j.1469-1809.1979.tb01550.x

Miller, J.M. and Coltman, D.W., 2014. Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: a meta-analysis. Mol. Ecol., 23:1899-1909.

Moore, J.H. & Williams S.M. (Eds.) 2016. *Epistasis*, New York: Humana Press.

Nei, M. & Li, W.H., 1973. Linkage disequilibrium in subdivided populations. Genetics, 75:213-219. doi:10.1093/genetics/75.1.213.

Neuman, R. J., Rice, J. P. & Chakravarti, A. 1992. Two-locus models of disease. Genet. Epidemiol., 9:347-365. doi: 10.1002/gepi.1370090506

Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University, Baltimore, MD. 604901: last edited: September 2, 2016. World Wide Web URL: https://omim.org/entry/604901

Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), (July 3, 2022). World Wide Web URL: https://omim.org/

Boyle, E.A., Li, Y.I. and Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell, 169:1177-1186. doi: 10.1016/j.cell.2017.05.038

Riazuddin, S., Castelein, C.M., Ahmed, Z.M., Lalwani, A.K., Mastroianni, M.A., Naz, S., Smith, T.N., Liburd, N.A., Friedman, T.B., Griffith, A.J. & Riazuddin, S., 2000. Dominant modifier DFNM1 suppresses recessive deafness DFNB26. Nature genetics, 26:431-434. doi: 10.1038/82558

Rosenberg, N.A. & Blum, M.G., 2007. Sampling properties of homozygosity-based statistics for linkage disequilibrium. Math. Biosci., 208:33-47. doi: 10.1016/j.mbs.2006.07.001

Sabatti, C. & Risch, N., 2002. Homozygosity and linkage disequilibrium. Genetics, 160:1707-1719. doi: 10.1093/genetics/160.4.1707

Schäffer, A. A. 2013. Digenic inheritance in medical genetics. J. Med. Genet., 50:641-652. doi: 10.1136/jmedgenet-2013-101713

Schnell, F.W., 1961. Some general formulations of linkage effects in inbreeding. Genetics, 46:947-957. doi:10.1093/genetics/46.8.947

Slatkin, M., 1994. Linkage disequilibrium in growing and stable populations. Genetics, 137:331-336. doi: 10.1093/genetics/137.1.331

Van Steen, K., 2012. Travelling the world of gene–gene interactions. Brief. Bioinformatics, 13:1-19. doi: 10.1093/bib/bbr012

Weir, B.S., 1979. Inferences about linkage disequilibrium. Biometrics, 35:235-254.

Weir, B.S., 2008. Linkage disequilibrium and association mapping. Annu. Rev. Genomics Hum. Genet., 9:129-142. doi: 10.1146/annurev.genom.9.081307.164347

Weir, B. S. & Cockerham, C. C. 1968. The two-locus inbreeding function. North Carolina State University. Dept. of Statistics.

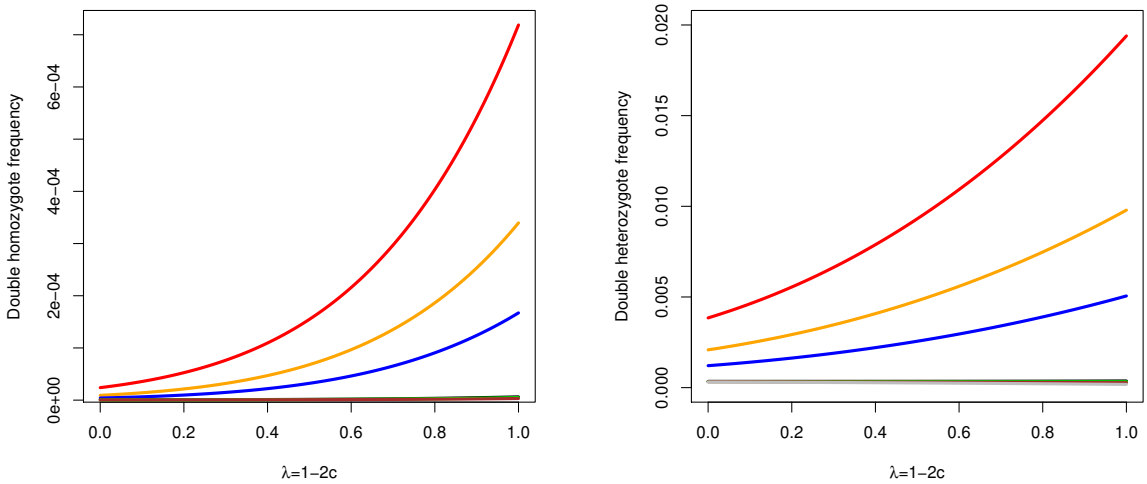
Weir, B.S. & Cockerham, C.C., 1974. Behavior of pairs of loci in finite monoecious populations. Theor. Popul. Biol., 6:323-354. doi: 10.1016/0040-5809(74)90015-X

Weir, B.S. & Cockerham, C.C., 1989. Complete characterization of disequilibrium at two loci. In: *Mathematical evolutionary theory*, Princeton: Princeton University Press, pp. 86-110.

Wright, S., 1922. Coefficients of inbreeding and relationship. Am. Nat., 56:330-338. doi: 10.1086/279872

Yengo, L., Wray, N.R. & Visscher, P.M., 2019. Extreme inbreeding in a European ancestry sample from the contemporary UK population. Nature communications, 10:1-11. doi: 10.1038/s41467-019-11724-6

$D \geq 0$



$D \leq 0$

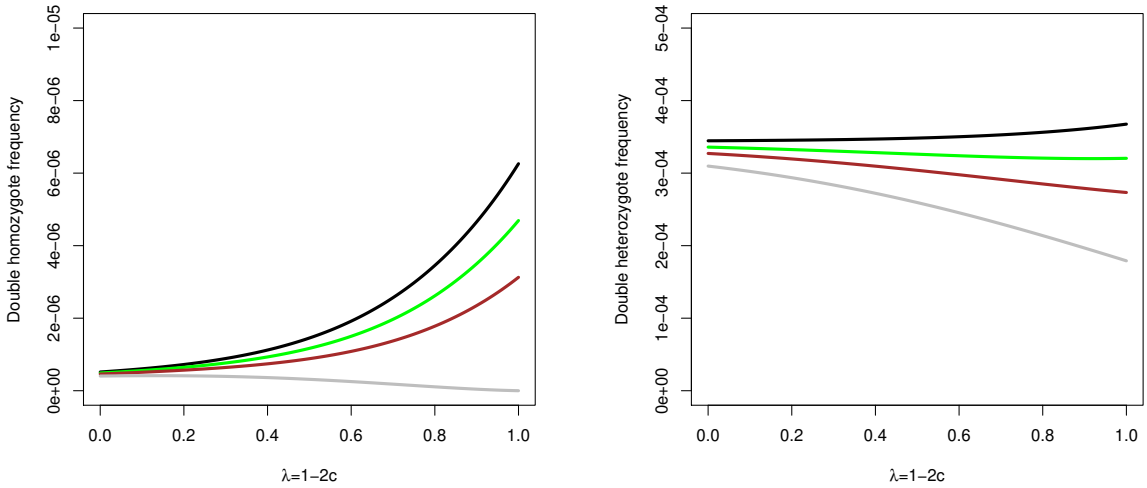


Figure 5: Plots of the frequencies of double homozygous and double heterozygous genotypes versus linkage where the frequency of the minor allele is 0.01 and the different colors represent different values of linkage disequilibrium. $D' = 1, D' = 0.5, D' = 0.25, D' = 0, D' = -0.25, D' = -0.5, D' = -1$. are represented by the line colors red, orange, blue, black, green, brown, and gray. Linkage is designated by the Schnell linkage value $\lambda = 1-2c$.

	BB	Bb	bb
AA	W_{22}	W_{21}	W_{20}
Aa	W_{12}	W_{11}	W_{10}
aa	W_{02}	W_{01}	W_{00}

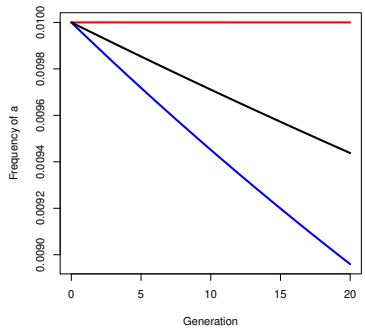
Figure 6: Variables of relative fitness by genotype.

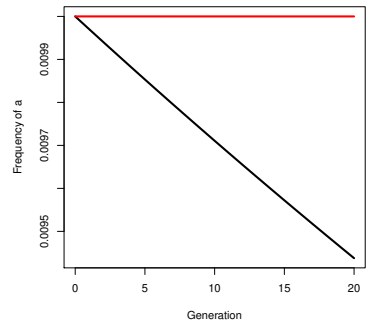
	BB	Bb	bb
AA	1	1	1
Aa	1	1	1
aa	$1-s_1$	$1-s_1$	$1-s_1-s_2$

	BB	Bb	bb
AA	1	1	1
Aa	1	$1-s_1$	1
aa	1	1	1

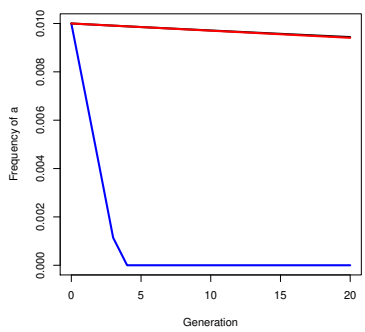
Figure 7: The relative fitness matrices for deleterious double homozygous and double heterozygous cases with the alleles at the primary locus A/a being essential in the digenic double homozygous recessive case. The selection coefficients s_1 and s_2 are greater than zero and less than 1 but also $s_1 + s_2 \leq 1$. For the double homozygous case, where this primary locus is homozygous recessive as aa, the relative fitness is $1-s_1$, where the double homozygous recessive genotype $aabb$ is present, relative fitness is $1-s_1-s_2$. For the double heterozygous case, the relative fitness is $1 - s_1$ for the $AaBb$ genotype.

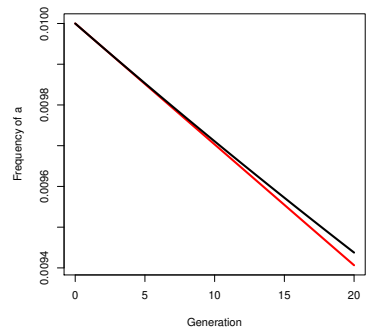
$D = 0$ and $c = 0.5$





$D = 0.099$ and $c = 0$





$D = -0.0001$ and $c = 0$

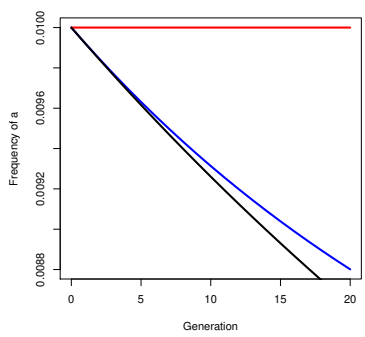


Figure 8: Simulations of the frequency of the allele a associated with a deleterious double homozygous (red), double heterozygous (blue) or single locus homozygous (black) genotypes. Alleles a and b (not graphed) have starting allele frequencies of 0.01 and linkage and linkage disequilibrium as shown in the graphs. The selection coefficient is 0.3 in all cases. The charts with only the double homozygous and single locus homozygous cases are given for clarity due to scale.

	BB	Bb	bb
AA	AxA	AxD	AxA
Aa	DxA	DxD	DxA
aa	AxA	AxD	AxA

Figure 9: Chart of type of two-locus epistasis by genotype.