

The correlation of substitution effects across populations and across generations

A. Legarra*, C.A. Garcia-Baccino*,†, Y.C.J. Wientjes‡, , Z.G. Vitezica*

* INRAE/INP, UMR 1388 GenPhySE, 31326, Castanet-Tolosan, France.

† Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, C1417DSQ, Buenos Aires, Argentina.

‡ Wageningen University & Research, Animal Breeding and Genomics, 6700 AH Wageningen, The Netherlands.



andres.legarra@inrae.fr



bioRxiv MS: 367227



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 772787



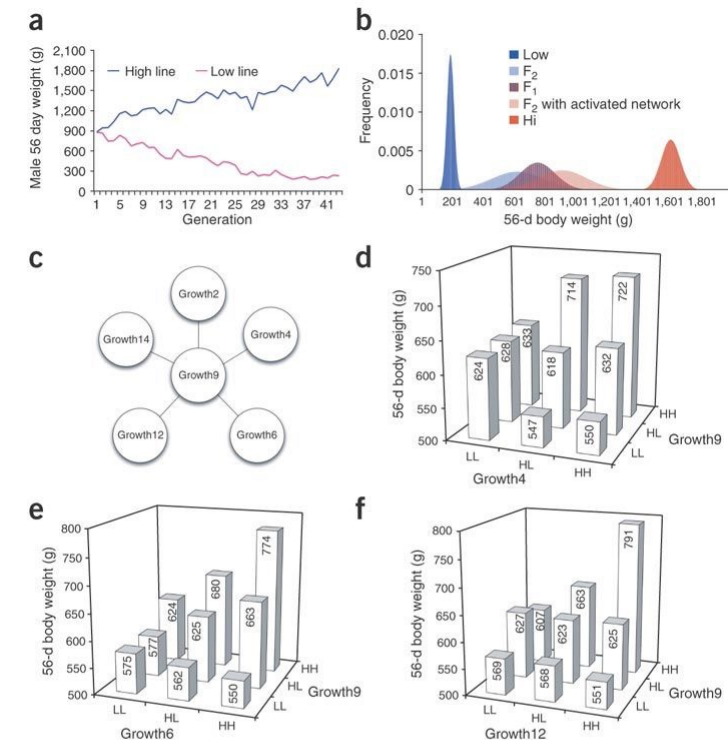
Empirical facts

- Marker estimates from one population do not predict another one
- Increasing marker density (700K, sequence) helps very little
- Gene annotation helps very little
- A hypothesis is that not only marker effects, but QTL effects are not identical across populations



Substitution effects α

- α is the basis of the Breeding Values
- However α is not the “biological” effect of genes
- α is the statistical (marginal) effect of the locus
- Genes and alleles interact (and it is hard to know what happens)
- α will differ across populations because of interactions through different allele frequencies



Carlborg, Örjan, et al. "Epistasis and the release of genetic variation during long-term selection." *Nature genetics* 38.4 (2006): 418.



Substitution effects

If we knew the “biological” effects, we could derive α . For instance:

- Biological additive and dominance effects only:

$$\alpha_i = a_i + (1 - 2p_i)d_i$$

- Biological additive, dominance and additive x additive gene effects:

$$\alpha_i = a_i + (1 - 2p_i)d_i + (2p_j - 1)[aa]_{ij}$$

- Biological multiplicative effects:

- $\alpha_i = \prod_{j \neq i} 2p_j - 1$

But we don't know biological networks and effects



Question

α_i depends on allele frequencies at locus i
and at other loci $j \neq i$

For populations b and b' we will have
different $\alpha_i^b, \alpha_i^{b'}$

Are $\alpha_i^b, \alpha_i^{b'}$ similar?

Answer?

Two populations are similar if they have
similar allele frequencies

Relationship across populations b and b'
(Robertson 1977, Legarra et al 2015):

$$\gamma_{b,b'} = 8Cov(p_i^b, p_j^{b'})$$



Objective

- Obtain similarity of substitution effects
 - actually the correlation $r(\alpha_i^b, \alpha_i^{b'})$
- in absence of data
- as a function of similarity of populations
- We assume the same genetic architecture (possibly very complex) in the different populations
- We assume no GxE which comes on top of this



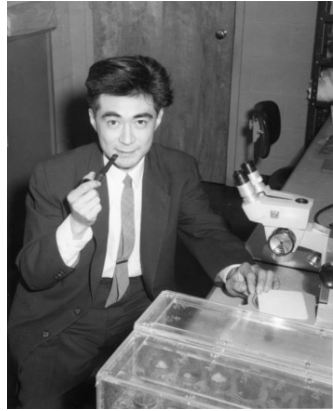
Idea



- We don't know α_i^b and $\alpha_i^{b'}$
- But we can approximate $\alpha_i^{b'}$ and α_i^b from α_i^0 in a population “0” using Taylor series expansions
- This approximation is universal: it holds for any form of epistasis/dominance



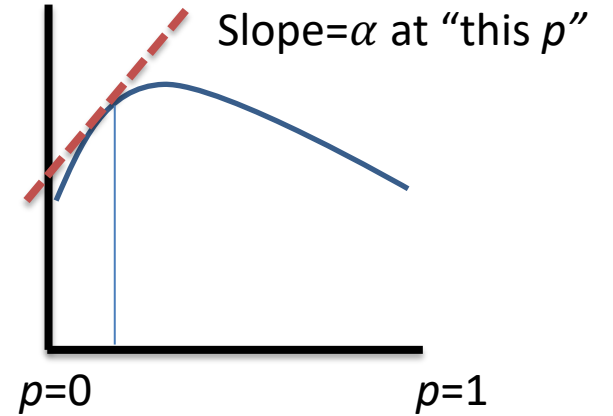
General definition of α for any genetic architecture



Kojima (1959, 1961): derivative of the mean of the population w.r.t. the allele frequency

$$\alpha_i = \frac{1}{2} \frac{\partial \mu}{\partial p_i}$$

interpretation: if I increase the frequency of the reference allele, does the population improve/deteriorate?



8

Analogous definitions for dominance and epistasis. Interestingly:

$$d_i^* = -\frac{1}{4} \frac{\partial \mu}{\partial p_i^2} = -\frac{1}{2} \frac{\partial \alpha_i}{\partial p_i}$$

$$[\alpha\alpha]_{ij} = \frac{1}{4} \frac{\partial \mu}{\partial p_i \partial p_j} = -\frac{1}{2} \frac{\partial \alpha_i}{\partial p_j}$$



Some gory details...

$$\epsilon^{(b)} = p^{(b)} - p_0, \epsilon^{(b')} = p^{(b')} - p_0$$

$$\alpha_i(p) \approx \alpha_i(p_0) + \nabla_i' \epsilon + \left(\frac{1}{2} \epsilon' \mathcal{H}_i \epsilon \dots \right)$$

$$\nabla_i = \begin{pmatrix} \frac{\partial \alpha_i}{\partial p_1} | (p_i = p_i^0) \\ \dots \\ \frac{\partial \alpha_i}{\partial p_i} | (p_i = p_i^0) \\ \dots \\ \frac{\partial \alpha_i}{\partial p_n} | (p_i = p_i^0) \end{pmatrix} = 2 \begin{pmatrix} (\alpha\alpha)_{i1}^0 \\ \dots \\ -d_i^{*0} \\ \dots \\ (\alpha\alpha)_{in}^0 \end{pmatrix}$$

$$\alpha_i^b \approx \alpha_i^0 + 2\epsilon_i^b(-d_i^{*0}) + 2\epsilon^{(b)'}(\alpha\alpha)_i^0$$

$$\alpha_i^{b'} \approx \alpha_i^0 + 2\epsilon_i^{b'}(-d_i^{*0}) + 2\epsilon^{(b')'}(\alpha\alpha)_i^0$$

$$\text{Cov}(\alpha_i^{(b)}, \alpha_i^{(b')}) = \text{Var}(\alpha_i^0) + 4\text{Cov}(\epsilon_i^{(b)}, \epsilon_i^{(b')})\text{Var}(d_i^{*0}) + 4\text{tr}\left(\text{Cov}(\epsilon^{(b)}, \epsilon^{(b')'})\text{Var}((\alpha\alpha)_i^0)\right)$$

Change in allele frequencies
from pop 0 to pop b

Keep 2 first terms of Taylor series

By Kojima's method

additive effect in population b is a function of:

- additive + dominant + epistatic statistical effects in population 0
- change in allele frequencies

the covariance is a function of
similarity of allele frequencies
and respective variances

The main expression (i)

The similarity of substitution effects

Similarity of allele frequencies

$$\begin{aligned} & Cov(\alpha_i^{(b)}, \alpha_i^{(b')}) \\ &= Var(\alpha_i^0) + 4Cov(\epsilon_i^{(b)}, \epsilon_i^{(b')})Var(d_i^{*0}) + 4tr\left(Cov(\epsilon^{(b)}, \epsilon^{(b')})Var((\alpha\alpha)_i^0)\right) \end{aligned}$$

Additive variance at QTL level

Dominance variance at QTL level

Additive by additive variance at QTL level

The main expression (ii)

$$r(\alpha_i^b, \alpha_i^{b'}) = \frac{\sqrt{\sigma_A^2}}{\sqrt{\sigma_A^2 + (\gamma_{b'} + \gamma_b - 2\gamma_{b,b'}) \left(\frac{1}{2} \frac{\bar{H}}{H^2} \sigma_D^2 + 2 \frac{\sigma_{AA}^2}{H} \right)}}$$

The similarity of substitution effects

Additive variance at animal level

Function of relationship across populations (similar to Fst)

Heterozygosities

Dominance variance at animal level

Additive by additive variance at animal level

In practice

- I have one population (say Texel)
- I want to estimate $r(\alpha_i^b, \alpha_i^{b'})$ Texel – Merino. I need
 - Estimates of additive, dominance and additive x additive variance $\sigma_A^2, \sigma_D^2, \sigma_{AA}^2$
 - From markers: relationship within and across populations:

$$\Gamma = \begin{pmatrix} \gamma^{(Texel, Texel)} & \gamma^{(Texel, Merino)} \\ \gamma^{(Merino, Texel)} & \gamma^{(Merino, Merino)} \end{pmatrix}$$

- From markers (or guesses): average heterozygosities \bar{H} and \bar{H}^2

$$\text{• Use then } r(\alpha_i^b, \alpha_i^{b'}) = \frac{\sqrt{\sigma_A^2}}{\sqrt{\sigma_A^2 + (\gamma_{b'} + \gamma_b - 2\gamma_{b,b'}) \left(\frac{1}{2H^2} \sigma_D^2 + 2 \frac{\sigma_{AA}^2}{\bar{H}} \right)}}$$

Example: across-breeds

- $\sigma_A^2 = 0.092$; $\sigma_D^2 = 0.020$; $\sigma_{AA}^2 = 0.016$ in pigs (Vitezica et al. 2018)
- $\Gamma = \begin{pmatrix} 0.756 & 0.259 \\ 0.259 & 0.730 \end{pmatrix}$ for Landrace and Yorkshire (Xiang et al. 2017)
- This gives Landrace - Yorkshire: $r(\alpha_i^b, \alpha_i^{b'}) \approx 0.6$
- Robust to assumptions on QTL allele frequency spectra
- Similar values are obtained for Jersey-Holstein
- These values are similar to actual estimates in literature



Example: within breed

- The relationship across generations can be measured by the increase in inbreeding; assuming $\Delta F = 0.01$ per generation
- Dairy cattle: $\sigma_A^2 = 0.2$; $\sigma_D^2 = 0.09$; $\sigma_{AA}^2 = 0.15$ (Fuerst and Solkner, 1994)
- Good agreement with empirical results

Number of generations away	$r(\alpha_i^b, \alpha_i^{b'})$
1	0.98
2	0.97
5	0.92
10	0.85

Conclusion

- The method gives good ballpark estimate
- Correlation across breeds is ~ 0.6
- Correlation across generations is from ~ 1 for close to ~ 0.8 for distant ones
- Explaining differences in QTL substitution effects:
 - distance of populations
 - dominance variance is not relevant due to the weighting of heterozygosities
 - additive by additive variance is more relevant
 - (typically, $\sigma_{AA}^2 \sim 10\%-50\%$ of additive variance)
 - agreement with simulations (Dai et al. 2020; Duenk et al. 2020)
- Instead of pretending that substitution QTL effects are the same across populations, we can check it



Part of this work was while Y. Wientjes was visiting the GenPhySE unit at INRAE, Toulouse, financed by the Netherlands Organisation of Scientific Research (NWO).

SMARTER PARTNERS



Thank you for your attention

www.smarterproject.eu



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 772787

