Biometrical genetics

Manuel Ferreira
Shaun Purcell
Pak Sham

Boulder Introductory Course 2006

Outline

1. Aim of this talk
2. Genetic concepts
3. Very basic statistical concepts
4. Biometrical model

1. Aim of this talk

- Revisit common genetic parameters - such as allele frequencies, genetic effects, dominance, variance components, etc
- Use these parameters to construct a biometrical genetic model

↓

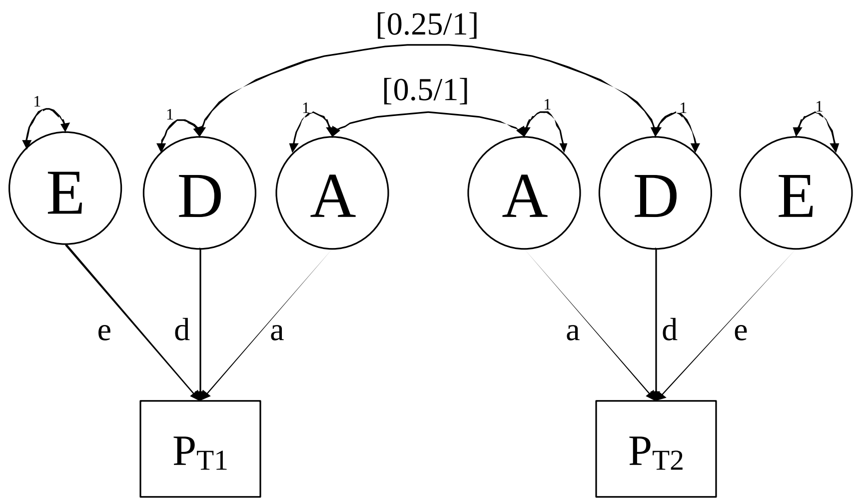
Model that expresses the:

(1) <u>Mean</u>

(2) <u>Variance</u>

(3) <u>Covariance between individuals</u>

for a quantitative phenotype as a function of genetic parameters.



2. Genetic concepts

▷ **Population level**

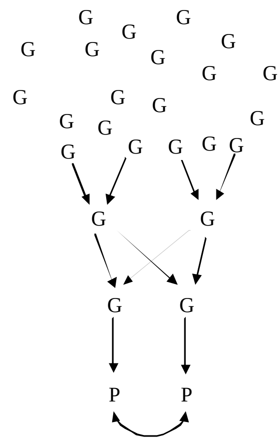
Allele and genotype frequencies

▷ **Transmission level**

Mendelian segregation
Genetic relatedness

▷ **Phenotype level**

Biometrical model
Additive and dominance components

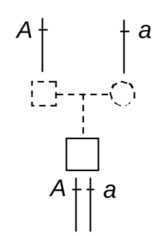


Population level

1. Allele frequencies

▷ A single <u>locus</u>, with two <u>alleles</u>
- Biallelic / diallelic
- Single nucleotide polymorphism, SNP

▷ Alleles ***A*** and ***a***
- Frequency of ***A*** is ***p***
- Frequency of ***a*** is ***q*** = 1 – ***p***



▷ Every individual inherits two alleles
- A genotype is the combination of the two alleles
- e.g. ***AA***, ***aa*** (the homozygotes) or ***Aa*** (the heterozygote)

Population level

2. Genotype frequencies (Random mating)

Allele 2 \ Allele 1	***A*** (***p***)	***a*** (***q***)
A (***p***)	***AA*** (p^2)	***Aa*** (***pq***)
a (***q***)	***aA*** (***qp***)	***aa*** (q^2)

Hardy-Weinberg Equilibrium frequencies

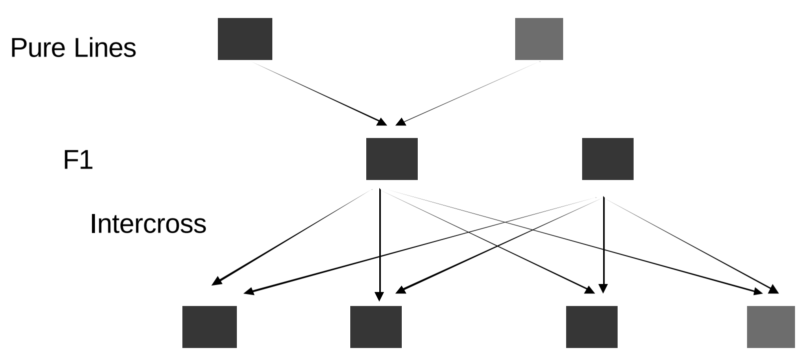
$P(AA) = p^2$

$P(Aa) = 2pq$ $\qquad p^2 + 2pq + q^2 = 1$

$P(aa) = q^2$

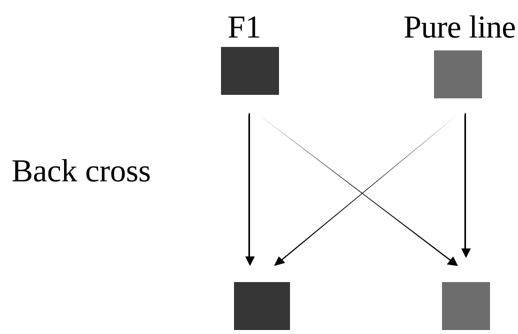
Transmission level

1. Mendel's experiments



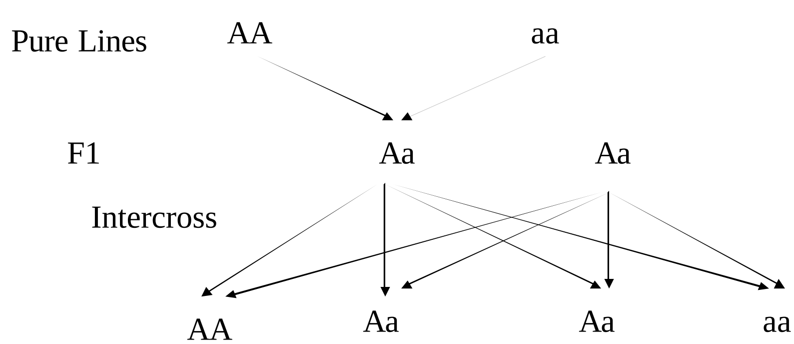
3:1 Segregation Ratio

Transmission level



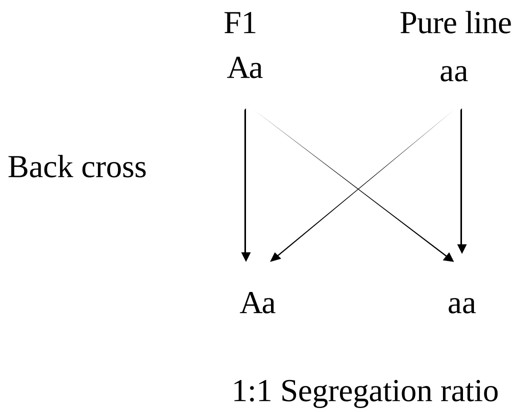
1:1 Segregation ratio

Transmission level



3:1 Segregation Ratio

Transmission level



Transmission level

1. Mendel's law of segregation

Mother (A_3A_4)

Segregation, Meiosis

Gametes

Father (A_1A_2)		A_3 (½)	A_4 (½)
	A_1 (½)	A_1A_3 (¼)	A_1A_4 (¼)
	A_2 (½)	A_2A_3 (¼)	A_2A_4 (¼)

Phenotype level

1. Classical Mendelian traits

- ▷ Dominant trait (**D** - presence, **R** - absence)
  - ***AA***, ***Aa*** **D**
  - ***aa*** **R**

- ▷ Recessive trait (**D** - absence, **R** - presence)
  - ***AA***, ***Aa*** **D**
  - ***aa*** **R**

- ▷ Codominant trait (**X, Y, Z**)
  - ***AA*** **X**
  - ***Aa*** **Y**
  - ***aa*** **Z**

Phenotype level

2. Dominant Mendelian inheritance

Mother (***Dd***)

Father (***Dd***)

	D (½)	***d*** (½)
D (½)	***DD*** (¼)	***Dd*** (¼)
d (½)	***dD*** (¼)	***dd*** (¼)

Phenotype level

3. Dominant Mendelian inheritance with incomplete penetrance and phenocopies

Mother (***Dd***)

Father (***Dd***)		***D*** (½)	***d*** (½)
	D (½)	***DD*** (¼)	***Dd*** (¼)
	d (½)	***dD*** (¼)	***dd*** (¼)

Incomplete penetrance

Phenocopies

Phenotype level

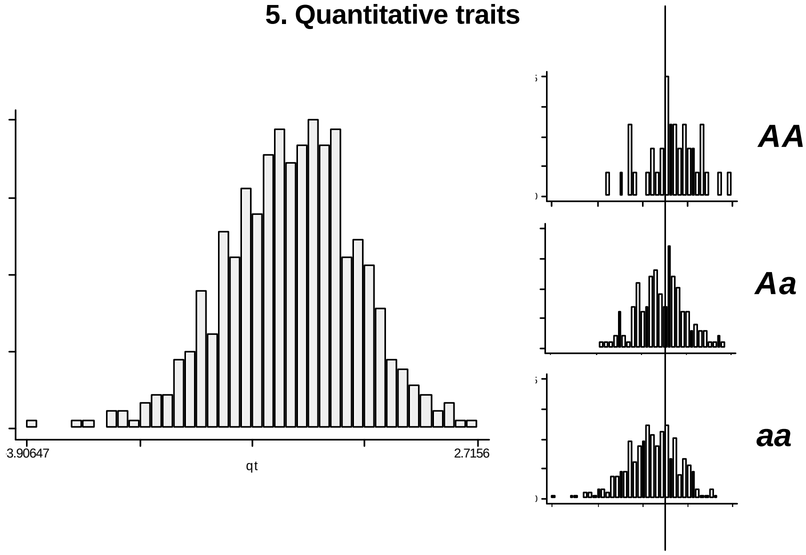
4. Recessive Mendelian inheritance

Mother (***Dd***)

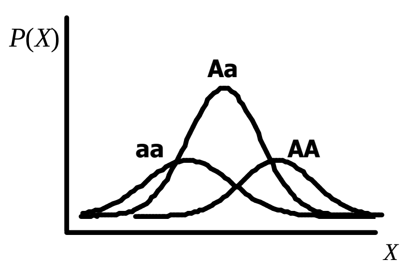
Father (***Dd***)		***D*** (½)	***d*** (½)
	D (½)	***DD*** (¼)	***Dd*** (¼)
	d (½)	***dD*** (¼)	***dd*** (¼)

Phenotype level

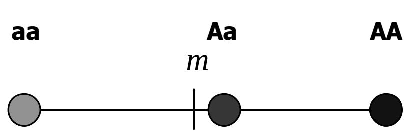
5. Quantitative traits



Phenotype level



Biometric Model



$-a$ d $+a$ ← Genotypic effect

$m-a$ $m+d$ $m+a$ ← Genotypic means

3. Very basic statistical concepts

Mean, variance, covariance

1. Mean (*X*)

$$m = E(X) = \frac{\sum_i x_i}{n} = \sum_i x_i f(x_i)$$

Mean, variance, covariance

2. Variance (*X*)

$$Var(X) = E(X-m)^2 = \frac{\sum_i (x_i - m)^2}{n-1} = \sum_i (x_i - m)^2 f(x_i)$$

Mean, variance, covariance

3. Covariance (*X,Y*)

$$Cov(X,Y) = E(X - m_X)(Y - m_Y) = \frac{\sum_i (x_i - m_X)(y_i - m_Y)}{n-1}$$

$$= \sum_i (x_i - m_X)(y_i - m_Y) f(x_i, y_i)$$

4. Biometrical model

Biometrical model for single biallelic QTL

- Biallelic locus
  - Genotypes: ***AA*, *Aa*, *aa***
  - Genotype frequencies: p^2, $2pq$, q^2
- Alleles at this locus are transmitted from P-O according to Mendel's law of segregation
- Genotypes for this locus influence the expression of a quantitative trait *X* (i.e. locus <u>*is*</u> a QTL)

↓

Biometrical genetic model that estimates the contribution of this QTL towards the **(1) Mean**, **(2) Variance** and **(3) Covariance between individuals** for this quantitative trait *X*

Biometrical model for single biallelic QTL

1. Contribution of the QTL to the Mean (X)

$$m = \sum_i x_i f(x_i)$$

Genotypes	***AA***	***Aa***	***aa***
Effect, x	***a***	***d***	***-a***
Frequencies, $f(x)$	p^2	$2pq$	q^2

$$Mean\ (X) = a(p^2) + d(2pq) - a(q^2) = a(p-q) + 2pqd$$

Biometrical model for single biallelic QTL

2. Contribution of the QTL to the Variance (X)

$$Var = \sum_i (x_i - m)^2 f(x_i)$$

Genotypes	***AA***	***Aa***	***aa***
Effect, x	***a***	***d***	***-a***
Frequencies, $f(x)$	p^2	$2pq$	q^2

$$Var\ (X) = (a-m)^2 p^2 + (d-m)^2 2pq + (-a-m)^2 q^2$$

$$= V_{QTL}$$

Broad-sense heritability of X at this locus = V_{QTL} / V_{Total}

Broad-sense total heritability of X = $\Sigma V_{QTL} / V_{Total}$

Biometrical model for single biallelic QTL

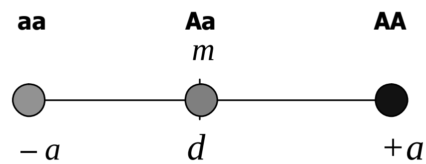
$$Var\,(X) = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$$

$$= \underline{2pq[a+(q-p)d]^2} + \underline{(2pqd)^2}$$

$$= V_{A_{QTL}} + V_{D_{QTL}}$$

Additive effects: the main effects of individual alleles

Dominance effects: represent the interaction between alleles



$d = 0$

Biometrical model for single biallelic QTL

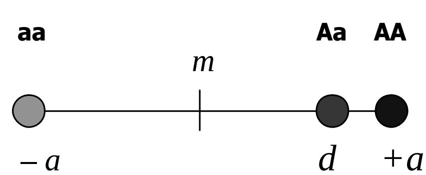
$$Var\,(X) = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$$

$$= \underline{2pq[a+(q-p)d]^2} + \underline{(2pqd)^2}$$

$$= V_{A_{QTL}} + V_{D_{QTL}}$$

Additive effects: the main effects of individual alleles

Dominance effects: represent the interaction between alleles



$d > 0$

Biometrical model for single biallelic QTL

$$Var(X) = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2q^2$$

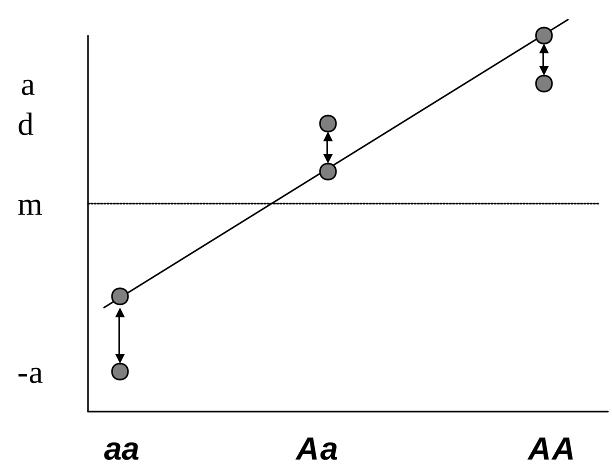
$$= \underline{2pq[a+(q-p)d]^2} + \underline{(2pqd)^2}$$

$$= V_{A_{QTL}} + V_{D_{QTL}}$$

Additive effects: the main effects of individual alleles

Dominance effects: represent the interaction between alleles

Biometrical model for single biallelic QTL

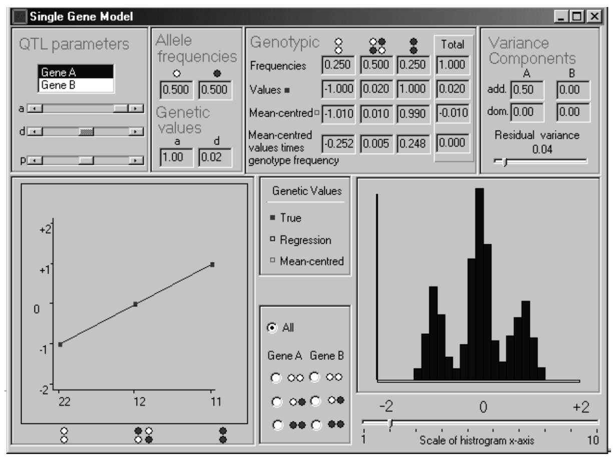


Var (X) = Regression Variance + Residual Variance

= Additive Variance + Dominance Variance

Practical

H:\manuel\Biometric\sgene.exe



Practical

▷ **Aim** Visualize graphically how allele frequencies, genetic effects, dominance, etc, influence trait mean and variance

Ex1

a=0, d=0, p=0.4, Residual Variance = 0.04, Scale = 2.
Vary a from 0 to 1.

Ex2

a=1, d=0, p=0.4, Residual Variance = 0.04, Scale = 2.
Vary d from -1 to 1.

Ex3

a=1, d=0, p=0.4, Residual Variance = 0.04, Scale = 2.
Vary p from 0 to 1.

Look at scatter-plot, histogram and variance components.

Some conclusions

1. Additive genetic variance depends on
   allele frequency p
   & additive genetic value a
   as well as
   dominance deviation d

2. Additive genetic variance typically greater than dominance variance

Biometrical model for single biallelic QTL

$$Var(X) = \underline{2pq[a+(q-p)d]^2} + \underline{(2pqd)^2}$$

↓ *Demonstrate*

$$V_{A_{QTL}} + V_{D_{QTL}}$$

2A. Average allelic effect

2B. Additive genetic variance

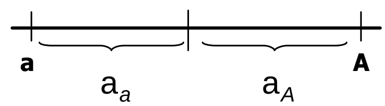
1/3

Biometrical model for single biallelic QTL

2A. Average allelic effect (a)

The deviation of the allelic mean from the population mean

	Allele ***a***	Population	Allele ***A***
Mean (X) →	?	$a(p\text{-}q) + 2pqd$	?



	AA ***a***	***Aa*** ***d***	***aa*** ***-a***	Allelic mean	Average allelic effect (a)
A	***p***	***q***		***ap+dq***	***q(a+d(q-p))***
a		***p***	***q***	***dp-aq***	***-p(a+d(q-p))***

2/3

Biometrical model for single biallelic QTL

- Denote the average allelic effects
  - $a_A = q(a+d(q-p))$
  - $a_a = -p(a+d(q-p))$
- If only two alleles exist, we can define the *average effect of allele substitution*
  - $a = a_A - a_a$
  - $a = (q-(-p))(a+d(q-p)) = (a+d(q-p))$
- Therefore:
  - $a_A = qa$
  - $a_a = -pa$

Biometrical model for single biallelic QTL

2A. Average allelic effect (a)
2B. Additive genetic variance

The variance of the average allelic effects

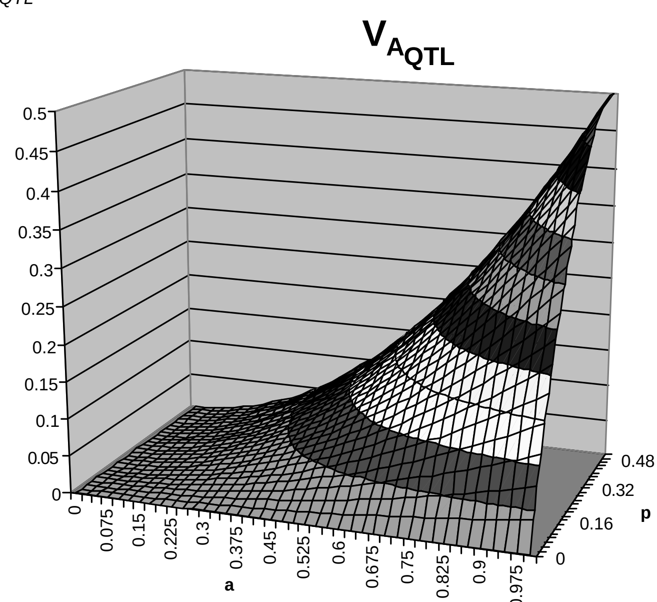
$a_A = qa$
$a_a = -pa$

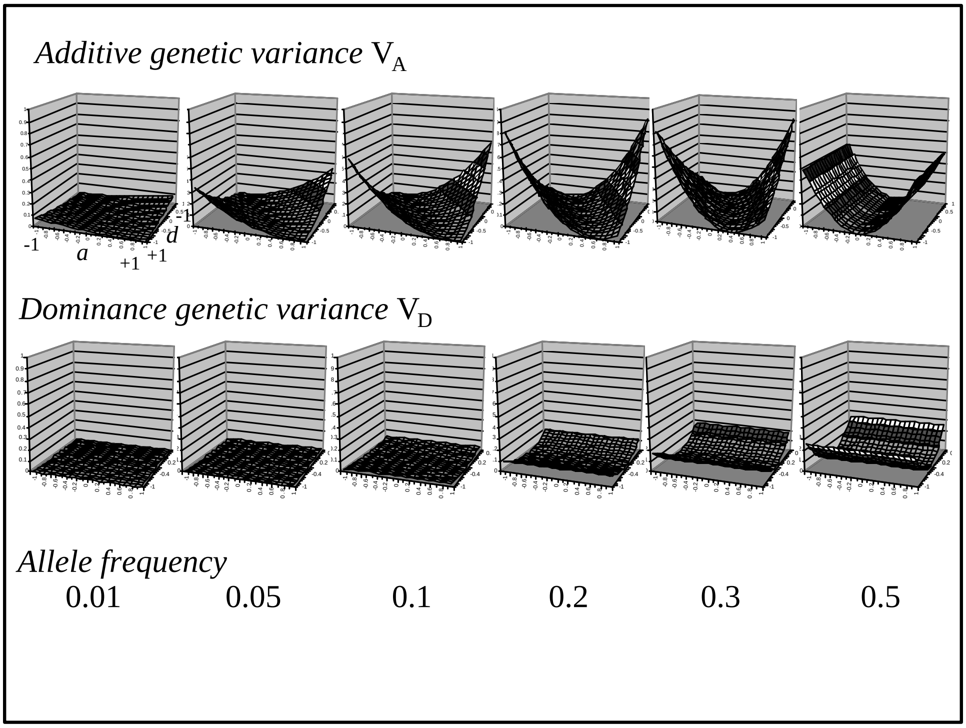
	Freq.	Additive effect	
AA	p^2	$2a_A$	$= 2qa$
Aa	$2pq$	$a_A + a_a$	$= (q-p)a$
aa	q^2	$2a_a$	$= -2pa$

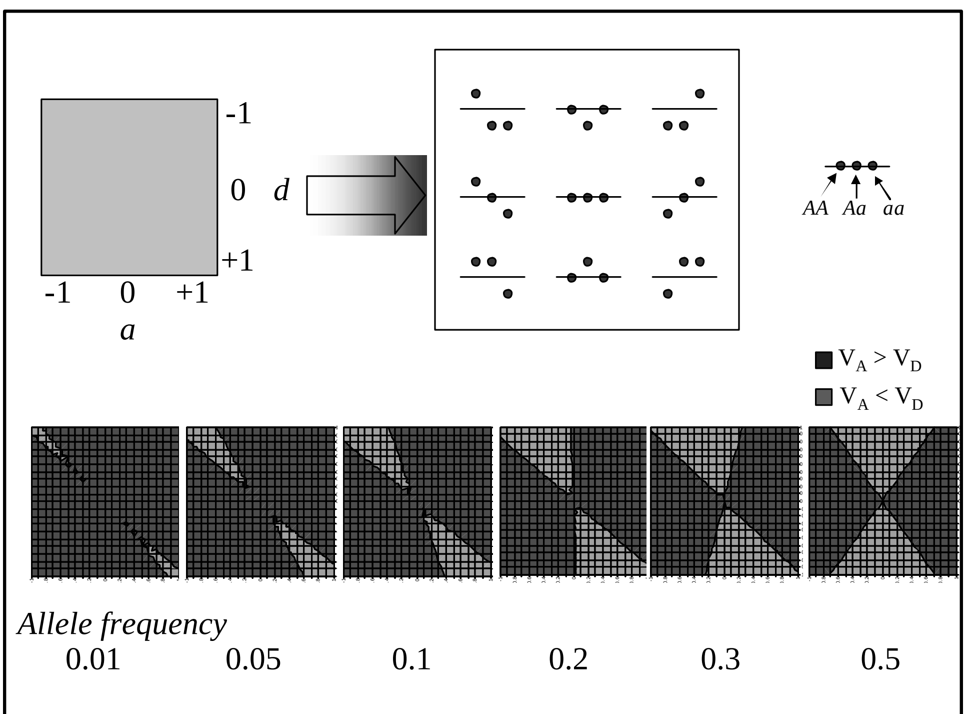
$$V_{A_{QTL}} = (2qa)^2p^2 + ((q-p)a)^2 2pq + (-2pa)^2q^2$$
$$= 2pqa^2$$
$$= 2pq[a+d(q-p)]^2$$

$d = 0, V_{A_{QTL}} = 2pqa^2$
$p = q, V_{A_{QTL}} = ½a^2$

$d = 0, V_{A_{QTL}} = 2pqa^2$




Additive genetic variance V_A
-1
a
+1
d
-1
+1
Dominance genetic variance V_D
Allele frequency
0.01
0.05
0.1
0.2
0.3
0.5


-1
0
d
+1
-1
0
+1
a
AA
Aa
aa
V_A > V_D
V_A < V_D
Allele frequency
0.01
0.05
0.1
0.2
0.3
0.5

Biometrical model for single biallelic QTL

1. Contribution of the QTL to the Mean (X)

2. Contribution of the QTL to the Variance (X)

2A. Average allelic effect (a)

2B. Additive genetic variance

3. Contribution of the QTL to the Covariance (X,Y)

Biometrical model for single biallelic QTL

3. Contribution of the QTL to the Cov (X,Y)

$$Cov(X,Y) = \sum_i (x_i - m_X)(y_i - m_Y) f(x_i, y_i)$$

	AA (*a-m*)	*Aa* (*d-m*)	*aa* (*-a-m*)
AA (*a-m*)	$(a-m)^2$		
Aa (*d-m*)	(*a-m*) (*d-m*)	$(d-m)^2$	
aa (*-a-m*)	(*a-m*) (*-a-m*)	(*d-m*) (*-a-m*)	$(-a-m)^2$

Biometrical model for single biallelic QTL

3A. Contribution of the QTL to the Cov *(X,Y)* – MZ twins

$$Cov(X,Y) = \sum_i (x_i - m_X)(y_i - m_Y) f(x_i, y_i)$$

	AA (a-m)	Aa (d-m)	aa (-a-m)
AA (a-m)	$p^2(a-m)^2$		
Aa (d-m)	0 (a-m) (d-m)	$2pq(d-m)^2$	
aa (-a-m)	0 (a-m) (-a-m)	0 (d-m) (-a-m)	$q^2(-a-m)^2$

$$Covar(X_i, X_j) = (a-m)^2p^2 + (d-m)^2 2pq + (-a-m)^2 q^2$$

$$= 2pq[a+(q-p)d]^2 + (2pqd)^2 = V_{A_{QTL}} + V_{D_{QTL}}$$

Biometrical model for single biallelic QTL

3B. Contribution of the QTL to the Cov *(X,Y)* – Parent-Offspring

	AA (a-m)	Aa (d-m)	aa (-a-m)
AA (a-m)	$p^3(a-m)^2$		
Aa (d-m)	p^2q (a-m) (d-m)	$pq(d-m)^2$	
aa (-a-m)	0 (a-m) (-a-m)	pq^2 (d-m) (-a-m)	$q^3(-a-m)^2$

- *e.g.* given an *AA* father, an *AA* offspring can come from either *AA* x *AA* or *AA* x *Aa* parental mating types

AA x *AA*	will occur $p^2 \times p^2 = p^4$ and have *AA* offspring Prob()=1
AA x *Aa*	will occur $p^2 \times 2pq = 2p^3q$ and have *AA* offspring Prob()=0.5 and have *Aa* offspring Prob()=0.5

Therefore, P(*AA* father & *AA* offspring) $= p^4 + p^3q$

$= p^3(p+q)$

$= p^3$

Biometrical model for single biallelic QTL

3B. Contribution of the QTL to the Cov (*X,Y*) – Parent-Offspring

	AA (**a**-**m**)	***Aa*** (**d**-**m**)	***aa*** (-**a**-**m**)
AA (**a**-**m**)	$p^3(a-m)^2$		
Aa (**d**-**m**)	$p^2q\,(a-m)(d-m)$	$pq\,(d-m)^2$	
aa (-**a**-**m**)	$0\,(a-m)(-a-m)$	$pq^2\,(d-m)(-a-m)$	$q^3(-a-m)^2$

$$Cov\,(X_i,X_j) = (\mathbf{a}-m)^2\mathbf{p^3} + \ldots + (-\mathbf{a}-m)^2\mathbf{q}^3$$

$$= \mathbf{pq[a+(q-p)d]}^2 = \tfrac{1}{2}V_{A_{QTL}}$$

Biometrical model for single biallelic QTL

<u>3C. Contribution of the QTL to the Cov</u> <u>(*X,Y*)</u> – Unrelated individuals

	AA *(a-m)*	***Aa*** *(d-m)*	***aa*** *(-a-m)*
AA *(a-m)*	$p^4(a-m)^2$		
Aa *(d-m)*	$2p^3q\,(a-m)\,(d-m)$	$4p^2q^2\,(d-m)^2$	
aa *(-a-m)*	$p^2q^2(a-m)\,(-a-m)$	$2pq^3\,(d-m)\,(-a-m)$	$q^4\,(-a-m)^2$

$$Cov\,(X_i,X_j) = (\mathbf{a}-m)^2\mathbf{p}^4 + \ldots + (-\mathbf{a}-m)^2\mathbf{q}^4$$
$$= 0$$

Biometrical model for single biallelic QTL

<u>3D. Contribution of the QTL to the Cov</u> <u>(*X,Y*)</u> – DZ twins and full sibs

	¼ genome	¼ genome	¼ genome	¼ genome
# identical alleles inherited from parents	**2**	**1** (father)	**1** (mother)	**0**

¼ (2 alleles) + ½ (1 allele) + *¼ (0 alleles)*

MZ twins — *P-O* — *Unrelateds*

$$Cov\,(X_i,X_j) = \tfrac{1}{4}\,Cov(MZ) + \tfrac{1}{2}\,Cov(P\text{-}O) + \tfrac{1}{4}\,Cov(Unrel)$$
$$= \tfrac{1}{4}(V_{A_{QTL}}+V_{D_{QTL}}) + \tfrac{1}{2}\,(\tfrac{1}{2}\,V_{A_{QTL}}) + \tfrac{1}{4}\,(0)$$
$$= \tfrac{1}{2}\,V_{A_{QTL}} + \tfrac{1}{4}V_{D_{QTL}}$$

Summary

- Biometrical model predicts contribution of a QTL to the mean, variance and covariances of a trait

1 QTL

$$Var(X) = V_{A_{QTL}} + V_{D_{QTL}}$$

$$Cov(MZ) = V_{A_{QTL}} + V_{D_{QTL}}$$

$$Cov(DZ) = \tfrac{1}{2}V_{A_{QTL}} + \tfrac{1}{4}V_{D_{QTL}}$$

Multiple QTL

$$Var(X) = S(V_{A_{QTL}}) + S(V_{D_{QTL}}) = V_A + V_D$$

$$Cov(MZ) = S(V_{A_{QTL}}) + S(V_{D_{QTL}}) = V_A + V_D$$

$$Cov(DZ) = S(\tfrac{1}{2}V_{A_{QTL}}) + S(\tfrac{1}{4}V_{D_{QTL}}) = \tfrac{1}{2}V_A + \tfrac{1}{4}V_D$$

- Biometrical model underlies the variance components estimation performed in Mx

$$Var\ (X) = V_A + V_D + V_E$$

$$Cov\ (MZ) = V_A + V_D$$

$$Cov\ (DZ) = \tfrac{1}{2}V_A + \tfrac{1}{4}V_D$$