bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Digenic traits under population admixture and inbreeding

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

The population genetics of digenic genotypes in diploid populations, genotypes based on alleles at two loci, have been studied theoretically for decades with relevant digenic traits of medical interest being known for over 25 years. Given the effects of linkage and linkage disequilibrium on two locus genotypes, it should be expected that these factors can change the expected frequencies of digenic genotypes in many, sometimes unexpected, ways. In particular, the combination of linkage disequilibrium and inbreeding can combine to increase the frequencies of double homozygotes and double heterozygotes significantly over outbred comparisons. Given the prevalence of linkage disequilibrium in recently admixed populations, this can lead to large shifts in trait prevalence such that it can sometimes exceed that of either original pre-admixed population with the combined effects of linkage disequilibrium and inbreeding. Here we investigate the frequencies of digenic genotypes under the combined effects of linkage, linkage disequilibrium, and inbreeding to analyze how these interact to increase or decrease the frequency of the genotypes across two loci.

Keywords: digenic traits, oligogenic traits, inbreeding, linkage, population admixture

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

more loci. Early work on multilocus genetics included Hogben (1932) and Li (1953) on double homozygous recessive traits and Haldane (1949) in the 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 stable equilibria 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).

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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, which 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 non-overlapping cases of digenic disorders were listed along with the method of inheritance for each locus. In the Online Mendelian Inheritance in Man (OMIM) database OMIM (2021), 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.

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 a high prevalence of linkage disequilibrium should be expected to have impacts on the frequencies of digenic traits, especially when mating between relatives occurs.

1.1. Population admixture and linkage disequilibrium

It has been long established that 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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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. The most obvious effect, that of linkage disequilibrium in an outbred population will be explained first and show how this linkage disequilibrium directly alters the frequencies of digenic genotypes. This is the most common outcome and is not really dependent on linkage except in how rapidly linkage disequilibrium decays.

Second, we will show how linkage disequilibrium accompanied by linkage between loci amplifies the effects of inbreeding. This shows that even compared to inbreeding in similar populations without linkage disequilibrium, inbreeding in populations with linkage disequilibrium between loci can substantially increase the frequency of digenic genotypes by orders of magnitude. Finally, we will demonstrate how these factors interplay in an example of expected digenic genotypes for the progeny of first cousins.

While the results derived here are valid for any two loci in any type of population that are in linkage disequilibrium, we focus on the case of admixed populations since linkage disequilibrium can be much more pervasive between loci, even those that are unlinked, and understanding the genetic epidemiology of digenic traits can be counterintuitive as with the increased frequencies of double heterozygotes due to inbreeding.

2. Digenic genotype frequencies

In this section, we will outline the effects on digenic genotype frequencies due to linkage disequilibrium and inbreeding. The first case will cover linkage disequilibrium in outbred populations alone and then incorporate the necessary adjustments for inbreeding.

2.1. The effects of linkage disequilibrium

When the population is outbred, the effect of linkage disequilibrium on genotype frequencies is clear as shown in Table 1. 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	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 1: Two loci genotype frequencies taking linkage disequilibrium into account. 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 6.7×10^{-6}, over a 600-fold increase though still relatively uncommon.

2.2. The effects of inbreeding at two loci

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 as well as the separation of a population into lines which have a steady decreasing genetic variance within the line and an increasing genetic variance between lines. 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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:

$$\begin{aligned} F_{11} &= P(a_1 \equiv a_2) \text{ and } P(b_1 \equiv b_2) \\ F_{10} &= P(a_1 \equiv a_2) \text{ and } P(b_1 \not\equiv b_2) \\ F_{01} &= P(a_1 \not\equiv a_2) \text{ and } P(b_1 \equiv b_2) \\ F_{00} &= P(a_1 \not\equiv a_2) \text{ and } P(b_1 \not\equiv b_2) \end{aligned} \tag{2}$$

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.

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. It can also be interpreted as the correlation between heterozygosity across loci caused by inbreeding at two loci. Where the inbreeding coefficient is the same for both loci, ID is defined as

$$\eta_{11} = F_{11} - F^2 \tag{3}$$

The main effect of identity disequilibrium 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 heterozygotes is in contrast to the effects of inbreeding at a single locus where the frequency of heterozygotes gradually decreases with increased inbreeding. This still occurs at single loci due to the effects of inbreeding but it is moderated or even reversed for double

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

heterozygotes given the value of the identity disequilibrium. The effects of η_{11} by genotype are shown in Table 2.

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

Table 2: Two loci genotype frequencies under inbreeding at linkage equilibrium taking identity disequilibrium into account. See also Cockerham & Weir (1973). 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, $P_{AA} = p_A^2 + p_A(1-p_A)F$, $P_{Aa} = 2p_A(1-p_A)(1-F)$, $P_{aa} = (1-p_A)^2 + p_A(1-p_A)F$, $P_{BB} = p_B^2 + p_B(1-p_B)F$, $P_{Bb} = 2p_B(1-p_B)(1-F)$, $P_{bb} = (1-p_B)^2 + p_B(1-p_B)F$.

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.

2.3. The combined impact of inbreeding and linkage disequilibrium

The sole addition of identity disequilibrium to the genotype frequencies described in Table 2 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.

The first, 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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

$$\begin{aligned} F^{11} &= P(a_1 \equiv b_1) \text{ and } P(a_2 \equiv b_2) \\ F^{10} &= P(a_1 \equiv b_1) \text{ and } P(a_2 \not\equiv b_2) \\ F^{01} &= P(a_1 \not\equiv b_1) \text{ and } P(a_2 \equiv b_2) \\ F^{00} &= P(a_1 \not\equiv b_1) \text{ and } P(a_2 \not\equiv b_2) \end{aligned} \tag{4}$$

The average parental coefficient across both individuals can be defined similarly to the inbreeding coefficient case and is designated as F^1.

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 other 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

$$\begin{aligned} {}_{11}F &= P(a_1 \equiv b_2) \text{ and } P(a_2 \equiv b_1) \\ {}_{10}F &= P(a_1 \equiv b_2) \text{ and } P(a_2 \not\equiv b_1) \\ {}_{01}F &= P(a_1 \not\equiv b_2) \text{ and } P(a_2 \equiv b_1) \\ {}_{00}F &= P(a_1 \not\equiv b_2) \text{ and } P(a_2 \not\equiv b_1) \end{aligned} \tag{5}$$

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 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.

2.4. 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

We will describe the double homozygous genotypes 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

$$\begin{aligned} P(ab|ab)^* =& (p_a^2 + p_A(1-p_A)F)(p_b^2 + p_B(1-p_B)F) + p_A(1-p_A)p_B(1-p_B)\eta_{11} \\ P(AB|ab)^* =& P(ab|AB)^* = P(Ab|aB)^* = P(aB|Ab)^* \\ =& p_A(1-p_A)p_B(1-p_B)(1-F)^2 + p_A(1-p_A)p_B(1-p_B)\eta_{11} \end{aligned} \tag{6}$$

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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

$$
\begin{aligned}
P(ab|ab) =& P(ab|ab)^* + D^2\Big[F^{11} + {}_{11}F - 2F^{11}_{11}\Big] \\
&+ D\Big[2(F^1 + {}_1F)p_a p_b \\
&+ F^{11}_{11}(1-2p_a)(1-2p_b) + 2{}_1F^1_1(p_a + p_b - 4p_a p_b)\Big] \\
P(AB|ab) =& P(ab|AB) = P(AB|ab)^* + D^2\Big[F^{11} + {}_{11}F - 2F^{11}_{11}\Big] \\
&+ D\Big[F^1(p_A p_B + (1-p_A)(1-p_B)) + {}_1F(p_A(1-p_B)+ \\
&(1-p_A)p_B) - 2{}_1F^1_1(1-2p_A)(1-2p_B) \\
&+ F^{11}_{11}(p_A p_B + (1-p_A)(1-p_B) + p_A(1-p_B) + (1-p_A)p_B)\Big] \\
P(Ab|aB) = P(aB|Ab) =& P(Ab|aB)^* + D^2\Big[F^{11} + {}_{11}F - 2F^{11}_{11}\Big] + D\Big[F^1(p_A(1-p_B)+ \\
&(1-p_A)p_B) + {}_1F(p_A p_B + (1-p_A)(1-p_B)) \\
&- 2{}_1F^1_1(1-2p_A)(1-2p_B) \\
&+ F^{11}_{11}(p_A p_B + (1-p_A)(1-p_B) + p_A(1-p_B) + (1-p_A)p_B)\Big]
\end{aligned} \tag{7}
$$

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

$$
\begin{aligned}
F^{11}_{11} =& P(a_1 \equiv a_2 \equiv b_1 \equiv b_2) \\
{}_1F^1_1 =& \frac{1}{4}[P(a_1 \equiv a_2 \equiv b_1) + P(a_1 \equiv a_2 \equiv b_2) + P(a_1 \equiv b_1 \equiv b_2) + P(a_2 \equiv b_1 \equiv b_2)]
\end{aligned} \tag{8}
$$

Again $\equiv$ means 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

$$\begin{aligned}P(AaBb) =& 2P(AB|ab) + 2P(Ab|aB) = 4P(AB|ab)^* + 4D^2\Big[F^{11} + {}_{11}F - 2F^{11}_{11}\Big] \\ &+ 2D\Big[(F^1 + {}_1F)(p_Ap_B + (1-p_A)(1-p_B) + p_A(1-p_B) \\ &+ (1-p_A)p_B) - 4{}_1F^1_1(1-2p_A)(1-2p_B) \\ &+ 2F^{11}_{11}(p_Ap_B + (1-p_A)(1-p_B) + p_A(1-p_B) + (1-p_A)p_B)\Big]\end{aligned} \tag{9}$$

The frequency difference between the two types of double heterozygosity can be given by

$$\begin{aligned}2P(AB|ab) - 2P(Ab|aB) =& 2D\Big[(F^1 - {}_1F)(p_Ap_B + (1-p_A)(1-p_B) \\ &- p_A(1-p_B) - (1-p_A)p_B)\Big]\end{aligned} \tag{10}$$

Equations 7 and 9 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 1. 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 1.

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 though it is usually negligible except when the loci are tightly linked. Thus, compared

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

to the case with inbreeding but no linkage disequilibrium, positive and negative linkage disequilibrium in the ancestor increase of decrease the genotype frequencies of double homozygotes and double heterozygotes.

To understand the magnitude of this difference, however, we will explore the effect of first cousin mating on the expected digenic genotypes under all conditions for a double homozygous trait inherited in an autosomal recessive manner as well as a double heterozygous trait inherited in an autosomal dominant manner. This will help demonstrate the magnitude of possible effects.

3. First cousin inbreeding example

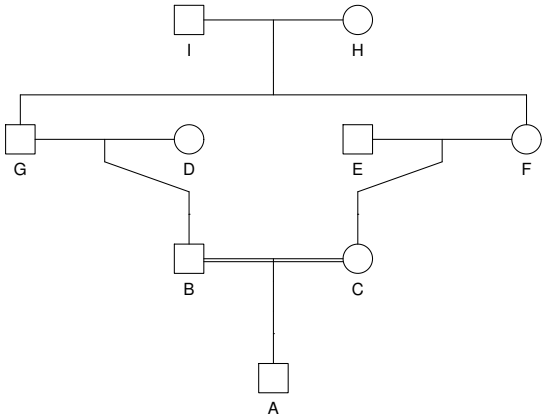


Figure 1: Pedigree of first cousin inbreeding.

In this example, we will examine the impact that the combined effects of linkage disequilibrium and inbreeding have on the frequency of double homozygous and double heterozygous genotypes at two loci with rare minor allele frequencies of 0.01. As will be demonstrated, not only does inbreeding increase double homozygosity greatly, as expected, but also double heterozygosity, even though inbreeding reduces heterozygosity and a single locus.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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:

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

These descent coefficients can then be used to calculate the double homozygous and double heterozygous genotype frequencies per equations 7 and 9. 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 2 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 (Schnell (1961)) $\lambda = 1 - 2c$ that ranges from $[0, 1]$ as c ranges from $[1/2, 0]$ since this more clearly demonstrates the increasing frequency with linkage.

Consider two types of digenic hereditary disorders with separate aetiologies. In both the minor allele of interest has a population frequency of 0.01. One condition is inherited in an autosomal recessive manner and only expressed when a double homozygous recessive genotype is present. The other is inherited in an autosomal dominant manner with a rare dominant allele and is expressed in double homozygotes, double heterozygotes, and mixed dominant homozgyote/heterozygotes. However, the largest frequency of occurrence is with double heterozygotes. If the rare alleles are considered the causative alleles in both cases, for the autosomal recessive condition,

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

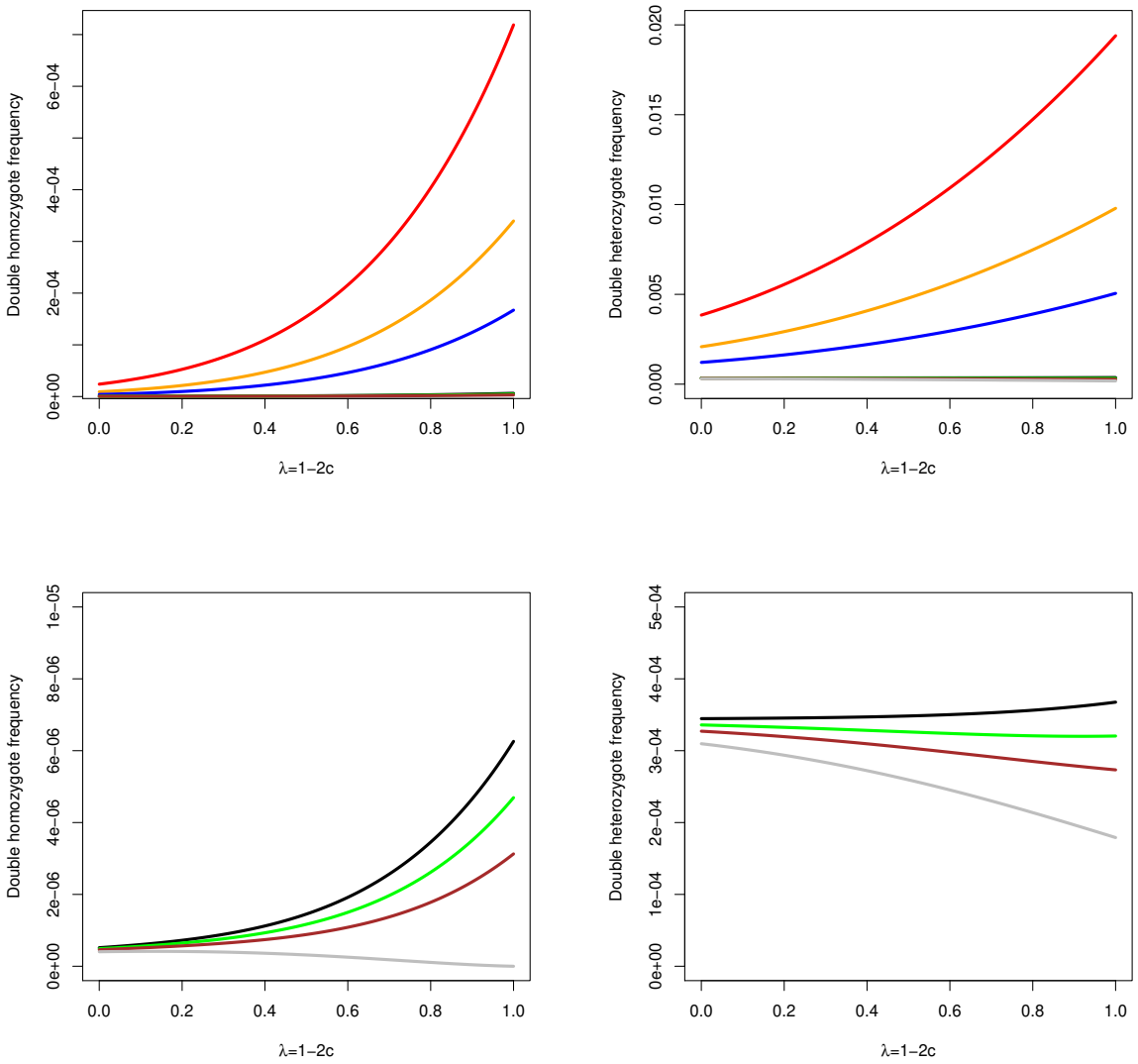


Figure 2: Plots of the frequencies of double homozygous and double heterozygous genotypes versus linkage where the frequency of the minor allele is 0.01 and the different colors represent different values of linkage disequilibrium.$D' = 1, D' = 0.5, D' = 0.25, D' = 0, D' = -0.25, D' = -0.5, D' = -1$.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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

there would only be one case in 100 million persons. 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, there is not a large change in incidence with double homozygosity rising to 5.2 cases per 10 million, a 50-fold increase, and double heterozygosity actually reducing to 34 cases per 100,000. This is solely due to the combined effects of single locus inbreeding genotype changes at each locus.

The effect of inbreeding increases with linkage however, especially for double homozygotes, and for $c = 1/4$ the rate becomes 0.14 per 100,000, a 2.7X increase, for double homozygotes and 35 per 100,000 for double heterozygotes. For completely linked loci the rates become 0.63 per 100,000 for double homozygotes and 37 per 100,000 for double heterozygotes. Thus when there is inbreeding without linkage disequilibrium, the effects of identity disequilibrium are more marked on double homozygotes. Double heterozygotes see a small decrease, though less than expected if one only considered the decline in heterozygosity at the individual loci.

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 120 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 500 per 100,000 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.

These rates are not reflective of the entire population since cousin marriages in most societies are relatively uncommon. However, for the partial linkage disequilibrium of $D' = 0.25$, for the double heterozygotes, unlinked

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

loci will have an incidence of about 1 in 833 offspring for cousin inbreeding with double homozygotes having a rate of 1 in 300,000. For linked loci, the double homozygote rate would be 1 in 5,882 offspring and the double heterozygote rate would be 1 in 200.

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}
		$c = 1/4$	2.9×10^{-5}	2.5×10^{-3}
		$c = 0$	1.7×10^{-4}	5.0×10^{-3}
	$D' = 1$	$c = 1/2$	1.3×10^{-5}	3.8×10^{-3}
		$c = 1/4$	1.3×10^{-4}	9×10^{-3}
		$c = 0$	7.2×10^{-4}	0.019

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 $D'(1-c)^3$.

4. Simplified expressions

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

$$\begin{aligned} P(ab|ab) =& P(ab|ab)^* + 2D(F^1 + {}_1F)p_a p_b \\ P(AaBb) =& 4P(AB|ab)^* + 2D(F^1 + {}_1F)(p_A p_B + (1-p_A)(1-p_B) + p_A(1-p_B) + (1-p_A)p_B) \end{aligned} \tag{12}$$

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

$$\begin{aligned} P(ab|ab) =& P(ab|ab)^* + 2D((1-c)^n + F)p_a p_b \\ P(AaBb) =& 4P(AB|ab)^* + 2D((1-c)^n + F)(p_A p_B + (1-p_A)(1-p_B) \\ &+ p_A(1-p_B) + (1-p_A)p_B) \end{aligned} \tag{13}$$

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

$$\begin{aligned} P(ab|ab) =& P(ab|ab)^* + 2D(\left(\frac{1}{2}\right)^n + F)p_a p_b \\ P(AaBb) =& 4P(AB|ab)^* + 2D(\left(\frac{1}{2}\right)^n + F)(p_A p_B + (1-p_A)(1-p_B) \\ &+ p_A(1-p_B) + (1-p_A)p_B) \end{aligned} \tag{14}$$

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 $F_{11} = F^{11}_{11} = F$.

$$\begin{aligned} P(ab|ab) =& P(ab|ab)^* + D^2(1-F) + 2D((1-c)^n + F)p_a p_b \\ P(AaBb) =& 4P(AB|ab)^* + 4D^2(1-F) + 2D((1-c)^n + F)(p_A p_B + (1-p_A)(1-p_B) \\ &+ p_A(1-p_B) + (1-p_A)p_B) \end{aligned} \tag{15}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

5. Discussion

The dynamics of two-locus systems are much more complex than those at a single locus and the analysis of the effects of inbreeding are not an exception. As shown by early work on two locus dynamics in Lewontin & Kojima (1960), linkage disequilibrium and linkage play integral parts in understanding solutions to problems of haplotypes and genotypes across two loci. While the analysis here can apply to any pair of loci in linkage disequilibrium and subject to consanguineous mating, the more pervasive presence of linkage disequilibrium in populations with substantial recent admixture gives emphasis to the results when applied to these cases.

In short, positive linkage disequilibrium can raise the expected frequency of some two-locus genotypes, especially double homozygous ones, above what would be expected based on allele frequencies alone. However, the impact of linkage and identity of alleles by descent across loci can increase it markedly through the identity disequilibrium and the other effects two-locus inbreeding coefficients have under linkage disequilibrium. This can give rise to double genotype frequencies higher in admixed populations than in either original population, a circumstance only possible at single loci for the frequency of heterozygotes under population admixture. In the single locus case, the heterozygote frequency in an admixed population can be higher if the admixed population allele frequency is closer to 1/2 than either original population. However, at two loci the various effects under inbreeding with linkage disequilibrium can give rise to a case where double homozygous and double heterozygous genotypes have higher frequencies than either original population.

The dual impact of inbreeding in linkage disequilibrium on digenic traits has two aspects. On one hand, inbreeding and linkage combined increase the frequency of double homozygotes and support the frequency of double heterozygotes against declines of heterozygosity at single loci. However, the impact of linkage disequilibrium has two contrasting effects on double homozygotes and double heterozygotes. For double homozygotes, linkage disequilibrium and linkage both combine to increase frequencies to ever higher levels if either linkage or linkage disequilibrium rise.

Under inbreeding, the frequency of double heterozygotes sees the greatest increase over the analogous outbred comparisons when there is linkage disequilibrium and the loci are unlinked. While the double heterozygote frequencies still increase with linkage, they do so at a slow rate and for linked

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

loci the effects of inbreeding are nearly identical to the situation with outbred individuals with the same linkage disequilibrium at completely linked loci. This is due to the negative term in the genotype frequency becoming larger with increasing linkage (${}_1F_1^1 \rightarrow F$) and thus counteracting other positive terms and the identity disequilibrium. Therefore, for double heterozygotes with alleles in linkage disequilibrium, the largest impacts are for unlinked and partially linked loci under inbreeding. For both double homozygotes and double heterozygotes, these effects are reversed when linkage disequilibrium is negative and rare genotypes can all but disappear from the population as shown with the frequencies for zero or negative linkage disequilibrium in Figure 2.

The examples for unlinked loci are particularly cogent for digenic traits of medical interest. The traits, while often on the same chromosome, may exist in different, though related, genes and can be unlinked. Therefore identity disequilibrium would play little role in the genotype frequencies except when loci are close on the same gene and have some linkage. On the other hand, comparing the frequencies in a population with linkage disequilibrium between the loci and one that does not should show a significantly higher prevalence, in all contexts, especially if there is inbreeding.

In summary, inbreeding across two loci raises the frequencies of double homozygotes as expected from analogy with the single locus case with the gap rising with increased linkage. Double heterozygotes do not see an increase from the outbred case unless linkage disequilibrium is involved though their frequencies under inbreeding can be supported near the outbred frequency by increasing linkage.

The effect of linkage disequilibrium is to increase or decrease genotypic frequencies for positive or negative linkage disequilibrium in all cases and causes a higher frequency in double homozygotes and double heterozygotes over their outbred comparisons. However, the interaction between inbreeding and linkage disequilibrium is much more unambiguous for double homozygous loci where linkage and linkage disequilibrium combine to strongly increase frequencies of double homozygotes. For double heterozygotes, the increased genotype frequencies under inbreeding are most prominent for loci with little or no linkage.

Digenic traits, like other oligogenic traits, are being found increasingly frequently in populations where unexplained phenotypes or medical conditions show no resolution as Mendelian or complex disorders. The frequencies of these disorders in populations is not yet understood similarly to the bet-

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ter understanding of population frequencies for some Mendelian traits. The foregoing should hopefully assist in clarifying the issues regarding disease aetiology and prevalence to allow us to better understand two locus conditions as well as their one locus counterparts.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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).

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

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}

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), (August 6, 2021). World Wide Web URL: https://omim.org/

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

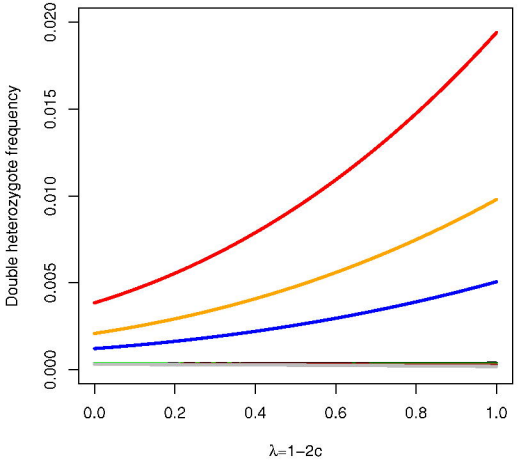
Weir, B. S. & Cockerham, C. C. 1968. The two-locus inbreeding function. North Carolina State University. Dept. of Statistics.

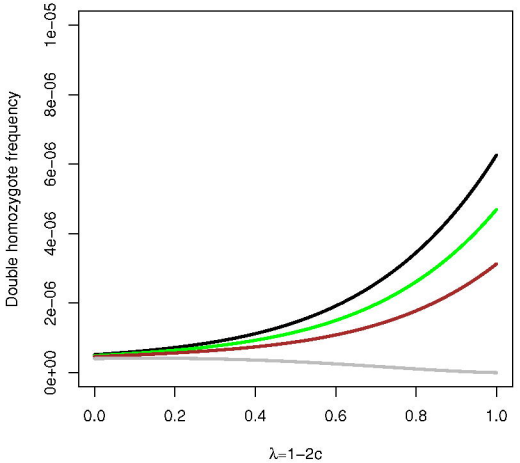
Weir, B.S. and Cockerham, C.C., 1974. Behavior of pairs of loci in finite monoecious populations. Theor. Popul. Biol., 6:323-354. doi: 10.1016/0040-5809(74)90015-X

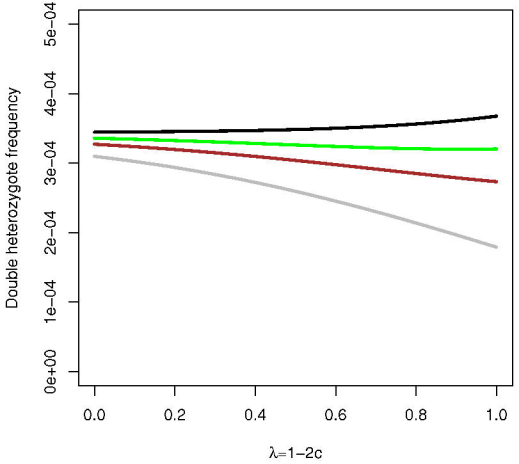
Weir, B.S., 2008. Linkage disequilibrium and association mapping. Annu. Rev. Genomics Hum. Genet., 9:129-142. doi: 10.1146/annurev.genom.9.081307.164347

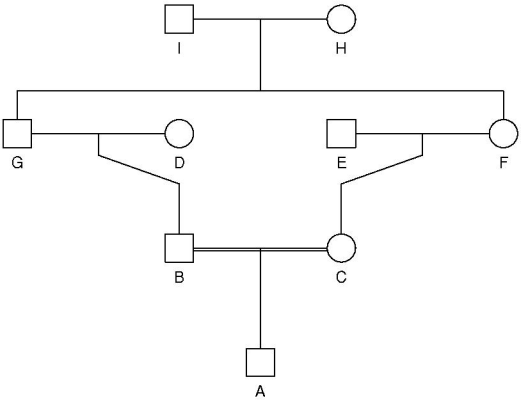
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.07.475385; this version posted January 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

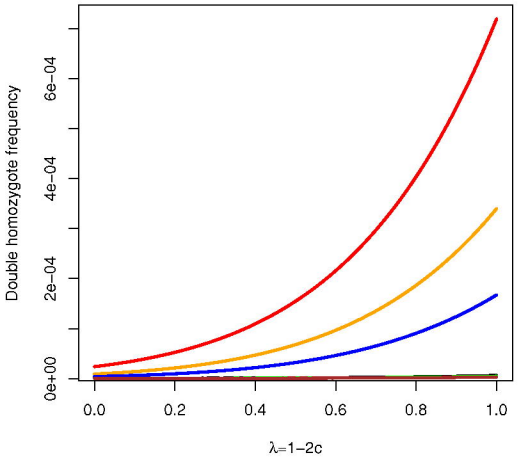
Wright, S., 1922. Coefficients of inbreeding and relationship. Am. Nat., 56:330-338. doi: 10.1086/279872


Double heterozygote frequency
0.000
0.005
0.010
0.015
0.020
0.0
0.2
0.4
0.6
0.8
1.0
λ=1−2c


Double homozygote frequency
0e+00
2e−06
4e−06
6e−06
8e−06
1e−05
0.0
0.2
0.4
0.6
0.8
1.0
λ=1−2c


Double heterozygote frequency
0e+00
1e−04
2e−04
3e−04
4e−04
5e−04
0.0
0.2
0.4
0.6
0.8
1.0
λ=1−2c


I
H
G
D
E
F
B
C
A


Double homozygote frequency
0e+00
2e−04
4e−04
6e−04
0.0
0.2
0.4
0.6
0.8
1.0
λ=1−2c