Building Blocks of Population Genetics

Jack Dekkers

Genotype frequencies, Allele frequencies, Haplotype frequencies, Linkage Disequilibrium, Linkage

Single locus → allele (or gene) **frequencies → genotype frequencies**

Consider a single locus in a random mating outbred population.
The locus has alleles A_1 and A_2 with allele (or gene) frequencies p and q

Under random mating (Hardy Weinberg Equilibrum), the allele received from one parent is *independent* of the allele received from the other parent, resulting in the following relationship between allele and genotype frequencies:

Table 1: Genotype probabilities, single locus two-allele case

	Maternal allele		
Paternal allele	***Pr(A_1) = p***	***Pr(A_2) = q***	***Marginal prob***
Pr(A_1) = p	p^2	pq	$p^2 + pq = p(p+q) = p$
Pr(A_2) = q	pq	q^2	$pq + q^2 = q(p+q) = q$
Marginal prob.	$p^2 + pq = p(p+q) = p$	$pq + q^2 = q(p+q) = q$	

This results in the HWE genotype frequencies: p^2 , $2pq$, q^2

With ***multiple*** loci we also need to consider ***haplotypes*** and their frequencies, and relationships between ***allele***, ***haplotype***, and ***genotype*** frequencies.

Haplotype = the combination of alleles at >1 locus that an individual inherited from a parent

E.g. an individual with (unordered) genotype A_1A_2 and B_1B_2 at loci A and B, can have the following combinations of haplotype pairs (separated by /):

A_1B_1/A_2B_2 → alleles A_1 and B_1 received from one parent and A_2 and B_2 from the other

A_1B_2/A_1B_2 → alleles A_1 and B_2 received from one parent and A_2 and B_1 from the other

Haplotype frequency = frequency of a given haplotype in a population

With two loci with two alleles, there are 4 possible haplotypes, 16 ordered genotypes (ordered based on haplotypes), and 9 unordered genotypes (see tables 2,3)

Table 2: Haplotype frequencies and genotype frequencies under random mating (HWE)

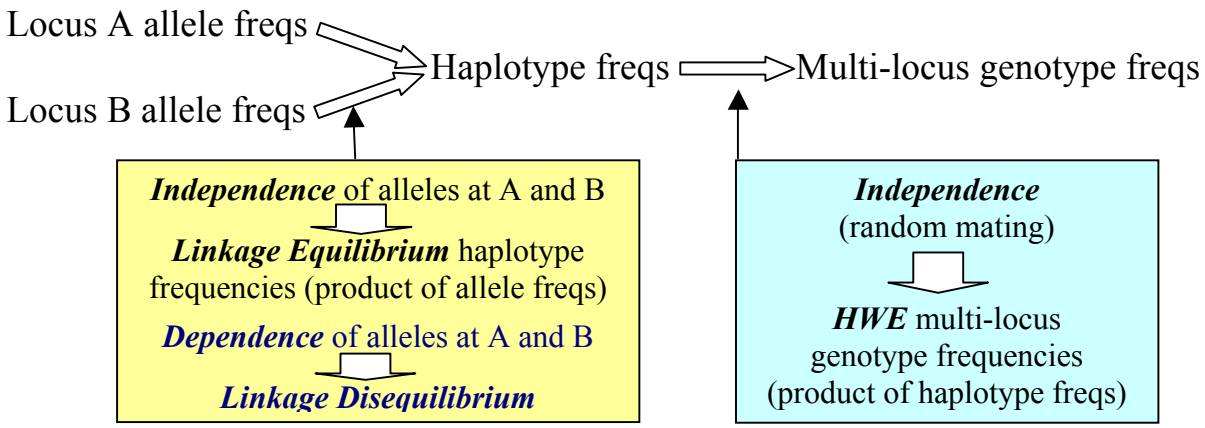
Haplotype - freq			***Maternal haplotype***							
			A_1B_1	r	A_1B_2	s	A_2B_1	t	A_2B_2	u
Paternal haplotype	A_1B_1	r	A_1B_1/A_1B_1	r^2	A_1B_1/A_1B_2	rs	A_1B_1/A_2B_1	rt	A_1B_1/A_2B_2	ru
	A_1B_2	s	A_1B_2/A_1B_1	sr	A_1B_2/A_1B_2	s^2	A_1B_2/A_2B_1	st	A_1B_2/A_2B_2	su
	A_2B_1	t	A_2B_1/A_1B_1	tr	A_2B_1/A_1B_2	ts	A_2B_1/A_2B_1	t^2	A_2B_1/A_2B_2	tu
	A_2B_2	u	A_2B_2/A_1B_1	ur	A_2B_2/A_1B_2	us	A_2B_2/A_2B_1	ut	A_2B_2/A_2B_2	u^2

Table 3: Unordered and ordered genotypes and their frequencies under random mating

Unordered genotypes	**Frequency** =sum of ordered frequencies	**Possible ordered genotypes** and their frequencies (from Table 1) 'ordered' based on parental origin (paternal haplotype/maternal haplotype)									
$A_1A_1B_1B_1$	r^2	A_1B_1/A_1B_1	r^2								
$A_1A_1B_1B_2$	$2rs$	A_1B_1/A_1B_2	rs	A_1B_2/A_1B_1	sr						
$A_1A_1B_2B_2$	s^2	A_1B_2/A_1B_2	s^2								
$A_1A_2B_1B_1$	$2rt$	A_1B_1/A_2B_1	rt	A_2B_1/A_1B_1	tr						
$A_1A_2B_1B_2$	$2ru+2st$	A_1B_1/A_2B_2	ru	A_1B_2/A_2B_1	st	A_2B_1/A_1B_2	ts	A_2B_2/A_1B_1	ur		
$A_1A_2B_2B_2$	$2su$	A_1B_2/A_2B_2	su	A_2B_2/A_1B_2	us						
$A_2A_2B_1B_1$	t^2	A_2B_1/A_2B_1	t^2								
$A_2A_2B_1B_2$	$2tu$	A_2B_1/A_2B_2	tu	A_2B_2/A_2B_1	ut						
$A_2A_2B_2B_2$	u^2	A_2B_2/A_2B_2	u^2								

The unordered genotype is what is obtained from genotyping, i.e. the genotype at each locus

What is the relationship between haplotype frequencies and the frequencies of alleles that make up each haplotype? This depends on whether the alleles at the two loci are <u>*dependent*</u> or <u>*independent*</u>:



Haplotype probabilities (= frequencies) - **two-allele case**:
What is the probability of a progeny to receive from a parents: allele A_i at locus A
and allele B_j at locus B ?

i) if the alleles at the two loci are <u>independent</u> from each other

➔ joint probability = product of marginal probabilities

Locus B allele freq's	***Locus A*** – allele frequencies $Pr(A_1) = p_A$	$Pr(A_2) = q_A$	***Marginal prob***
$Pr(B_1) = p_B$	$\Pr(A_1B_1) = p_Ap_B$	$\Pr(A_2B_1) = q_Ap_B$	$p_Ap_B + q_Ap_B = p_B\,(p_A + q_A) = p_B$
$Pr(B_2) = q_B$	$\Pr(A_1B_2) = p_Aq_B$	$\Pr(A_2B_2) = q_Aq_B$	$p_Aq_B + q_Aq_B = q_B\,(p_A + q_A) = q_B$
Marginal prob	$p_Ap_B + p_Aq_B = p_A(p_B + q_B) = p_A$	$q_Ap_B + q_Aq_B = q_A(p_B + q_B) = q_A$	

ii) What if the alleles at the two loci are <u>NOT</u> independent ?

➔ joint probabilities deviate from product of marginal probabilities (by <u>±</u>D)

Locus B	***Locus A*** $Pr(A_1) = p_A$	$Pr(A_2) = q_A$	***Marginal prob***
$Pr(B_1) = p_B$	$\Pr(A_1B_1) = r = p_Ap_B + D$	$\Pr(A_2B_1) = t = q_Ap_B - D$	$p_Ap_B + D + q_Ap_B - D = p_B\,(p_A + q_A) = p_B$
$Pr(B_2) = q_B$	$\Pr(A_1B_2) = s = p_Aq_B - D$	$\Pr(A_2B_2) = u = q_Aq_B + D$	$p_Aq_B - D + q_Aq_B + D = q_B\,(p_A + q_A) = q_B$
Marginal prob	$p_Ap_B + D + p_Aq_B - D = p_A(p_B + q_B) = p_A$	$q_Ap_B - D + q_Aq_B + D = q_A(p_B + q_B) = q_A$	

If alleles are dependent ➔ loci are in **Linkage Disequilibrium**
or in **Gametic Phase Disequilibrium**

The term 'linkage' in Linkage disequilibrium is actually not quite correct and a bit misleading because disequilibrium can occur between <u>unlinked</u> loci, although it is more likely to be present (and persist) between linked loci (see later). Thus, 'Gametic phase' disequilibrium is a better term; gametic phase refers to the haploid phase of chromosomes and disequilibrium refers to dependence between alleles that make up the haplotypes that are present in the current generation and which originated from the haploid gametes produced by their parents.

D = measure of disequilibrium = $D = r - p_Ap_B$

$\Pr(A_1B_1) - \Pr(A_1)\Pr(B_1)$

The value obtained for $|D|$ is the same irrespective of the haplotype used.

You can also calculate D as: $D = \frac{1}{2}\left[\Pr\left(\frac{A_1B_1}{A_2B_2}\right) - \Pr\left(\frac{A_1B_2}{A_2B_1}\right)\right] = ru - st$

Coupling heterozygote

Repulsion heterozygote

Other measures of LD:

D' = D standardized to make it less dependent on allele frequencies

$D' = D/D_{max}$ where $D_{max} = \text{Min}(p_A p_B , q_A q_B)$ if $D<0$

$D_{max} = \text{Min}(p_A q_B , q_A p_B)$ if $D>0$

r^2 = squared correlation between allele at locus A and allele at locus B

- also measures ability (R^2) to predict allele at locus A from allele at locus B

To derive r^2: Let $X = 1$ when allele A_1 present, $X = 0$ if A_2 present (= Bernoulli var.)

$Y = 1$ when allele B_1 present, $Y = 0$ if B_2 present (= Bernoulli var.)

Then: $\text{cov}(X,Y) = E(XY) - E(X)\,E(Y)$

$= \quad r \quad - \quad p_A \; p_B \quad = D$

	A_1 X=1	A_2 X=0
B_1 Y=1	$\Pr(A_1B_1) = r$ $XY = 1$	$\Pr(A_2B_1) = t$ $XY = 0$
B_2 Y=0	$\Pr(A_1B_2) = s$ $XY = 0$	$\Pr(A_2B_2) = u$ $XY = 0$

➔ $\text{Corr.} = r_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} = \frac{D}{\sqrt{(p_A q_A)(p_B q_B)}}$

➔ $r^2 = r_{XY}^2 = \frac{D^2}{p_A q_A p_B q_B}$ (Note: this r is different than r in the table above)

$|D'|$ and r^2 range between 0 and 1

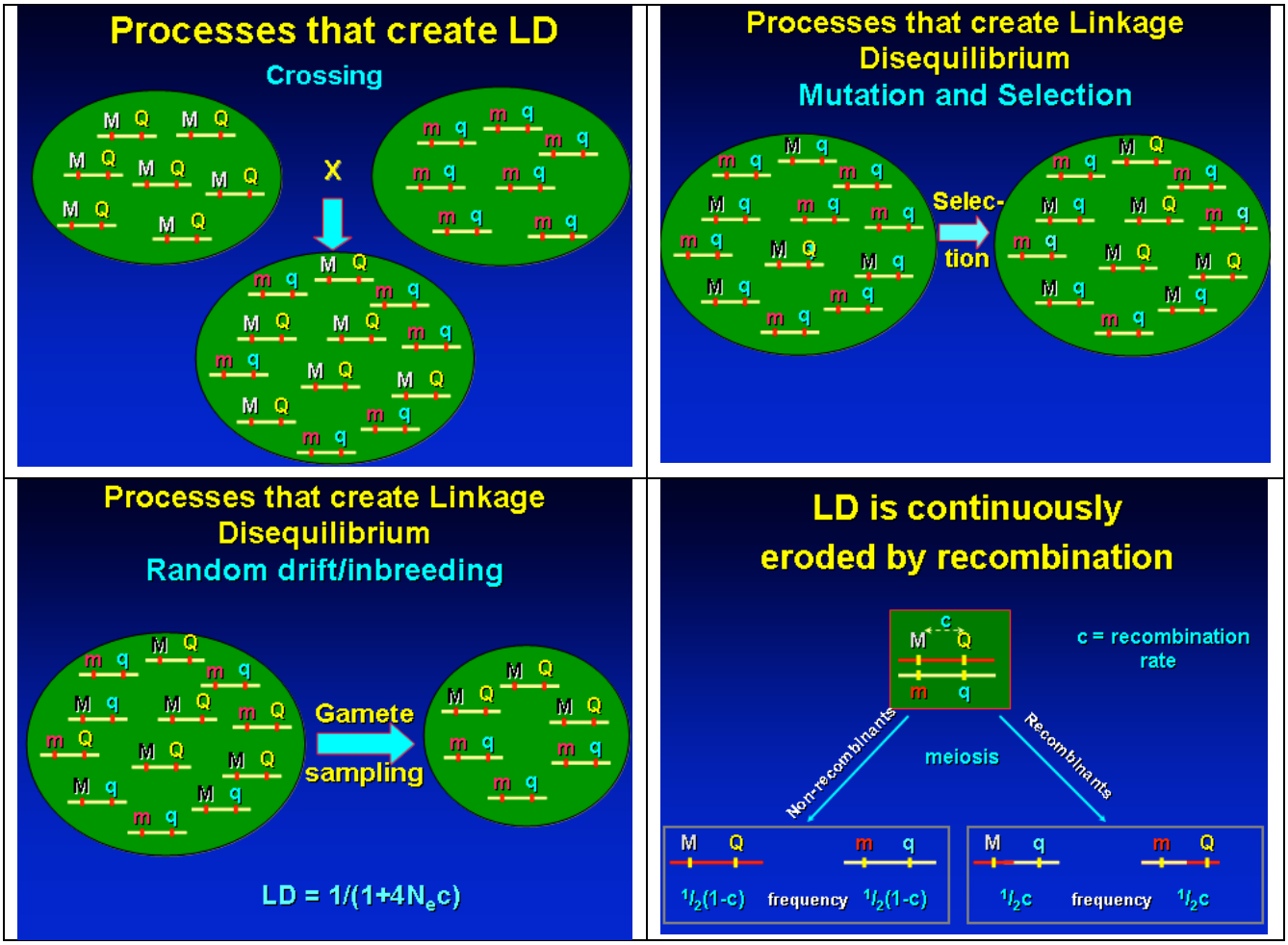
$|D'|$ is strongly inflated if one haplotype has a very low frequency

r^2 is the preferred measure of LD for most uses

Mechanisms that generate Linkage Disequilibrium (LD)

A variety of mechanisms generate linkage disequilibrium, and several of these can operate simultaneously. They can be separated into:

1. **Recurrent factors** – operate to create LD each generation
   a. ***Drift*** (inbreeding) in small populations – by chance or sampling, haplotypes passed on to the next generation are not in LE frequencies
   b. ***Recurrent migration*** – continuous mixing of populations in which haplotypes occur in different frequencies (e.g. $Pr(A_1B_1)$=1 for pop. 1 and =0 for pop. 2)
   c. ***Selection*** – certain haplotypes may be selected upon and increase in frequency
      – selection creates LD between loci that are selected upon (= Bulmer effect)
      – selection with epistasis (certain combinations of alleles are favorable) also creates LD between loci involved.
2. **Punctual factors** – operate only sporadically over time to create LD
   a. ***Mutation*** – occurs in a specific haplotype, which is then the only haplotype that contains that mutation, resulting it to be in LD with the mutation.
   b. ***One-time admixture/migration/crossing*** (e.g. producing F_1/F_2) – results in mixing populations with different haplotype frequencies
   c. ***Population bottleneck / founder effects*** – severe drift from 1-time sampling effects



LD is continuously eroded by recombination: how does D change over time?

Let r_0 = frequency of A_1B_1 haplotypes in generation 0 → $D_0 = r_0 - p_A p_B$

What is the frequency of A_1B_1 haplotypes in generation 1?

In the following derivation, we will consider parental origin of haplotypes and will let $_\bullet$ indicate 'any' allele, so $A_1B_1/A_\bullet B_\bullet$ indicates an individual that received the A_1B_1 from its father and any haplotype (A_1B_1 or A_1B_2 or A_2B_1 or A_2B_2) from its mother)

There are four ways that parents from generation 0 can generate gametes that carry the A_1B_1 haplotype and that will produce generation 1:

1. non-recombinant A_1B_1 haplotype produced by a $A_1B_1/A_\bullet B_\bullet$ parent
2. non-recombinant A_1B_1 haplotype produced by a $A_\bullet B_\bullet/A_1B_1$ parent
3. recombinant A_1B_1 haplotype produced by a $A_1B_\bullet/A_\bullet B_1$ parent
4. recombinant A_1B_1 haplotype produced by a $A_\bullet B_1/A_1B_\bullet$ parent

For case 1, the frequency of $A_1B_1/A_\bullet B_\bullet$ parents is r_0. The frequency of non-recombinant A_1B_1 haplotypes produced by these parents is ½(1-c). Thus, the frequency of A_1B_1 haplotype produced by $A_1B_1/A_\bullet B_\bullet$ parents = Prob(1.) = ½(1-c)r_0.

Case 2 results in the same frequency: Prob(2.) = ½(1-c)r_0.

For case 3, the frequency of $A_1B_\bullet/A_\bullet B_1$ parents is $p_A p_B$. The frequency of recombinant A_1B_1 haplotypes produced by these parents is ½c, so the overall frequency is ½$cp_A p_B$.

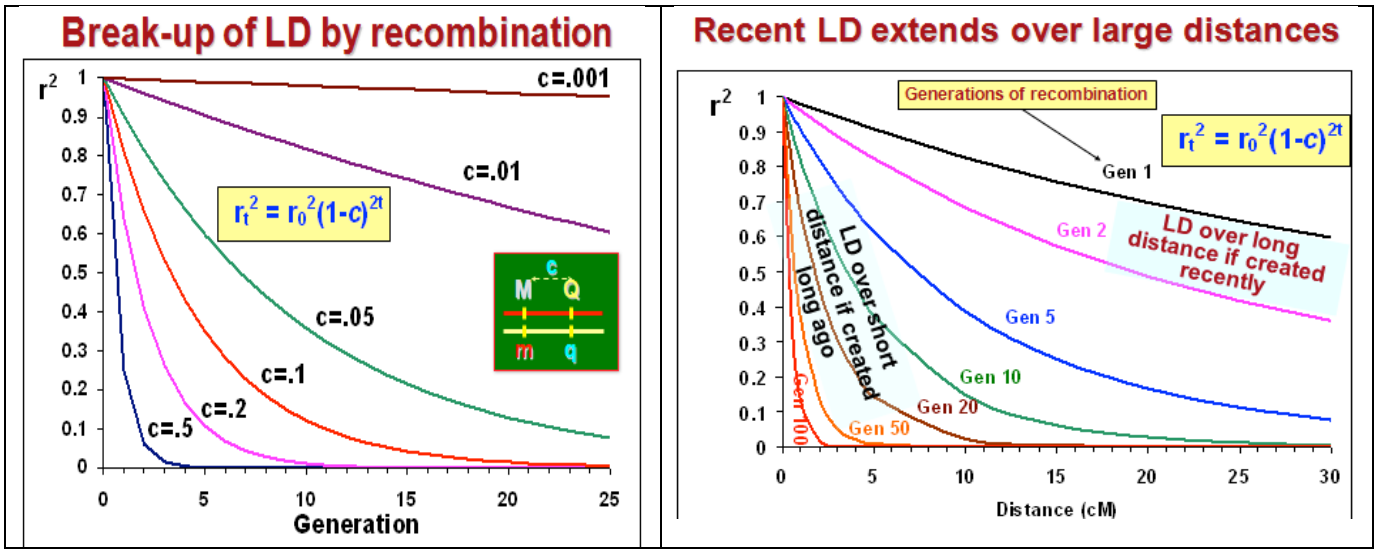
Case 4.results in the same frequency: Prob(4.) = ½$cp_A p_B$..

Thus, the overall frequency of A_1B_1 gametes produced by generation 0 is the some of these four mutually exclusive cases:

- → $r_1 = r_0(1-c) + p_A p_B\, c$
- → $D_1 = r_1 - p_A p_B = r_0(1-c) + p_A p_B c - p_A p_B = r_0(1-c) - p_A p_B(1-c) = (r_0 + p_A p_B)(1-c) = D_0(1-c)$
- → $D_2 = D_1(1-c) = \{D_0(1-c)\}\ (1-c) = D_0(1-c)^2$
- → $D_t = D_0(1-c)^t$ → $D_\infty = 0$

→ Erosion of LD by recombination occurs faster when loci are further apart.
LD is halved each generation if loci are unlinked (c = ½).

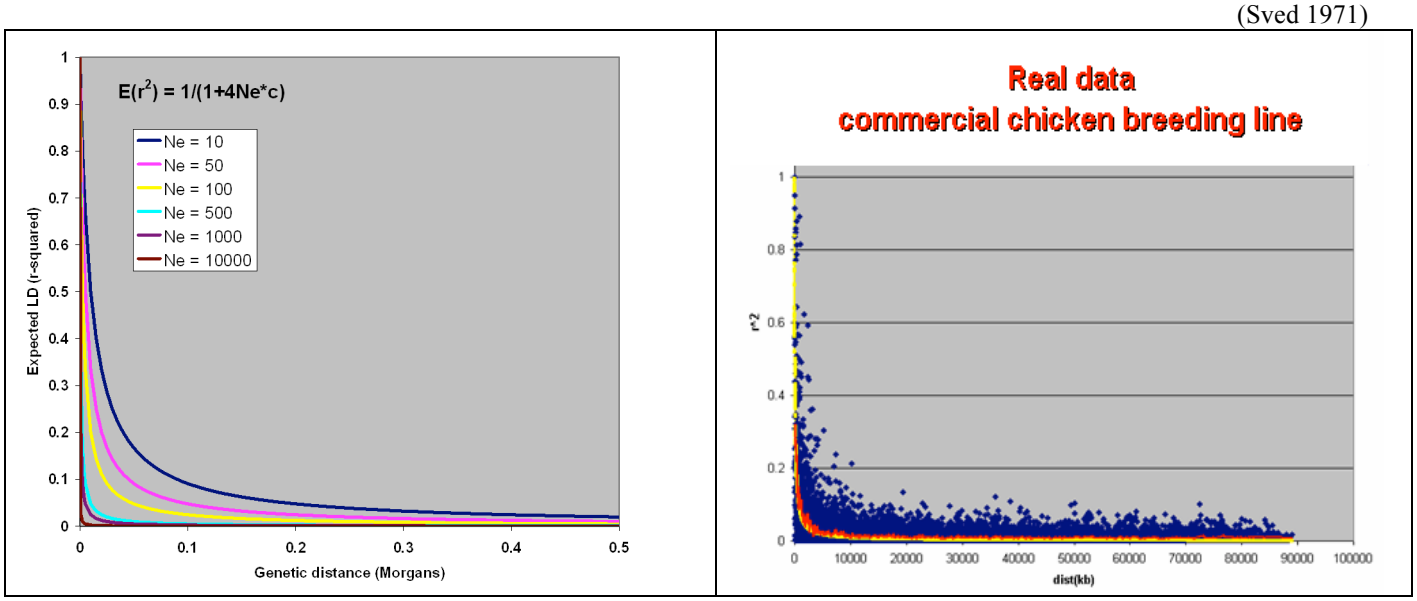
Since $r^2 = \dfrac{D^2}{p_A q_A p_B q_B}$, LD measured by r^2 will decline at a rate of $(1-c)^2$ per generation: $\mathbf{r_t^2 = r_0^2(1-c)^{2t}}$



Balance between drift and recombination: in *small(er) closed populations*

- LD is continuously created by drift (sampling) (small effective population size, N_e)
- LD is continuously eroded by recombination – faster at longer distances

This results in a balance/equilibrium of average LD at a given distance: $E(r^2_{\infty,c}) = \frac{1}{1+4N_e c}$ (Sved 1971)



Most outbred domesticated plant and animal populations have small(er) (historical) effective population size and drift-recombination balance is expected to be the main contributor to LD ➔ LD expected to be sizeable at short distances, but small at longer distances.

Most human populations have large (historical) N_e ➔ $E(r^2_{\infty,c}) = \frac{1}{1+4N_e c}$ is smaller at given distance.

Building Blocks of Quantitative Genetics

Jack Dekkers

1 locus case:

Genotype A_2A_2 A_1A_2 A_1A_1

Genotypic value $\mu-a$ μ $\mu+d$ $\mu+a$ μ = mid-homozygote value

Single-locus frequencies, genotypic values (deviated from μ), and expectation

T	Frequency, $\Pr(T)$	Genotypic value, G_T	$\Pr(T)\times G_T$
A_1A_1	p^2	a	p^2a
A_1A_2	$2pq$	d	$2pqd$
A_2A_2	q^2	a	$-q^2a$

Population mean $= \mu + E(G_T) = M = \mu + p^2a + 2pqd + -q^2a = \mu + a(p-q) + 2pqd$

Extension to two loci (*without* epistasis):

The genotypic value of an individual is the sum of the genotypic values at each locus:

$G_T = \mu + G_A + G_B$ G_i = genotypic value locus i, as defined for 1-locus case

Now the homozygote "midpoint" μ is midway between the best and worst double homozygote ($A_1A_1B_1B_1$ and $A_2A_2B_2B_2$) = (17+3)/2 = 10 in the example below.

Pop. mean $= \mu + E(G_T) = M = \mu + E(G_A + G_B) = E(G_A) + E(G_B)$

$= \mu + \{a_A(p_A - q_A) + 2p_Aq_Ad_A\} + \{a_B(p_B - q_B) + 2p_Bq_Bd_B\}$

Spreadsheet Genotypic_values_models_v10.xls

Input parameters:

Homozygote midpoint μ =		**10**	
a_A =	**4**	a_B =	**3**
d_A =	**2**	d_B =	**-1**
p_A =	**0.6**	p_B =	**0.3**
q_A =	**0.4**	q_B =	**0.7**
Linkage Disequilibrium D =		**0**	
Recomb. Rate =	**0.2**		

Input matrix for epistatic effects	A_1A_1	A_1A_2	A_2A_2
B_1B_1	**0**	**0**	**0**
B_1B_2	**0**	**0**	**0**
B_2B_2	**0**	**0**	**0**

Output:

GENOTYPE-BASED MODEL FOR GENOTYPIC VALUES

2-locus genotypic values and frequencies (random mating)			A locus genotype		
			A_1A_1	A_1A_2	A_2A_2
	B locus genotype	$\mu+G_T$	**4**	**2**	**-4**
		freq	*0.36*	*0.48*	*0.16*
	B_1B_1	**3**	**17**	**15**	**9**
		0.09	*0.0324*	*0.0432*	*0.0144*
	B_1B_2	**-1**	**13**	**11**	**5**
		0.42	*0.1512*	*0.2016*	*0.0672*
Population mean M = **10.14**	B_2B_2	**-3**	**11**	**9**	**3**
		0.49	*0.1764*	*0.2352*	*0.0784*
new μ = **10**					
μ_A = 8.38	Average at A locus		12.38	10.38	4.38
μ_B = 11.76	Average at B locus		14.76	10.76	8.76
Re-calculated 1-locus additive, dominance			+a	d	-a
and genotypic values		G_A =	**4**	**2**	**-4**
with epistasis		G_B =	**3**	**-1**	**-3**

Extension to many loci: $G_T = \Sigma G_i$ summation is over all loci

Homozygote "midpoint" μ is average of the best and the worst multi-homozygote

Population mean = $\mu + E(G_T) = M$ $= \mu + E(\Sigma G_i) = \mu + \Sigma E(G_i)$

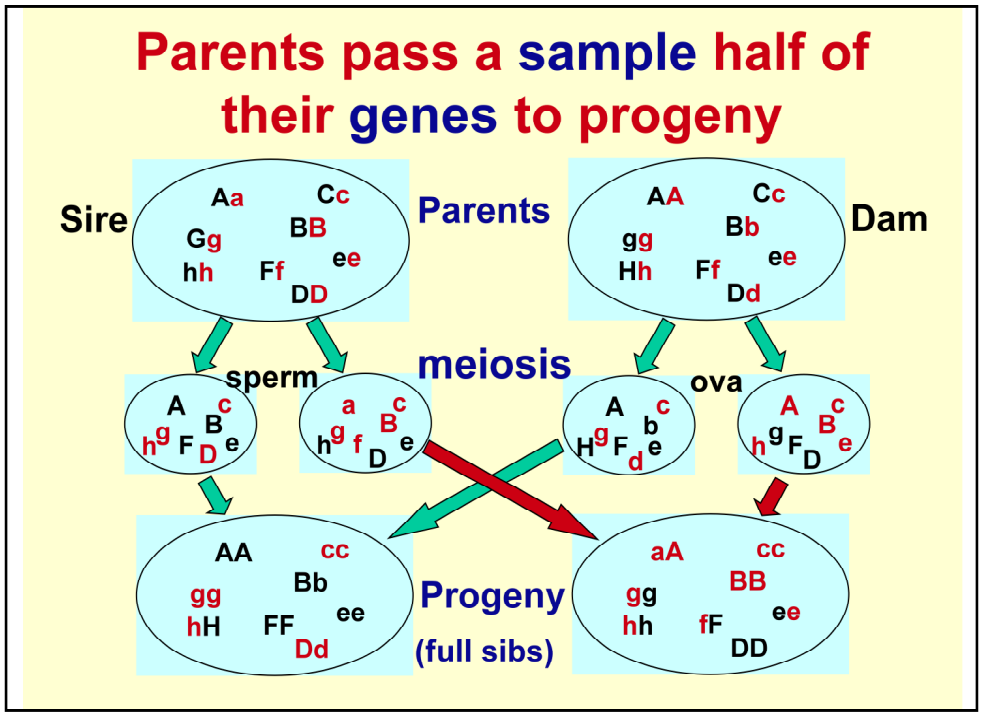
$= \mu + \Sigma\{a_i(p_i - q_i) + 2p_iq_id_i\} = \Sigma a_i(p_i - q_i) + 2\Sigma p_iq_id_i$

Allele-based models for additive effects

In practice, we are interested in selecting the 'best' individuals to be used as parents to breed the next generation; we want to select individuals whose progeny have the highest expected phenotype, i.e. whose progeny have the highest expected genotypic value.

To identify these individuals we need to know how the genotypic value of progeny relates to the genotypic value of their parents.

Models described in terms of *a* and *d* are for genotypic values for whole genotypes.
But individuals pass on alleles NOT genotypes



The breeding value of an individual is defined to quantify an individual's value as a parent. It is related to the expected genotypic or phenotypic values of that individual's progeny.

An individual's **breeding value** = 2 x expected deviation of the mean phenotype of an individual's progeny from population mean (M) *when mated at random to other individuals from the population.*

$$A_i = 2\,E(P_{progeny} - M)$$

In general: an individual's breeding value is the sum of the average effects of the alleles that the individual carries: $A_{ij} = \alpha_i + \alpha_j$

Average effect α_i = Average deviation from the population mean of individuals who received allele *i* (i.e. A_1 or A_2) from one parent and the other allele at random (i.e. A_1 with freq. *p* and A_2 with freq. *q*)

"Allele *i*"	"Other Allele" (=random) A_1 Pr(A_1) = *p*	A_2 Pr(A_2) = *q*	Mean G_T of resulting individuals	Mean of G_T deviated from Population mean
A_1	$G_{A1A1} = a$	$G_{A1A2} = d$	$pa + qd$	$\alpha_1 = pa + qd - M$ $= q[a + (q-p)d] = q\alpha$
A_2	$G_{A2A1} = d$	$G_{A2A2} = -a$	$pd - qa$	$\alpha_2 = pd - qa - M$ $= -p[a + (q-p)d] = -p\alpha$

Another important concept/quantity is the average allele substitution effect (α):

Average allele substitution effect = $\alpha = \alpha_1 - \alpha_2 = a + (q - p)d$

= average effect on the genotypic value of substituting a random A_2 allele for an A_1 allele

One locus multiple alleles: an individual's breeding value is the sum of the average effects of the alleles that the individual carries: $A_{ij} = \alpha_i + \alpha_j$

For two loci: individual with alleles *i* and *j* at locus A and alleles *k* and *l* at locus B:

$A_{ijkl} = \alpha_{Ai} + \alpha_{Aj} + \alpha_{Bk} + \alpha_{Bl}$ with each α_{ni} derived as above for 1-locus case *(no epistasis)*

Many loci: $A_{ijkl} = \sum_{locus\ l=1}^{n} \sum_{allele\ i=1}^{2} \alpha_{li}$ sum average effects over all *n* loci and the individual's two alleles at each locus

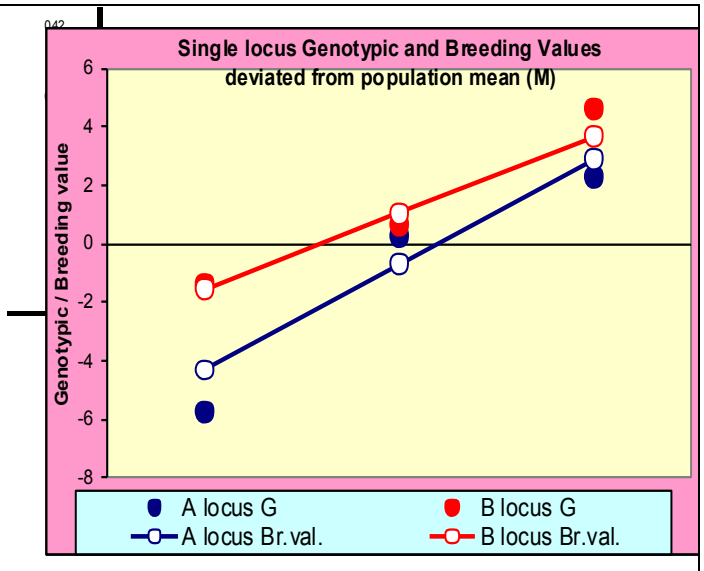
Spreadsheet 'Genotypic_value_models.v10.xls'

ALLELE-BASED MODEL FOR GENOTYPIC VALUES

	Average allele effects		Substitution effect
Locus A	α_{A1} = 1.44	α_{A2} = -2.16	α_A = 3.6
Locus B	α_{B1} = 1.82	α_{B2} = -0.78	α_B = 2.6

All values are now deviated from the population mean, M.

		A locus genotype		
		A_1A_1	A_1A_2	A_2A_2
B locus Genotype	($+G_T$)-*M*	2.24	0.24	-5.76
	Additive	2.88	-0.72	-4.32
	Dom.dev.	*-0.64*	*0.96*	*-1.44*
	Epistasis	--	--	--
B_1B_1	4.62	6.86	4.86	-1.14
	3.64	6.52	2.92	-0.68
	0.98	0.34	1.94	-0.46
	–	*0.00*	*0.00*	*0.00*
B_1B_2	0.62	2.86	0.86	-5.14
	1.04	3.92	0.32	-3.28
	-0.42	-1.06	0.54	-1.86
	–	*0.00*	*0.00*	*0.00*
B_2B_2	-1.38	0.86	-1.14	-7.14
	-1.56	1.32	-2.28	-5.88
	0.18	-0.46	1.14	-1.26
	–	*0.00*	*0.00*	*0.00*



Alternate derivation of allele substitution effect based on

Linear regression on number of '1' alleles

Allele substitution effects can also be derived by analyzing phenotype (or the genotypic value) by a linear regression model on the number of '1' alleles that an individual carries (as we have done is some of the homeworks):

Linear regression of Y on X: $Y = M + \hat{b}_{YX}(X - \bar{X}) + e$

$$\hat{Y} = M + \hat{b}_{YX}(X - \bar{X})$$

In this case, X = # 1 alleles: $X_T = \{2, 1, 0\}$ for $T = \{A_1A_1, A_1A_2, A_2A_2\}$

Genotype T	G_T	# A_1 X_T	Frequency f	$f*X_T^2$	$F*G_T*X_T$
A_1A_1	a	2	p^2	$4p^2$	$2p^2a$
A_1A_2	d	1	$2pq$	$2pq$	$2pqd$
A_2A_2	-a	0	q^2	0	0
SUM				$4p^2+2pq$	$2p^2a+2pqd$

$$\hat{b} = \frac{\text{cov}(G_T, X_T)}{\text{var}(X_T)} = \frac{E(G_T X_T) - E(G_T)E(X_T)}{E(X_T^2) - E(X_T)^2} = \frac{\left[2p^2a + 2pqd\right] - \left[a(p-q) + 2pqd\right]\left[2p^2 + 2pq\right]}{\left[4p^2 + 2pq\right] - \left[2p^2 + 2pq\right]^2}$$

Interpretation of regression coefficient $\hat{b}$:

- When X_T increases by 1, $\hat{Y}$ increases by $a + (q - p)d = \alpha$ = allele substitution effect
- When X_T increases by 1, an allele substitution has occurred ➔ $\hat{b} = \alpha$

Allele-based models for dominance and epistatic effects

When d is not 0, breeding values will not explain everything about the genotypic value:

Single locus example: p=0.6; a=+4; d=1; M=0 ➔ $\alpha = a + (q - p)d$ = +3.8

➔ $\alpha_1 = q\alpha$ = 0.4*3.8 =+1.52

➔ $\alpha_2 = -p\alpha$ = -0.6*3.8 = -2.28

Genotype T	Fre-quency	Genotypic value (G) deviated from M	Breeding value A	Dominance deviation $\delta = G-A$
A_1A_1	0.36	+2.72	$2\alpha_1$ = +3.04	-0.32
A_1A_2	0.48	-0.28	$\alpha_1 + \alpha_2$ = -0.76	+0.48
A_2A_2	0.16	-5.28	$2\alpha_2$ = -4.56	-0.72

These differences between the single-locus genotypic and breeding values are called '***dominance deviations***'.

Based on this, the genotypic value at a single locus of an individual that has alleles *i* and *j* at that locus (deviated from the population mean, *M*) can be written as:

$G_{ij} = \alpha_i + \alpha_j + \delta_{ij} \;=\; A_{ij} \;+ \delta_{ij}$

α_i = average effect of allele *i* $\quad$ α_j = average effect of allele *j*

δ_{ij} = dominance deviation effect of the interaction of alleles *i* and *j*

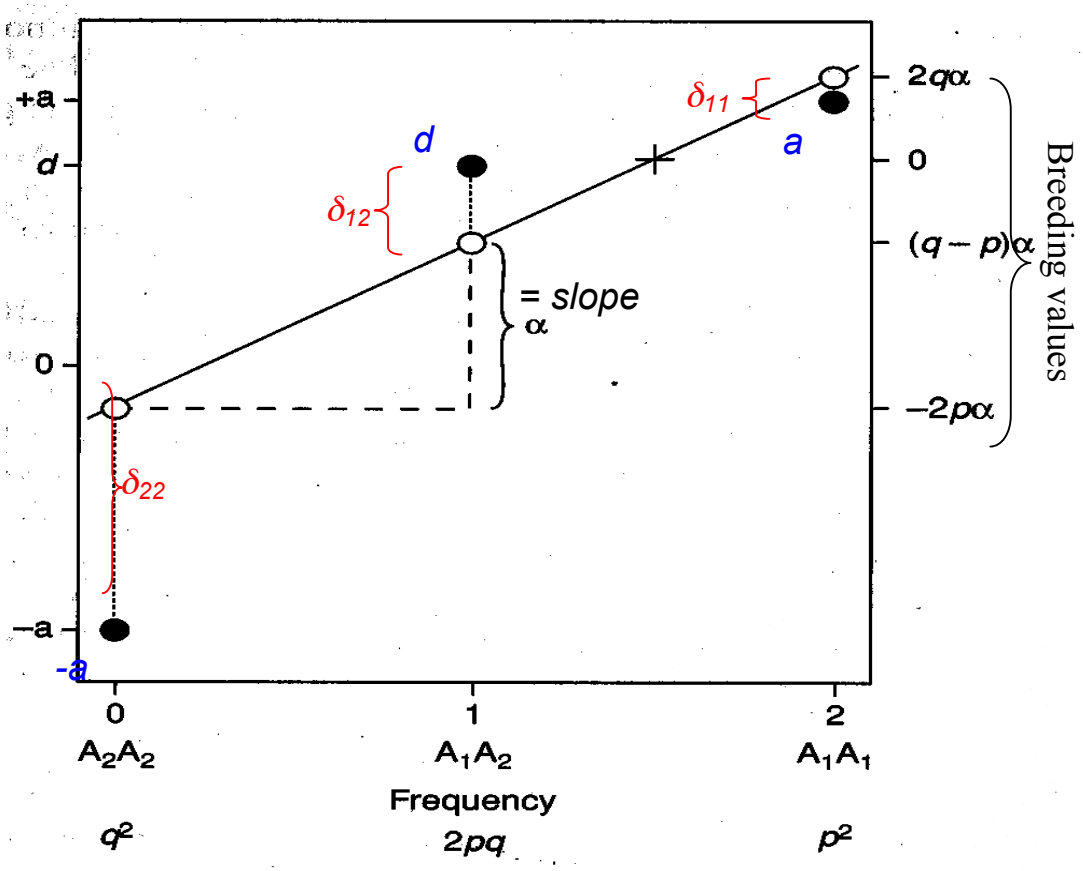
Dominance devations can also be derived as a function of allele frequencies and *d*

Genotype *T*	G_T *(deviated from μ)*	*Average allele effects*		*Dom.dev.*
		α_i	α_j	δ_{ij}
A_1A_1	a	$\alpha_1 = q\alpha$	$\alpha_1 = q\alpha$	$-2q^2d$
A_1A_2	d	$\alpha_1 = q\alpha$	$\alpha_2 = -p\alpha$	$2pqd$
A_2A_2	-a	$\alpha_2 = -p\alpha$	$\alpha_2 = -p\alpha$	$-2p^2d$

Graphical representation of average effects and dominance deviations

Dominance deviations are the residuals from the regression of genotypic values on the number of A_1 alleles.

The regression line represents the breeding values



Extension to two loci (no epistasis)

With two loci, the genotypic value is the sum of the individual's genotypic value at each locus: $G_T = G_A + G_B$ G_i = genotypic value locus *i*, as defined for 1-locus case

And the genotypic value at each locus can be partitioned into additive and dominance effects:

$$G_{Aij} = \alpha_{Ai} + \alpha_{Aj} + \delta_{Aij}$$

and $$G_{Bij} = \alpha_{Bi} + \alpha_{Bj} + \delta_{Bij}$$

Thus the overall genotypic value can be written as the sum of average allele effects and dominance deviations as:

$$G_T = \alpha_{Ai} + \alpha_{Aj} + \delta_{Aij} + \alpha_{Bi} + \alpha_{Bj} + \delta_{Bij}$$

$$G_T = \alpha_{Ai} + \alpha_{Aj} + \alpha_{Bi} + \alpha_{Bj} + \delta_{Aij} + \delta_{Bij}$$

The sum of average allele effects define the ***breeding value***: $A_T = \alpha_{Ai} + \alpha_{Aj} + \alpha_{Bi} + \alpha_{Bj}$

The sum of dominance deviations define the ***dominance effect***: $D_T = \delta_{Aij} + \delta_{Bij}$

Thus the genotypic value can be written as the breeding value and its dominance effect:

$$G_T = A_T + D_T$$

Homozygote midpoint =	**10**
a_A = **4**	a_B = **3**
d_A = **2**	d_B = **-1**
p_A = **0.6**	p_B = **0.3**
q_A = **0.4**	q_B = **0.7**
Linkage Disequilibrium D =	**0**
Recomb. Rate =	**0.2**

Input matrix for epistatic effects

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	0	0	0
B_1B_2	0	0	0
B_2B_2	0	0	0

ALLELE-BASED MODEL FOR GENOTYPIC VALUES

	Average allele effects		Substitution effect
Locus A	α_{A1} = 1.44	α_{A2} = -2.16	α_A = 3.6
Locus B	α_{B1} = 1.82	α_{B2} = -0.78	α_B = 2.6

All values are now deviated from the population mean, M.

		A locus genotype		
		A_1A_1	A_1A_2	A_2A_2
	($+G_T$)-*M*	**2.24**	**0.24**	**-5.76**
	Additive	2.88	-0.72	-4.32
B locus	*Dom.dev.*	*-0.64*	*0.96*	*-1.44*
Genotype	*Epistasis*	--	--	--
B_1B_1	**4.62**	**6.86**	**4.86**	**-1.14**
	3.64	6.52	2.92	-0.68
	0.98	0.34	1.94	-0.46
	–	*0.00*	*0.00*	*0.00*
B_1B_2	**0.62**	**2.86**	**0.86**	**-5.14**
	1.04	3.92	0.32	-3.28
	-0.42	-1.06	0.54	-1.86
	–	*0.00*	*0.00*	*0.00*
B_2B_2	**-1.38**	**0.86**	**-1.14**	**-7.14**
	-1.56	1.32	-2.28	-5.88
	0.18	-0.46	1.14	-1.26
	–	*0.00*	*0.00*	*0.00*

Extension to multiple loci (no epistasis):

$$G_T = A_T + D_T \qquad A_T = \sum_{locus\ l=1}^{n} \sum_{allele\ i=1}^{2} \alpha_{li} \qquad D_T = \sum_{locus\ l=1}^{n} \delta_{lij}$$

Epistatic Deviations

When epistatic effects are present, the genotypic value of an individual can not be written as a simple sum of the genotypic value at each locus but an effect of the interaction between loci needs to be added: For two loci: $G_T = G_A + G_B + G_{AxB}$

Similarly, the genotypic value of an individual can also not be written as the sum of a breeding value and a dominance deviation but an epistatic deviation effect (I_T) needs to be added: $G_T = A_T + D_T + I_T$

Epistatic deviation effects for each individual (or multi-locus genotype) can be calculated by subtraction, after the additive and dominance deviation effects have been computed as described before: $I_T = G_T - A_T - D_T$

Homozygote midpoint =	10	
a_A =	4	a_B = 3
d_A =	2	d_B = -1
p_A =	0.6	p_B = 0.3
q_A =	0.4	q_B = 0.7

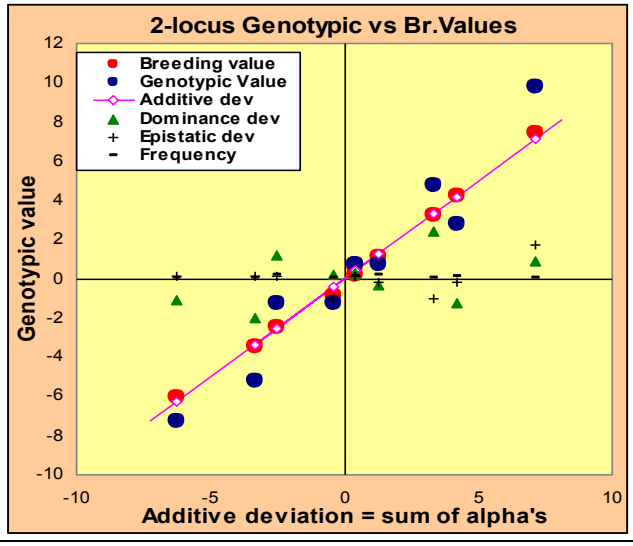
Linkage Disequilibrium D =	0
Recomb. Rate =	0.2

Input matrix for epistatic effects

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	3	0	0
B_1B_2	0	0	0
B_2B_2	0	0	0

Genotype-based epistatic effects G_{AxB}

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	0.825	-1.905	-1.905
B_1B_2	-1.095	-0.825	-0.825
B_2B_2	-1.095	-0.825	-0.825



ALLELE-BASED MODEL FOR GENOTYPIC VALUES

	Average allele effects		Substitution effect
Locus A	α_1 = 1.5048	α_2 = -2.2572	α = 3.762
Locus B	β_1 = 2.0468	β_2 = -0.8772	β = 2.924

		A locus genotype		
		A_1A_1	A_1A_2	A_2A_2
	($+G_T$)-*M*	**2.41**	**0.14**	**-5.86**
	Additive	3.01	-0.75	-4.51
B locus	*Dom.dev.*	*-0.60*	*0.90*	*-1.34*
Genotype	*Epistasis*	--	--	--
B_1B_1	**5.60**	**9.76**	**4.76**	**-1.24**
	4.09	7.10	3.34	-0.42
	1.51	0.91	2.40	0.17
	--	*1.75*	*-0.98*	*-0.98*
B_1B_2	**0.52**	**2.76**	**0.76**	**-5.24**
	1.17	4.18	0.42	-3.34
	-0.65	-1.24	0.25	-1.99
	--	*-0.17*	*0.10*	*0.10*
B_2B_2	**-1.48**	**0.76**	**-1.24**	**-7.24**
	-1.75	1.26	-2.51	-6.27
	0.28	-0.32	1.17	-1.07
	--	*-0.17*	*0.10*	*0.10*

Genetic Variance Components

Single-locus model

Genotype A_2A_2 A_1A_2 A_1A_1

Genotypic value $-a$ 0 d a

T	Fre-quency Pr(T)	Genotypic value G_T	Pr(T) × G_T	$(G_T)^2$	Pr(T) × $(G_T)^2$	Breeding value $A_T=\alpha_i+\alpha_j$	Domi-nance dev. $D_T=\delta_{ij}$
A_1A_1	p^2	a	p^2a	a^2	p^2a^2	$2q\alpha$	$-2q^2d$
A_1A_2	$2pq$	d	$2pqd$	d^2	$2pqd^2$	$(q-p)\,\alpha$	$2pqd$
A_2A_2	q^2	$-a$	$-q^2a$	a^2	q^2a^2	$-2p\alpha$	$-2p^2d$

Genetic model for genotypic values: $G_T = A_T + D_T$ = Breeding value + Dominance dev.

Variance of genotypic values in a population = (Total) **Genetic variance** = V_G

$V_G = \text{var}(G_T) = p^2a^2 + 2pqd^2 + q^2a^2 - E(G_T)^2 = 2pq[a + (q-p)d]^2 + (2pqd)^2$

Using $\alpha = a + (q-p)d$ = allele substitution effect: $V_G = 2pq\ \alpha^2 + (2pqd)^2$

Additive genetic variance = **variance of breeding values** in a population = V_A

$V_A = \text{var}(A_T) = p^2(2q\alpha)^2 + 2pq[2(q-p)\alpha]^2 + q^2(-2p\alpha)^2 - 0^2 = 2pq\alpha^2$ (Note that $E(A_T)=0$)

Dominance variance = **variance of Dominance deviations** in a population = V_D

Using the table on p1, the variance of dominance deviations in the population is :

$V_D = \text{var}(D_T) = \text{var}(\delta_{ij}) = p^2(-2q^2d)^2 + 2pq(2pqd)^2 + q^2(-2p^2d)^2 - 0^2 = (2pqd)^2$ ($E(D_T)=0$)

➔ Genotypic variance = $V_G = 2pq\alpha^2 + (2pqd)^2 = V_A + V_D$
= Additive Variance + Dominance Variance

Note : $\text{cov}(A_T, D_T) = 0$; i.e. breeding values and dominance deviations are independent

Extension to two loci – first without epistasis:

Genotypic value = $G_T = G_A + G_B$ G_i = genotypic value locus i

$V_G = \text{var}(G_T) = \text{var}(G_A + G_B) = \text{var}(G_A) + \text{var}(G_B) + 2\text{cov}(G_A, G_B)$

$= \text{var}(G_A) + \text{var}(G_B) + 0$ cov=0 if loci are in LE

$= 2p_Aq_A\alpha_A^2 + (2p_Aq_Ad_A)^2 + 2p_Bq_B\alpha_B^2 + (2p_Bq_Bd_B)^2$

$= \{2p_Aq_A\alpha_A^2 + 2p_Bq_B\alpha_B^2\} + \{(2p_Aq_Ad_A)^2 + (2p_Bq_Bd_B)^2\}$

$= \{ V_{A_A} + V_{A_B} \} + \{ V_{D_A} + V_{D_B} \}$

$= V_A + V_D$

Homozygote midpoint =	10
a_A = 4	a_B = 3
d_A = 2	d_B = -1
p_A = 0.6	p_B = 0.3
q_A = 0.4	q_B = 0.7

Linkage Disequilibrium D =	0
Recomb. Rate =	0.2

Input matrix for epistatic effects

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	0	0	0
B_1B_2	0	0	0
B_2B_2	0	0	0

ALLELE-BASED MODEL FOR GENOTYPIC VALUES

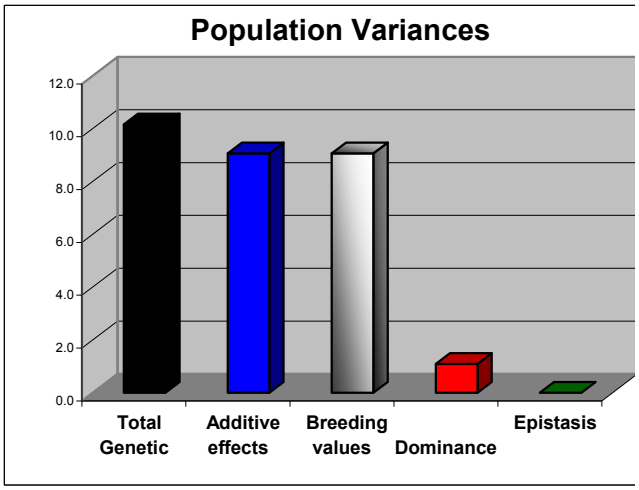
	Average allele effects		Substitution effect
Locus A	α_{A1} = 1.44	α_{A2} = -2.16	α_A = 3.6
Locus B	α_{B1} = 1.82	α_{B2} = -0.78	α_B = 2.6

All values are now deviated from the population mean, M.

		A locus genotype		
		A_1A_1	A_1A_2	A_2A_2
B locus Genotype	($+G_T$)-*M*	**2.24**	**0.24**	**-5.76**
	Additive	2.88	-0.72	-4.32
	Dom.dev.	*-0.64*	*0.96*	*-1.44*
	Epistasis	--	--	--
B_1B_1	**4.62**	**6.86**	**4.86**	**-1.14**
	3.64	6.52	2.92	-0.68
	0.98	0.34	1.94	-0.46
	–	*0.00*	*0.00*	*0.00*
B_1B_2	**0.62**	**2.86**	**0.86**	**-5.14**
	1.04	3.92	0.32	-3.28
	-0.42	-1.06	0.54	-1.86
	–	*0.00*	*0.00*	*0.00*
B_2B_2	**-1.38**	**0.86**	**-1.14**	**-7.14**
	-1.56	1.32	-2.28	-5.88
	0.18	-0.46	1.14	-1.26
	–	*0.00*	*0.00*	*0.00*

Population variances	A locus	B locus	Population	Percent
Total Genetic	**7.142**	**3.016**	**10.158**	100.0%
Additive effects	6.221	2.839	9.060	89.2%
Breeding values	6.221	2.839	9.06	
Dominance	*0.922*	*0.176*	*1.098*	*10.8%*
Epistasis	--	--	*0.000*	*0.0%*

"Additive effects" refer to breeding values computed as the sum of average allele effects.
"Breeding values" are computed based on the expected progeny means



Extended to **>2 loci**, this gives:

$$V_G = \Sigma V_{Gi} = \Sigma\{2p_iq_i\alpha_i^2 + (2p_iq_id_i)^2\} = \Sigma V_{Ai} + \Sigma V_{Di} \quad = V_A + V_D$$

with: $V_A = \Sigma V_{Ai} = \Sigma 2p_iq_i\alpha_i^2$ and $V_D = \Sigma V_{Di} = \Sigma(2p_iq_id_i)^2$

➔ the genetic, additive, and dominance variances for a quantitative trait are the simple sum of the genetic, additive, and dominance variances at each locus that affect the trait.

With Epistatic effects = Interactions between the effects that loci have on phenotype

Two locus example: $G_T = G_A + G_B + G_{AxB}$ Genotype-based model

$G_T = \alpha_{Ai} + \alpha_{Aj} + \alpha_{Bi} + \alpha_{Bj} + \delta_{Aij} + \delta_{Bij} + I_{AB}$ Allele-based model

$= A + D + I$

Epistatic variance = variance of epistatic deviations in a population = $V_I = \text{var}(I_{AB})$

➔ Complete partitioning of genetic variance: $V_G = V_A + V_D + V_I$ Note: all cov's = 0

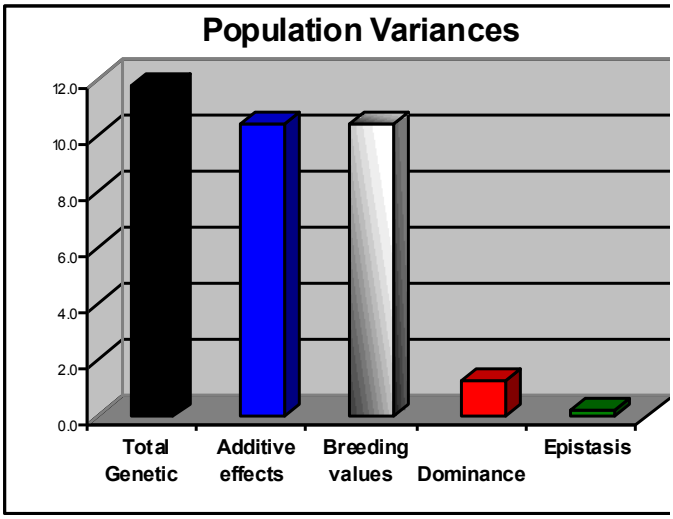
➔ Epistatic variance can be obtained by difference: $V_I = V_G - V_A - V_D$ see spreadsheet for ex.

Input matrix for epistatic effects			
	A_1A_1	A_1A_2	A_2A_2
B_1B_1	3	0	0
B_1B_2	0	0	0
B_2B_2	0	0	0

Population variance	A locus	B locus	Population	Percent
Total Genetic	7.595	4.009	11.774	100.0%
Additive effects	6.793	3.591	10.384	88.2%
Breeding values	6.793	3.591	10.41357	
Dominance	*0.801*	*0.418*	*1.220*	*10.4%*
Epistasis	--	--	*0.170*	*1.4%*

		A locus genotype		
		A_1A_1	A_1A_2	A_2A_2
	$(\mu+G_T)$-*M*	**2.41**	**0.14**	**-5.86**
	Additive	3.01	-0.75	-4.51
B locus	*Dom.dev.*	*-0.60*	*0.90*	*-1.34*
Genotype	*Epistasis*	--	--	--
B_1B_1	**5.60**	**9.76**	**4.76**	**-1.24**
	4.09	7.10	3.34	-0.42
	1.51	*0.91*	*2.40*	*0.17*
	--	*1.75*	*-0.98*	*-0.98*
B_1B_2	**0.52**	**2.76**	**0.76**	**-5.24**
	1.17	4.18	0.42	-3.34
	-0.65	*-1.24*	*0.25*	*-1.99*
	--	*-0.17*	*0.10*	*0.10*
B_2B_2	**-1.48**	**0.76**	**-1.24**	**-7.24**
	-1.75	1.26	-2.51	-6.27
	0.28	*-0.32*	*1.17*	*-1.07*
	--	*-0.17*	*0.10*	*0.10*

Population means	A locus	B locus	Population	Population means are always equal to zero
Genotypic values	**0.00**	**0.00**	**0.00**	
Additive	0.00	0.00	0.00	
Dominance deviations	*0.00*	*0.00*	*0.00*	
Epistasic deviations	--	--	*0.00*	



Covariances	Addit.	Dom.dev	Covariances are always zero
Dom.dev	0.00		
Epist.dev	0.00	0.00	

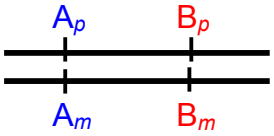
Additive, dominance, and epistatic effects are independent (no covariances)

In a typical population, most genetic variance is additive – see also Hill et al. PLOS Genetics (2008)

Impact of Linkage disequilibrium on genetic variances

Consider 2 linked loci, A and B:
Each individual has a paternal and a maternal gamete

Paternal gamete *p*
Maternal gamete *m*



The breeding value is the sum of the average effects of the paternal and maternal alleles:

$$A = \alpha_{Ap} + \alpha_{Bp} + \alpha_{Am} + \alpha_{Bm}$$

If the loci are in LD ➔ allele states (0/1) at two loci on the same gamete are not independent
➔ they have a non-zero covariance (r^2 = squared correl. of allele states (0/1) is > 0)

Then, the variance caused by the **additive effect** of *the paternal* (or *maternal*) *gamete* is

$$\text{var}(\alpha_{Ap} + \alpha_{Bp}) = \text{var}(\alpha_{Ap}) + \text{var}(\alpha_{Bp}) + 2\text{cov}(\alpha_{Ap}, \alpha_{Bp})$$

From bottom p.1: $\text{var}(\alpha_i) = pq\alpha^2 = \frac{1}{2}V_A$

$$= \tfrac{1}{2}V_{A_A} + \tfrac{1}{2}V_{A_B} + 2D_{AB}\alpha_{Ap}\alpha_{Bp}$$

where D_{AB} = LD A,B

➔ Additive genetic variance = $V_A = V_{A_A} + V_{A_B} + 4D_{AB}\alpha_{Ap}\alpha_{Bp}$ see spreadsheet for example

➔ Dominance genetic variance =

$$V_D = \text{var}(\delta_{Apm} + \delta_{Bpm}) = \text{var}(\delta_{Apm}) + \text{var}(\delta_{Bpm}) + 2\text{cov}(\delta_{Apm}, \delta_{Bpm})$$

$$= V_{D_A} + V_{D_B} + 8D_{AB}^2 d_A d_B$$

(D^2 because dominance is based on combinations of paternal and maternal alleles)

Homozygote midpoint =	10
a_A = 4	a_B = 3
d_A = 2	d_B = -1
p_A = 0.6	p_B = 0.3
q_A = 0.4	q_B = 0.7
Linkage Disequilibrium D =	0
Recomb. Rate =	0.2

D = +0.1

Population variances	A locus	B locus	Population	Percent
Total Genetic	7.142	3.016	13.742	100.0%
Additive effects	6.221	2.839	12.804	93.2%
Breeding values	6.221	2.839	12.804	
Dominance	*0.922*	*0.176*	*0.938*	*6.8%*
Epistasis	--	--	*0.000*	*0.0%*

D = 0

Population variances	A locus	B locus	Population	Percent
Total Genetic	7.142	3.016	10.158	100.0%
Additive effects	6.221	2.839	9.060	89.2%
Breeding values	6.221	2.839	9.06	
Dominance	*0.922*	*0.176*	*1.098*	*10.8%*
Epistasis	--	--	*0.000*	*0.0%*

D = - 0.1

Population variances	A locus	B locus	Population	Percent
Total Genetic	7.142	3.016	6.254	100.0%
Additive effects	6.221	2.839	5.316	85.0%
Breeding values	6.221	2.839	5.316	
Dominance	*0.922*	*0.176*	*0.938*	*15.0%*
Epistasis	--	--	*0.000*	*0.0%*

Note that individual locus variances are not affected by LD but across locus variances are (because of non-zero covariances).

Whether V_A increases or decreases depends on whether the favorable alleles are in repulsion or coupling phase

Whether V_D increases or decreases depends on whether *d* has the same sign for both loci.

Phenotypic and Environmental Effects and Variances

Models for Environmental Effects

Phenotype for a quantitative trait is determined by genetic and environmental factors:

$$P = \mu + G + E$$

μ includes the mean and ***systematic*** (*environmental*) ***effects***
G = genotypic value
E = ***Random environmental effects***

Partitioning of phenotypic variance

Phenotypic variance = var. of phenotypes in a pop. after removal/adjustment for syst. effects

$$= V_P = \text{var}(P-\mu) = \text{var}(G+E) = \text{var}(G) + \text{var}(E) + 2\text{cov}(G,E)$$

If genotypes are distributed at random relative to random environmental effects ➔ $\text{cov}(G,E)=0$

$$\rightarrow \quad V_P = V_G + V_E = V_A + V_D + V_I + V_E$$

Relative importance of the genetic component

Genetic variance as a fraction of the phenotypic variance:

Broad sense heritability $H^2 = \frac{V_G}{V_P}$ = proportion of phen. var. in a pop. that is genetic

Narrow sense heritability $h^2 = \frac{V_A}{V_P}$ = proportion of phen. var. that is additive genetic

MODELS FOR TRAITS WITH REPEATED MEASURES

General environmental effects = Eg =Effects that are common to each measurement

Special environmental effects = Es = Effects that are specific to a given measurement

$P_{ij} = G_i + Eg_i + Es_{ij}$ $P_{ij} = j^{th}$ measurement of phenotype on i^{th} individual
G_i and Eg_i are common to all measurements on individual i
Es_{ij} = special envir. effect for j^{th} measurement on i^{th} individual

This also allows random environmental variance to be separated into variances due to General versus Special environmental effects:

$V_P = V_G + V_{Eg} + V_{Es}$ Note: cov(Eg,Es)=0

Repeatability = r = correlation between repeated measures on the same individual
(Assume that V_{Es} and therefore V_P is the same for each measurement)

$$r = \text{corr}(P_{ij}, P_{ik}) = \frac{\text{cov}(P_{ij}, P_{ik})}{\sqrt{\text{var}(P_{ij})\text{var}(P_{ik})}} = \frac{\text{cov}(P_{ij}, P_{ik})}{\sqrt{V_P V_P}} = \frac{\text{cov}(P_{ij}, P_{ik})}{V_P}$$

$$\text{cov}(P_{ij}, P_{ik}) = \text{cov}(G_i + Eg_i + Es_{ij},\ G_i + Eg_i + Es_{ik})$$

$$= \text{cov}(G_i, G_i) + \text{cov}(Eg_i, Eg_i) + \text{Cov}(Es_{ij}, Es_{ik})$$

$$= \quad V_G \quad + \quad V_{Eg} \quad + \quad 0$$ special env. effects are independent

➔ repeatability = $r = (V_G + V_{Eg}) / V_P$ = prop. of V_P that is due to effects that are consistent across measurements ($G + Eg$)

➔ $1 - r = V_{Es} / V_P$ = prop. of V_P that is due to effects that differ between measurements (Es)

CORRELATED TRAITS

Phenotypic correlation = correlation between phenotypes on traits 1 and 2 on same individual
= caused by genetics and environment

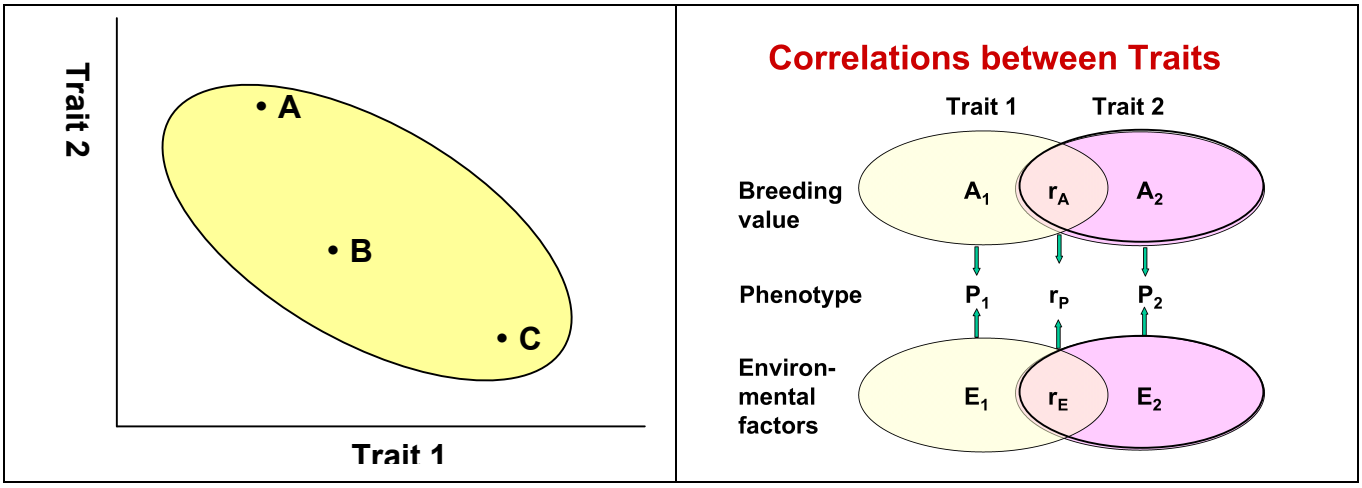
Phenotype trait 1 $P_1 = A_1 + D_1 + I_1 + E_1$

(Additive) genetic correlation = $\mathbf{r_A}$
= Corr(A_1,A_2)

Environmental correlation
= $\mathbf{r_E}$

Phenotype trait 2 $P_2 = A_2 + D_2 + I_2 + E_2$

$$r_P = \frac{Cov(P_1,P_2)}{\sigma_{P_1}\sigma_{P_2}} \qquad r_A = \frac{Cov(A_1,A_2)}{\sigma_{A_1}\sigma_{A_2}} \qquad r_E = \frac{Cov(E_1,E_2)}{\sigma_{E_1}\sigma_{E_2}}$$



Genetic correlation – caused by - **pleiotropic genes** = genes with effect on both traits
- linkage – a gene that affects trait 1 is in LD with a gene that affects trait 2

→ transient correlation – disappears with loss of LD

- quantifies the overall effect on both traits, across <u>all</u> loci

➔ r_A = 0 does <u>not</u> imply that there are <u>no</u> pleiotropic genes

Environmental correlation – caused by random environmental factors that affect both traits
– measures the overall effect of <u>all</u> environmental factors

Some quantitative genetic math to show relationships among correlations:

$$Cov(P_1,P_2) = Cov(A_1+E_1, A_2+E_2) = Cov(A_1,A_2) + Cov(E_1,E_2)$$

➔ $r_P\sigma_{P_1}\sigma_{P_2} = r_A\sigma_{A_1}\sigma_{A_2} + r_E\sigma_{E_1}\sigma_{E_2}$

➔ $r_P\sigma_{P_1}\sigma_{P_2} = r_A h_1\sigma_{P_1} h_2\sigma_{P_2} + r_E e_1\sigma_{P_1} e_2\sigma_{P_2}$ $\quad e^2 = 1\text{-}h^2$ = prop.of phen.var. that is <u>not</u> add.genetic

➔ $r_P = r_A h_1 h_2 + r_E e_1 e_2$

GENETIC RELATIONSHIPS AND INBREEDING

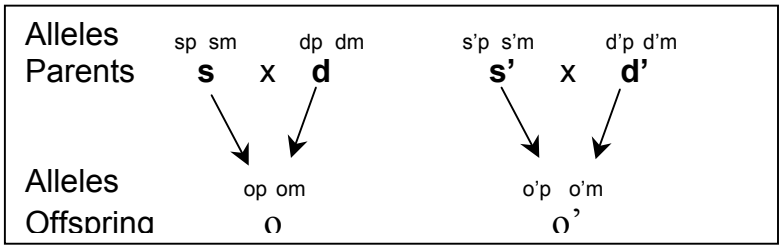
Are two alleles the same? Identity By State (IBS) *versus* Identity By Descent (IBD)

- ***IBS***: if we can genotype individuals o and o' for this locus (QTL), then we can directly determine whether the alleles the two individuals carry are indeed the same – if they are the same, this is referred to as the alleles being ***IBS***.
- ***IBD***: if we cannot genotype the locus (ie. the usual case), then we cannot determine IBS directly but, if **o** and **o'** have a ***common ancestor***, than we can determine the ***probability*** that the two alleles are identical because they may have originated from a common ancestor

IBD probabilities from pedigree:

Prob(op is IBD to o'p) = P(op $\stackrel{=}{}$ o'p)

= ***probability that alleles*** op **and** o'p ***originated from the same allele of the common ancestor***



Example IBD probabilities, coefficients of coancestry and additive and dominance coefficients

	IBD probabilities for pairs of alleles				Coancestry coefficient	Additive relationship coefficient	Dominance relationship coefficient
Individual o – o'	**op–o'p**	**om–o'm**	**op–o'm**	**om–o'p**	$f_{oo'}$	$r_{oo'}$	$u_{oo'}$
Sire(o) – Offspring(o')	½	0	0	½	¼	½	0
Dam – Offspring	0	½	½	0	¼	½	0
Paternal half-sibs	½	0	0	0	$^1/_8$	¼	0
Full sibs	½	½	0	0	¼	½	¼
Identical twins	1	1	0	0	½	1	1

Coefficient of coancestry (also coeff. of kinship or consanguinity) between **o** and **o'** (See also Ch 5 p85)

= $f_{oo'}$ = probability that an allele drawn at random from **o** is IBD to an allele drawn random from **o'**
= average of the 4 possible IBD probabilities between alleles at o and o'

$r_{oo'}$ = $2f_{oo'}$ = **coefficient of relationship = additive genetic relationship coefficient**

NOTE: $f_{oo'}$ is also equal to the **coefficient of inbreeding** of a progeny produced by **o** and **o'**
= probability that an individual's alleles are IBD

From IBD probabilities to covariances of between relatives:

$$Cov(G_o,G_{o'}) = r_{oo'}V_A + u_{oo'}V_D$$

This equation applies to each locus that affects the trait but also to total genetic value; summing variances over loci ($\mathbf{V_A} = \Sigma\mathbf{V_{Ai}}$), this equation also applies to multiple loci

Thus, the genetic covariance (resemblance) between relatives is a function of their genetic relationships and genetic variance components.

- the additive genetic cov. between relatives = genetic relationship x add. genetic var.= $r_{oo'}V_A$
  - relatives 'share' a portion $r_{oo'}$ of their additive genetic variance because they share a portion $r_{oo'}$ of their alleles
- the dominance genetic cov. betw. relatives = dom. relationship x dom. genetic var. = $u_{oo'}V_D$

ALTERNATIVE DERIVATION OF ADDITIVE COVARIANCES based on *quantitative genetics algebra*

Model of phenotype: $P = A + E$ E includes D, I, environment

Offspring phenotype: $P_o = A_o + E_o = \frac{1}{2}A_s + \frac{1}{2}A_d + RA_s + RA_d + E_o$

$\frac{1}{2}$ * breeding value of parents

- Breeding value = $2*E(P_O-M)$ (by definition)
- Includes some dominance and epistatic effects

RA_s , RA_d = random assortment / Mendelian sampling terms

- sampling of 1 of 2 parent alleles at each locus during meiosis
- by definition independent from other terms: $Cov(A_s,RA_s) = 0$

Without inbreeding: $Var(RA_s) = \frac{1}{4}V_A$ $Var(RA_d) = \frac{1}{4}V_A$ (see derivation below)

With inbreeding: $Var(RA_s) = \frac{1}{4}(1-F_s)V_A$ $Var(RA_d) = \frac{1}{4}(1-F_d)V_A$

Thus: $Var(A_o) = Var(\frac{1}{2}A_s + \frac{1}{2}A_d + RA_s + RA_d) =$

$= \frac{1}{4}V_A + \frac{1}{4}V_A + \frac{1}{4}V_A + \frac{1}{4}V_A = V_A$ (no inbreeding or selection)

Single locus derivation of Var(RA)

Parent Geno-type	Fre-quen-cy	Genotypic value of parent [$\alpha=a+(q-p)d$]		Offspring mean phenotype = ½*breeding value parent	Trans-mitted allele	Fre-quen-cy	Offspring mean phenotype	Mendelian sampling term (RA)
A_1A_1	p^2	a	$2q(\alpha-qd)$	$q\alpha$	A_1	1	$q\alpha$	0
A_1A_2	$2pq$	d	$(q-p)\alpha+2qd$	$\frac{1}{2}(q-p)\alpha$	A_1	½	$q\alpha$	$\frac{1}{2}\alpha$
					A_2	½	$-p\alpha$	$-\frac{1}{2}\alpha$
A_2A_2	q^2	$-a$	$-2p(\alpha+pd)$	$-p\alpha$	A_2	1	$-p\alpha$	0

$E(RA_s) = p^2(0) + 2pq\frac{1}{2}(\frac{1}{2}\alpha) + 2pq\frac{1}{2}(-\frac{1}{2}\alpha) + q^2(0) = 0$

Without inbreeding: $Var(RA_s) = p^2(0)^2 + 2pq\frac{1}{2}(\frac{1}{2}\alpha)^2 + 2pq\frac{1}{2}(-\frac{1}{2}\alpha)^2 + q^2(0)^2$

$= \frac{1}{2}pq\alpha^2 = \frac{1}{4}V_A$ ($V_A = 2pq\alpha^2$)

With inbreeding: F_s = Pr(two alleles in s are ibd) ➔ $RA_s = 0$

$1-F_s$ = Pr(two alleles in s not ibd) ➔ RA_s = as in Table above

$Var(RA_s) = F_s(0) + (1-F_s)\frac{1}{4}V_A = \frac{1}{4}(1-F_s)V_A$