

Quantitative Genetics:
The “Genetic” Analysis of Phenotypic Data



Tom Van Dooren 11/2012
vandoore@biologie.ens.fr

WE SIMPLIFY

Phenotypes are *filtrations* over all past environments experienced by ancestor individuals and all heritable information these ancestors contained.

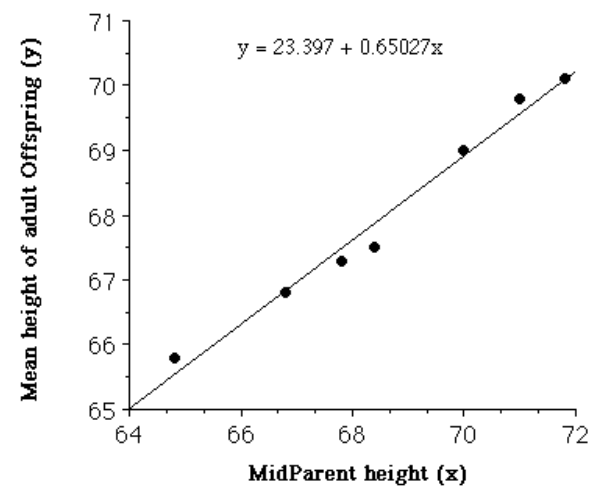
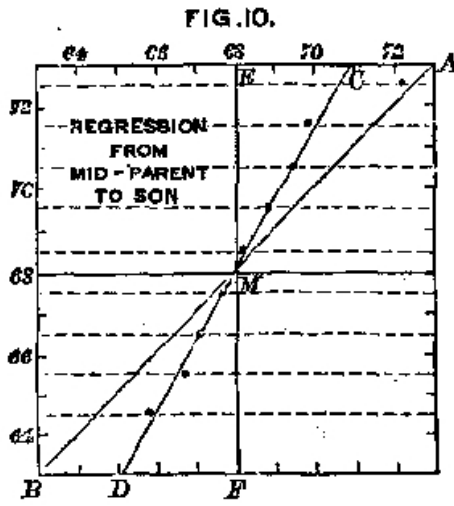
functions of history



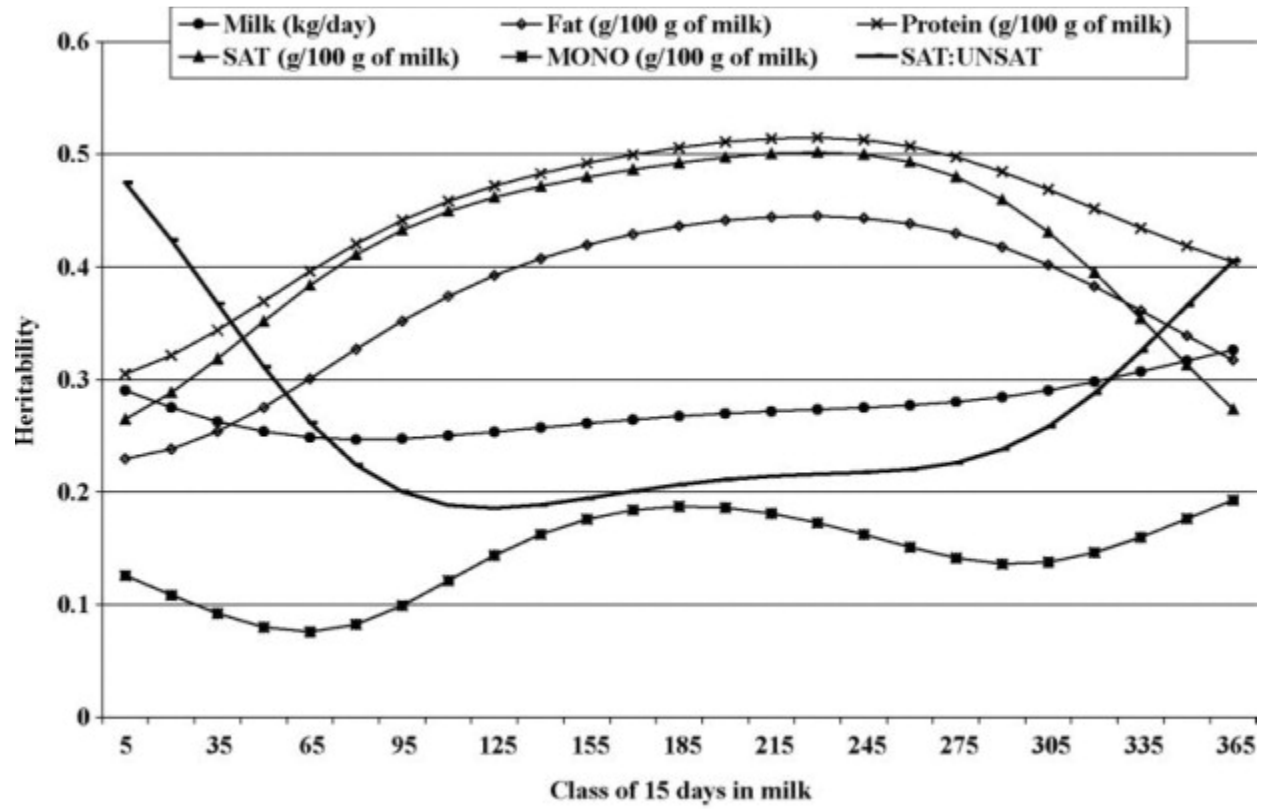
functions of current genotype, environment experienced during development

$$z = g + e$$

when allowing for pluralistic heritability: also of some other genotypes and environments of recent ancestors



Galton's parent offspring data
Galton 1894, Bruce Walsh online lecture material



Soyeurt et al. (2008)

1. What can you use quantitative genetics for?

2. Decomposition of phenotypic value

3. Relatedness

4. Decomposition of genotypic value

Covariance between genotypic values in different individuals

5. Decomposition of genotypic value, again ...

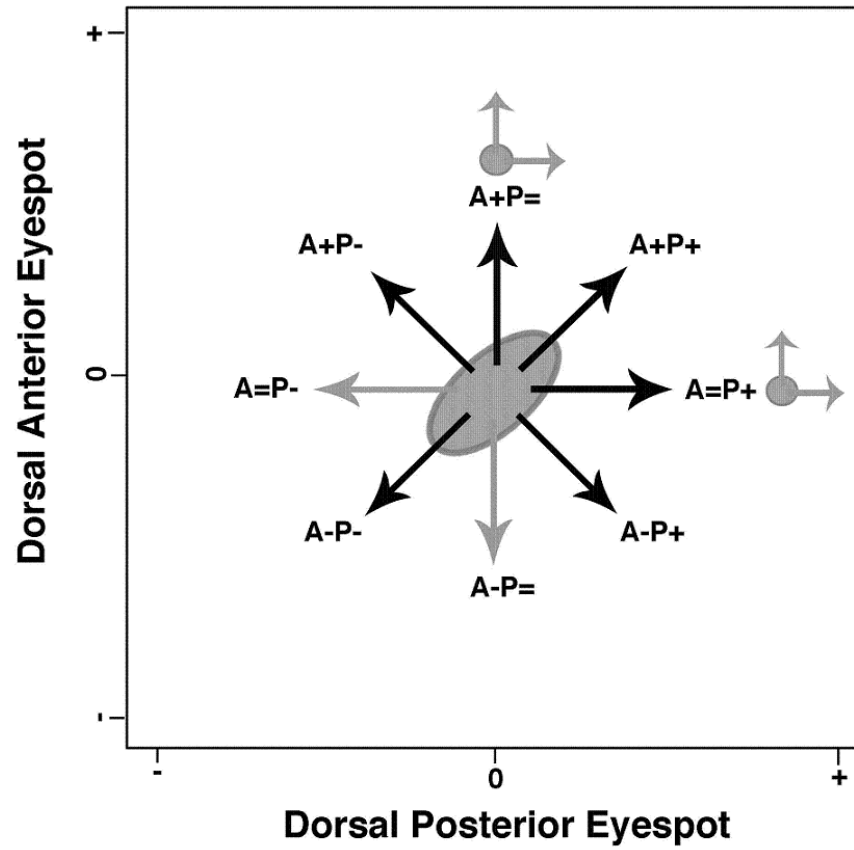
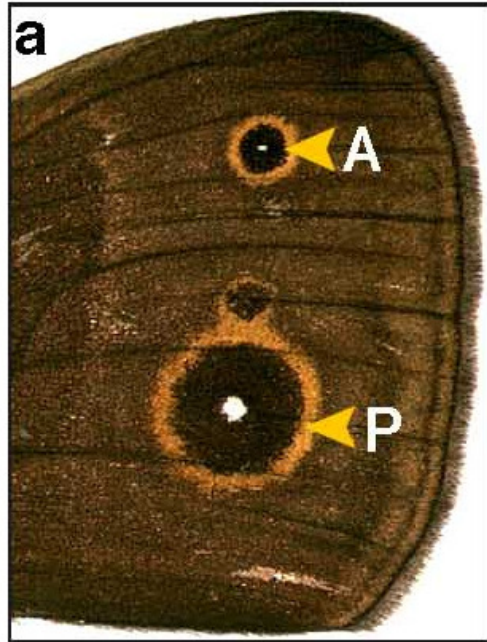
G A P

PS. Function-valued traits

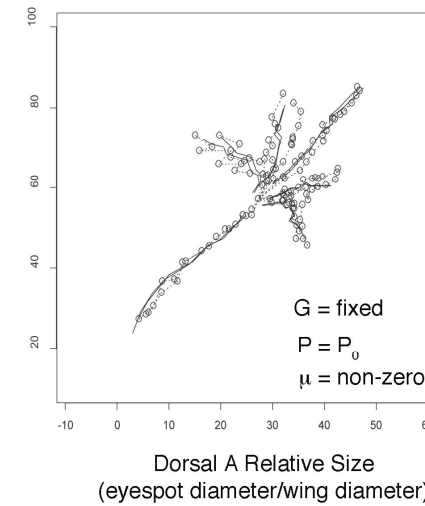
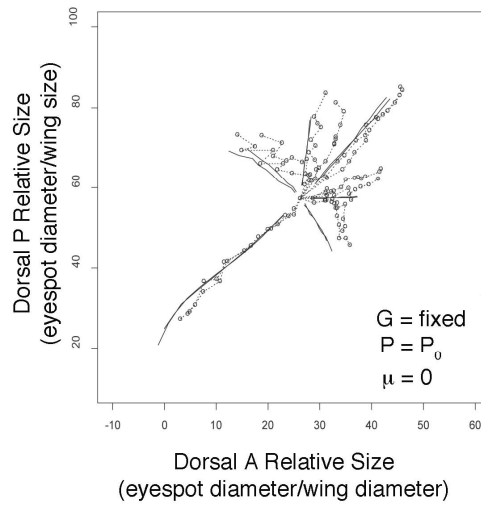
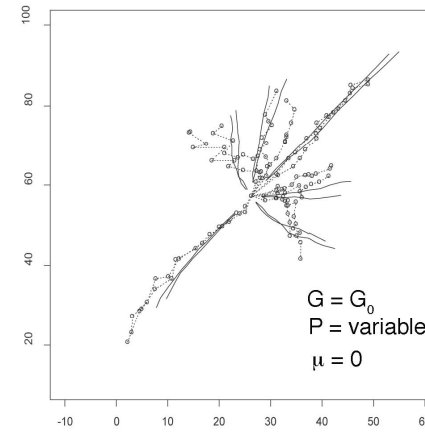
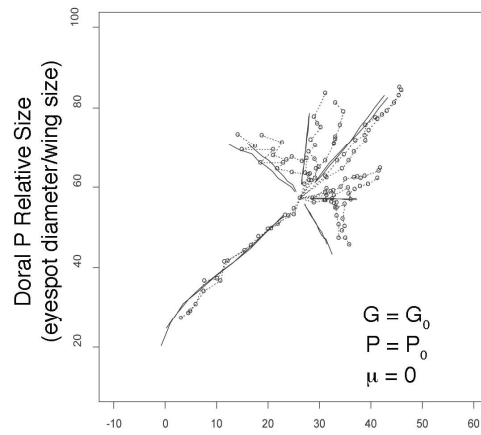
1. Without using molecular or chromosomal markers

You can

- Determine whether there is heritable variation for a phenotypic trait or probably not.
- Find out whether the amount of heritable variation is smaller or larger in different environments.
- Predict the genotypic/epigenotypic/phenotypic value of an individual compared to the population mean, using phenotypic information on relatives.
- Predict evolutionary response caused by selection



Predicted and observed responses to artificial selection on the size of dorsal forewing eyespots (Beldade et al. 2002. Allen et al. submitted).



Predicted and observed responses to artificial selection on the size of dorsal forewing eyespots (Beldade et al. 2002. Allen et al. submitted).

You can not

- Show that a trait is (partially) genetically determined. Any trait of a biological organism is genetically determined.
- Determine what the phenotypic effects are of all separate individual genes that contribute to phenotypic variation.
- Compare genetic analyses from different experiments or populations very well.

What is typical for traditional QG is

- The load of population genetic assumptions
- The stress on good experimental design
- The use of all kinds of linear models
- The neglect of the real functional dependence between genes and phenotypes
- There is no reluctance to adjust model structure depending on feasibility of parameter estimation.

Quantitative genetic data are expensive, and you really want to get something out for all the effort and money!

2. Decomposition of phenotypic value

A phenotype is composed from a genetic and an environmental contribution. These contributions are for quantitative traits usually represented by real or integer numbers.

$$\underline{z} = \underline{g} + \underline{\varepsilon}$$

random variables: phenotype \underline{z} , genotype \underline{g} and environment $\underline{\varepsilon}$

or in terms of a sample of observations, each are a realisation of these stochastic variables:

$$z_{i,j} = g_i + \varepsilon_{ij}$$

phenotype $i, j =$ genotype $i +$ environmental effect j , which is depending on genotype i

- The average environmental contribution $\underline{\varepsilon}$ that comes with a specific genotype is assumed to be zero:

$$E(\underline{\varepsilon} \mid \underline{g} = g_i) = 0 \Rightarrow E(\underline{z} \mid \underline{g} = g_i) = g_i$$

\Rightarrow *The mean phenotypic value is equal to the mean genotypic value: $E(\underline{z}) = E(\underline{g})$*

- The phenotypic variance $V(\underline{z}) = V(\underline{g}) + E_g(V_{\underline{\varepsilon}|\underline{g}}(\underline{\varepsilon}))$
- The covariance of phenotypes of two individuals \underline{z}' and \underline{z}'' becomes, with independence of $\underline{\varepsilon}'$ and $\underline{\varepsilon}''$ and independence of \underline{g}' and $\underline{\varepsilon}''$ (\underline{g}'' and $\underline{\varepsilon}'$),

$$COV(\underline{z}', \underline{z}'') = COV(\underline{g}', \underline{g}'')$$

Conclusion:

*If environments are correctly **randomized**, covariances between phenotypes depend on covariances between genotypic values only.*



3. Relatedness

Alleles are *identical by descent* (**i.b.d.**) \equiv The alleles are direct descendants of the same gene carried by a common ancestral individual.

The *inbreeding coefficient* f_a of an individual with index a is the probability that the two physically different genes at a locus in that individual are identical by descent.

An allele is drawn at random from the same locus in each of two individuals a and b :



JBS Haldane



Would you give your life to save a drowning brother?

“No, but I would to save two brothers or eight cousins. “

- Reply when asked if he would give his life to save a drowning brother, as quoted in *Mathematical Models of Social Evolution : A Guide for the Perplexed* (2007) by Richard McElreath and Robert Boyd, p. 82.

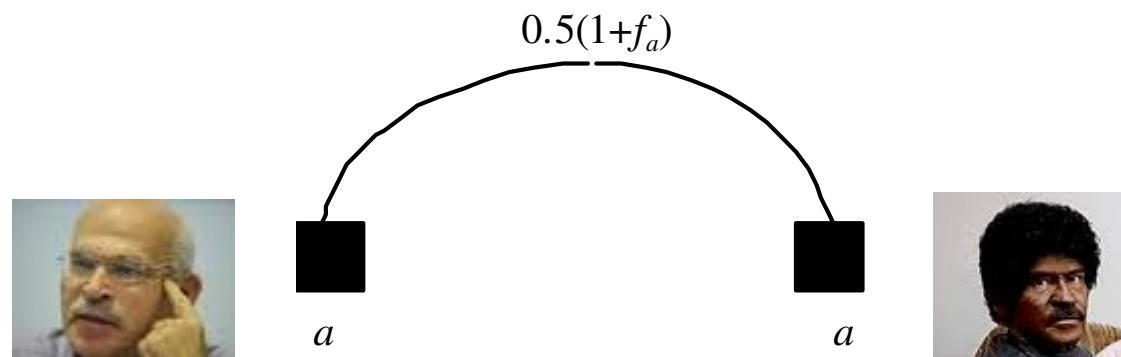
$\Theta_{a, b} \equiv \text{P}[\text{the gene from } a \text{ and the gene from } b \text{ are identical by descent}]$
= *coefficient of coancestry of } a \text{ and } b*

The inbreeding coefficient of an individual a equals the coefficient of coancestry of the parents p and m : $f_a = \Theta_{p, m}$.

$\Phi_{a, b} \equiv \text{P}[\text{the genotypes of } a \text{ and } b \text{ at the focal locus are identical by descent}]$
= *coefficient of fraternity of } a \text{ and } b*

Coancestry Diagrams

Coefficient of relatedness of an individual a with itself. Pick one allele ■ at random from a , and another one ■ from a .

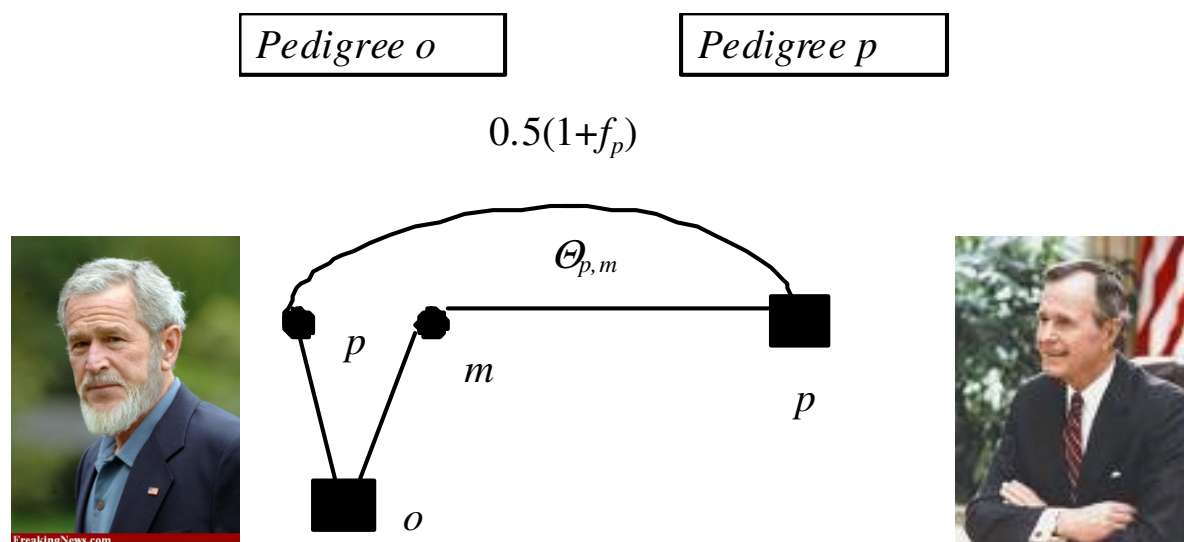


$$\theta_{a,a} = \frac{1}{4}(1 + f_a + f_a + 1)$$

From individual a with alleles $A_p A_m$, you either draw the same allele twice, or the alleles A_p and A_m from paternal and maternal origin. When the parents of individual a are related, the alleles from parental and maternal origin are identical with probability f_a .

Coancestry Diagrams

Father p and offspring o (with some help from mother m). Pick one allele ■ from o and one allele ■ from p .



$$\theta_{po} = \frac{1}{4}(1 + f_p + 2f_o) \text{ (using } f_o = \Theta_{p,m} \text{)}$$

Coefficient of Fraternity

The coefficient of fraternity can be written as a function of coefficients of coancestry.

$$\Phi_{a,b} = \theta_{m_a, m_b} \theta_{p_a, p_b} + \theta_{m_a, p_b} \theta_{p_a, m_b}$$

where m_i and p_i are the mother and father of individual i ($i = a$ or b).

This equation translates into: “The coefficient of fraternity of individuals a and b is the sum of two probabilities: 1) the maternal alleles of a and b are identical by descent, and the paternal alleles also; 2) the maternal allele of a is i.b.d. to the paternal one of b and vice versa.

4. Decomposition of genotypic value (Bulmer 1980)

- individual 1 with genotypic value g' and alleles $B_p' B_m'$
- individual 2 with genotypic value g'' and alleles $B_p'' B_m''$

\underline{g} = mean + additive effects ($\underline{\alpha}_p + \underline{\alpha}_m = \underline{A}$) + dominance interaction ($\underline{\alpha}_{mp} = \underline{D}$)

$$\underline{g}' = \mu + \underline{\alpha}'_p + \underline{\alpha}'_m + \underline{\alpha}'_{mp} = \mu + \underline{A}' + \underline{D}'$$

$$\underline{g}'' = \mu + \underline{\alpha}''_p + \underline{\alpha}''_m + \underline{\alpha}''_{mp} = \mu + \underline{A}'' + \underline{D}''$$

- $E(\underline{A}) = 0$, $E(\underline{D}) = 0$, $\text{COV}(\underline{A}, \underline{D}) = 0$ are assumed.
- the distributions of allelic effects of paternal and maternal origin are the same, and thus $V(\underline{\alpha}_p) = V(\underline{\alpha}_m) = \frac{1}{2}V(\underline{A})$.

What is the **covariance of g' and g''** ?

We need this quantity to describe patterns of variation in populations where individuals can be related.

- First calculate $COV(\underline{g}', \underline{g}'')$ conditional on a particular situation = pattern of identity by descent
- Average over probabilities that these situations occur

There are three possible **patterns of allelic identities** in two individuals, each pattern occurring with respective probability p_0 , p_1 and p_2 :

0. No identical alleles:

$$\text{COV}(g', g'' | \text{no identical alleles}) = 0$$

1. One pair of identical genes. For instance, $B_p' = B_p''$ and maternal alleles independent yields $\underline{\alpha}'_p = \underline{\alpha}''_p$. Therefore

$$\text{COV}(\underline{g}', \underline{g}'' | \underline{\alpha}'_p = \underline{\alpha}''_p) = \frac{1}{2}V(\underline{A})$$

2. Two pairs of identical alleles, $B_p' = B_p''$ and $B_m' = B_m''$:

$$\text{COV}(g', g'' | B_p' = B_p'' \text{ and } B_m' = B_m'') = V(g)$$

$$\text{COV}(\underline{A}', \underline{A}'' | B_p' = B_p'' \text{ and } B_m' = B_m'') = V(\underline{A})$$

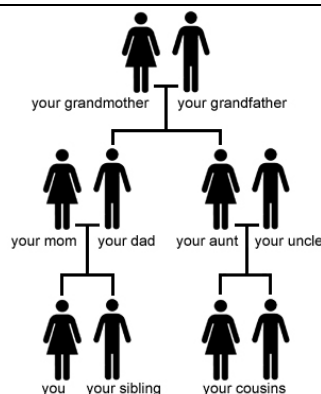
$$\text{COV}(\underline{D}', \underline{D}'' | B_p' = B_p'' \text{ and } B_m' = B_m'') = V(\underline{D})$$

Averaging over these three possibilities gives

$$COV(\underline{g}', \underline{g}'') = \left(\frac{1}{2} p_1 + p_2 \right) V(\underline{A}) + p_2 V(\underline{D})$$

- Given a set of customary assumptions, $0.5p_1 + p_2$ is twice the *coefficient of coancestry* of the two individuals

Conclusion: The covariances between individuals will depend on their relatedness.



4. Fisher's decomposition of genotypic value (after Lynch and Walsh 1998)

An individual has genotype B_1B_1 , B_1B_2 or B_2B_2 .

“gene content” N_i is the number of alleles of type B_i ($i = 1$ or 2)

$$\underline{g} = \mu + \alpha_1 \underline{N}_1 + \alpha_2 \underline{N}_2 + \underline{\delta} = \mu + \underline{A} + \underline{D}$$

- in diploids, there are two alleles at each locus: $\underline{N}_1 = 2 - \underline{N}_2$

$$\underline{g} = \mu + \alpha_1 (2 - \underline{N}_2) + \alpha_2 \underline{N}_2 + \underline{\delta}$$

$$\underline{g} = (\mu + 2\alpha_1) + (\alpha_2 - \alpha_1) \underline{N}_2 + \underline{\delta}$$

1. intercept $(\mu + 2\alpha_1)$

2. slope $\alpha = (\alpha_2 - \alpha_1) = \frac{COV(\underline{g}, \underline{N}_2)}{V(\underline{N}_2)}$

3. The genetic variance is the sum of additive genetic variance and dominance variation. $V(\underline{g}) = V(\underline{A}) + V(\underline{D})$

4. $E(\underline{g}) = \mu + \alpha_1 E(\underline{N}_1) + \alpha_2 E(\underline{N}_2)$. Usually, it is assumed that $E(\underline{g}) = \mu$.

Then the average additive genetic effect must be zero: $\alpha_1 E(\underline{N}_1) + \alpha_2 E(\underline{N}_2) = 0$.

- Assuming random mating, $E(\underline{N}_1) = 2p_1$ and $E(\underline{N}_2) = 2p_2$ and $p_1\alpha_1 + p_2\alpha_2 = 0$.

Conclusion: The usual interpretation of additive effects makes them dependent on population composition – allele frequencies.

5. Decomposition of genotypic values - multi-locus

genotype γ = paternal haplotype/maternal haplotype

$$\gamma = \underline{B}_1 \underline{B}_2 \dots \underline{B}_n / \underline{B}_{n+1} \underline{B}_{n+2} \dots \underline{B}_{2n}$$

A genotype is composed of alleles that are more or less randomly drawn from a base population of alleles.

$$\begin{aligned} \underline{g} = & \mu + \sum_{i=1 \dots n} (\underline{\alpha}_i + \underline{\alpha}_{n+i}) + \sum_{i=1 \dots n} \underline{\alpha}_{i,i+n} + \sum_{i,j=1 \dots n, i < j} (\underline{\alpha}_{i,j} + \underline{\alpha}_{i,j+n} + \underline{\alpha}_{i+n,j} + \underline{\alpha}_{i+n,j+n}) \\ & + \sum_{i,j=1 \dots n, i \neq j} (\underline{\alpha}_{i,j,j+n} + \underline{\alpha}_{i+n,j,j+n}) + \sum_{i,j=1 \dots n, i < j} \underline{\alpha}_{i,j,i+n,j+n} + \dots \end{aligned}$$

The genotypic value can be represented by a linear series of terms, as in a Taylor expansion.

- $\sum_{i=1\dots n} (\alpha_i + \alpha_{n+i})$ or A are the effects of individual genes.
- $\sum_{i=1\dots n} \alpha_{i,i+n}$ or D are the interaction effects of pairs of genes at the same locus.
- $\sum_{i,j=1\dots n, i < j} (\alpha_{i,j} + \alpha_{i,j+n} + \alpha_{i+n,j} + \alpha_{i+n,j+n})$ or AA are the effects of pairs of genes at different loci (4 terms per pair of loci).
- $\sum_{i,j=1\dots n, i \neq j} (\alpha_{i,j,j+n} + \alpha_{i+n,j,j+n})$ or AD are effects of triplets of which two alleles are at the same locus (2 terms per pair of loci)
- etc...

$$\underline{g} = \mu + \underline{A} + \underline{D} + \underline{AA} + \underline{AD} + \underline{DD} + \underline{AAA} + \dots$$

$$\underline{g} = \mu + \underline{A} + \underline{D} + \underline{AA} + \underline{AD} + \underline{DD} + \underline{AAA} + \dots$$

Independence of types of effects within individuals is assumed, that implies independence of types of effects between individuals.

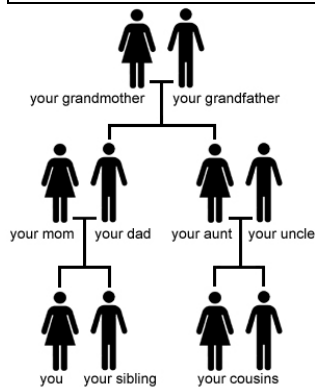
$$V(\underline{g}) = V(\underline{A}) + V(\underline{D}) + V(\underline{AA}) + V(\underline{AD}) + V(\underline{DD}) + V(\underline{AAA}) + \dots$$

$$COV(\underline{g}', \underline{g}'') = COV(\underline{A}', \underline{A}'') + COV(\underline{D}', \underline{D}'') + COV(\underline{AA}', \underline{AA}'') + COV(\underline{AD}', \underline{AD}'') + \dots$$

5. The covariance between genotypic values of relatives

Again, without linkage and so on (the effects of different alleles are independent unless they are identical by descent), this shows that

$$COV(\underline{g}', \underline{g}'') = \sum_{i,j=0\dots n} \left(2\Theta_{\underline{g}'\underline{g}''}\right)^i \left(\Phi_{\underline{g}'\underline{g}''}\right)^j V(\underline{A}^i \underline{D}^j)$$



Important for selection dynamics: in sexual populations the coefficient of fraternity is often (usually) zero in pairs of individuals of successive generations.

Mutational effects

Usually treated independently of standing variance –
Rare mutations - which are therefore only present in heterozygotes

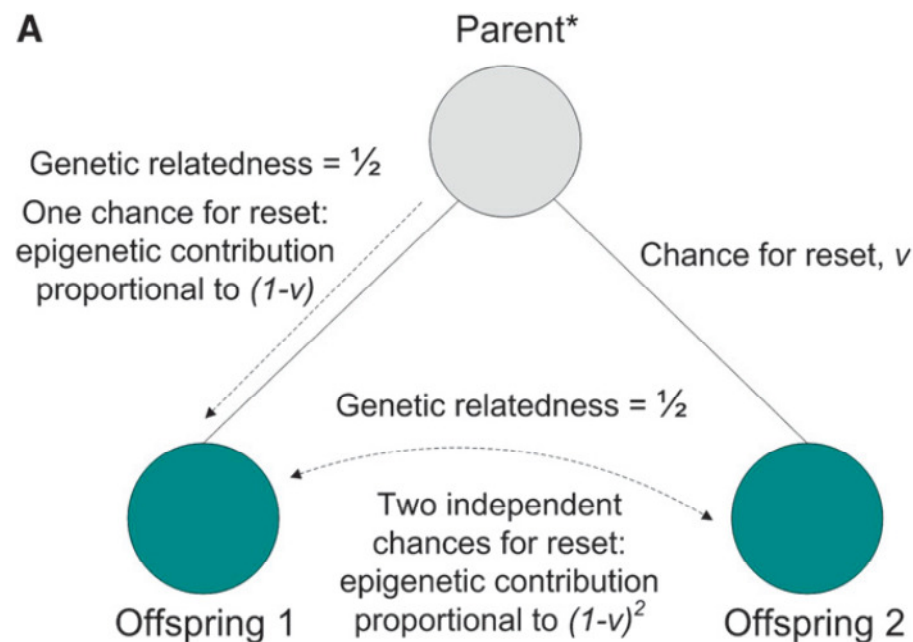
Mutational variance

$$\sigma_M^2 = 2 \sum_i u_i E[a_i^2]$$

a_i squared heterozygotic effect of mutation at locus i
 u mutation rate at locus i

Genotypic and epigenetic effects

When certain specific models for the determination of effects are assumed, epigenetic and genotypic heritability can be estimated.

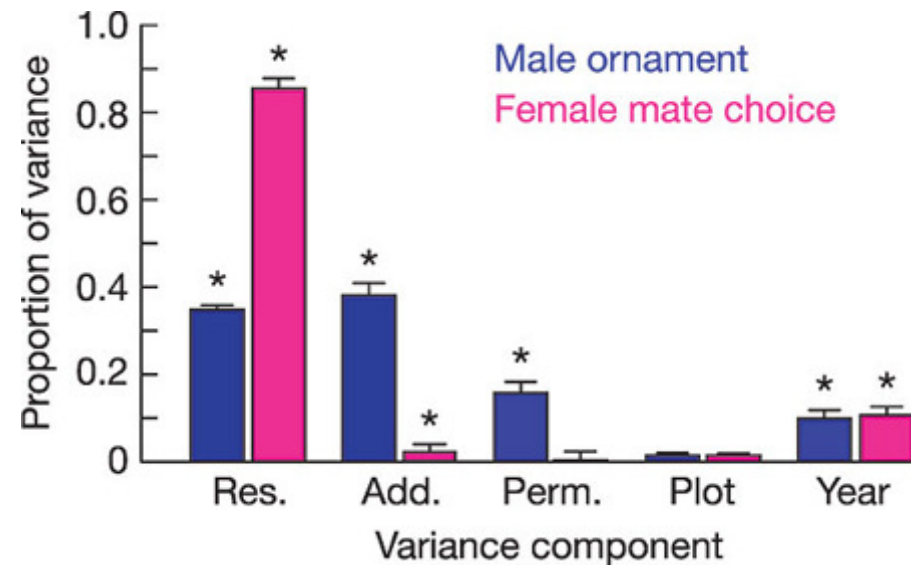


**Sexual: Showing only one of the two parents*

Tal et al. (2010) DOI: 10.1534/genetics.109.112466



"Testing the genetics underlying the co-evolution of mate choice and ornament in the wild". Nature 441: 84- (2006)



“A total of 4,220 male (ornament, blue) and 4,345 female (choice for the ornament, pink) collared flycatchers were measured. The proportion of the total variance in each trait attributed to residual (Res), additive genetic (Add), permanent environment (Perm), study plot and year were calculated using an animal model^{2,3}. “

5. Multiple Traits

Traits indexed 1 and 2 within the same individual:

$$\underline{g}_1 = \mu_1 + \underline{A}_1 + \underline{D}_1 + \underline{AA}_1 + \underline{AD}_1 + \underline{DD}_1 + \underline{AAA}_1 + \dots$$

$$\underline{g}_2 = \mu_2 + \underline{A}_2 + \underline{D}_2 + \underline{AA}_2 + \underline{AD}_2 + \underline{DD}_2 + \underline{AAA}_2 + \dots$$

$$COV(\underline{g}_1, \underline{g}_2) = COV(\underline{A}_1, \underline{A}_2) + COV(\underline{D}_1, \underline{D}_2) + COV(\underline{AA}_1, \underline{AA}_2) + COV(\underline{AD}_1, \underline{AD}_2) + \dots$$

In estimation problems we use,
for the covariance between individuals indexed by ' and ":

$$COV(\underline{g}_1', \underline{g}_2'') = \sum_{i,j=0 \dots n} \left(2\Theta_{\underline{g}_1' \underline{g}_2''} \right)^i \left(\Phi_{\underline{g}_1' \underline{g}_2''} \right)^j COV(\underline{A}^i \underline{D}^j_1, \underline{A}^i \underline{D}^j_2)$$

PS. Reaction Norms

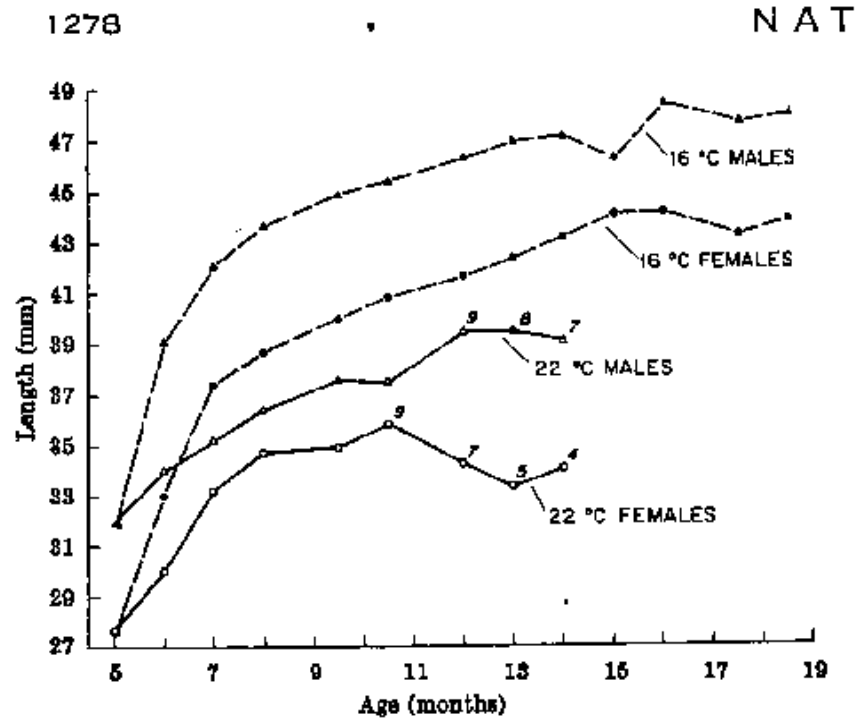


Fig. 3. Mean lengths of male and female *C. adloffi* maintained at 16° and 22° C. Each point is the mean measurement of ten fish except where indicated above the point.



Liu, R.K. and Walford, R.L.: Increased Growth and Lifespan with Lowered Ambient Temperature in the Annual Fish, *Cynolebis Adloffi*. *Nature* 212:1277-1278



G Used for multivariate covariance matrix of genotypic values

A Used for covariance matrix of additive component of genotypic values

P Multivariate covariance matrix of phenotypic values

PS. Reaction Norms



The effect of environment on phenotypic traits can be dealt with by

- using conditional traits: one trait per value of the environment
- describing the environmental dependence with a function that has a low number of parameters, these parameters are trait
- using an infinite-dimensional function-valued approach

A **function-valued trait** is a stochastic process \equiv family of random variables $z(e)$ with $e \in [a, b]$

Mean function $\mu_z(e) = E[z(e)]$

Covariance function $\sigma_z(e_i, e_j) = COV[z(e_i), z(e_j)]$ with $e_i, e_j \in [a, b]$

Covariances between traits developed in different environments.
Expectations are taken per state of the environment e .

A covariance function must be non-negative definite. For any finite set of environments e_i and any set of real numbers b_i indexed from 1 to n ,

$$\sum_{i=1}^n \sum_{j=1}^n b_i b_j \sigma_z(e_i, e_j) \geq 0$$

Variance components of function valued traits

Stochastic processes are *Gaussian* if the joint distribution of every finite set $\{z(e)\}$ is multivariate normal.

$$z(e) = \mu_z(e) + g(e) + \varepsilon(e)$$

The phenotypic covariance function can be decomposed as

$$\sigma_z(e_i, e_j) = \sigma_g(e_i, e_j) + \sigma_\varepsilon(e_i, e_j)$$

Further simplification

(A) *Character state approach*

An indexed array or vector of function values in different environments

$$\mathbf{z}^T = (z_1, z_2, \dots, z_n)^T$$

(B) *Polynomial approach*

A trait vector of coefficients of a polynomial function of e :

$$\mathbf{v}^T = (v_0, v_1, v_2, \dots, v_m)^T$$

- Correspondence between (A) and (B): $z_i = v_0 + v_1e + v_2e^2 + \dots + v_me^m$

$$\mathbf{z} = \mathbf{E}\mathbf{v}$$

$$\mathbf{E}_i = \left(1 \quad e_i \quad e_i^2 \quad \dots e_i^m \right)$$

Matrix \mathbf{E} is usually not square and thus does not have an inverse. By using a pseudoinverse one obtains the reverse transformation

$$\mathbf{v} = (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \mathbf{z} = \mathbf{U}\mathbf{z}$$

$$\mathbf{G}_z = \mathbf{E}\mathbf{G}_v\mathbf{E}^T$$

$$\mathbf{G}_v = \mathbf{U}\mathbf{G}_z\mathbf{U}^T$$

References

- Lynch, M and B Walsh. 1997. Genetics and Analysis of Quantitative Genetic Traits. Sinauer Ass.
- Bulmer, MG. The mathematical Theory of Quantitative Genetics. Clarendon Press, Oxford.
- Galton, F .1894. Natural Inheritance (5th ed.), New York: Macmillan and Company.
- Via, S et al. 1995. Adaptive phenotypic plasticity: consensus and controversy. TREE 10: 212-.