

Variance and Covariance of Actual Relationships between Relatives at One Locus

2013 PLoS ONE 8(2):
e57003. doi:10.1371/journal.pone.0057003



andres.legarra@toulouse.inra.fr

Luis Alberto Garcia-Cortés, Andrés Legarra,
Claude Chevalet, Miguel Angel Toro,



EAAP 2013

Why

Realized relationships = expected relationships + deviations

$$\mathbf{G} = \mathbf{A} + \mathbf{E}$$

Estimated with markers (Van Raden 2008 and many others)

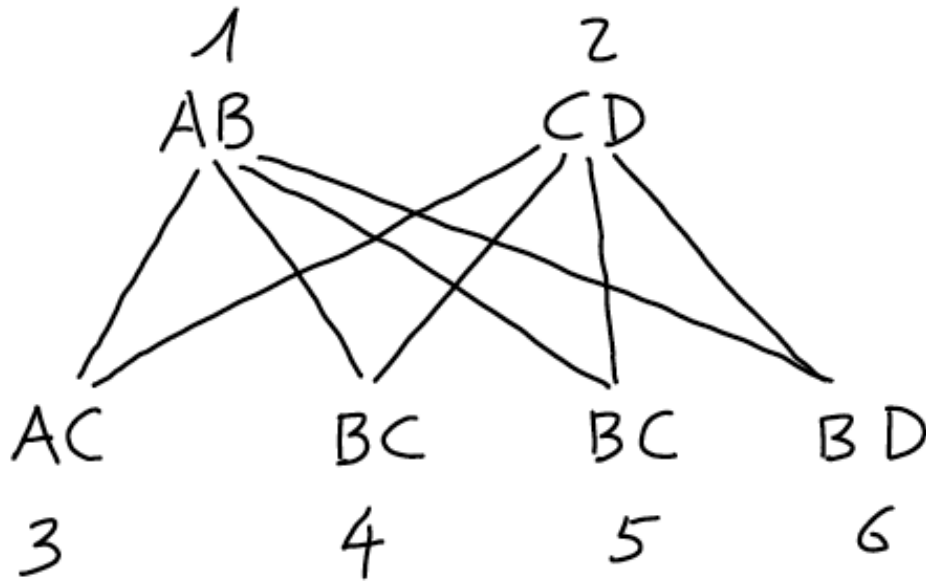
Estimated with pedigree (Emik & Terrill 1949 and many others)

????

Why

- Increased accuracy of genomic selection relies on using a more accurate **G** instead of **A**
 - But we don't know *a priori* how much will they differ
- Errors in **A** cumulate with generations !!
 - How many generations should one use?

$$r_{ij} = \frac{1}{4} [\Pr(P_i \equiv P_j) + \Pr(P_i \equiv M_j) + \Pr(M_i \equiv P_j) + \Pr(M_i \equiv M_j)]$$



$$r_{34} = 1/4$$

$$r_{35} = 1/4$$

$$r_{36} = 0$$

$$r_{45} = 1/2$$

$$r_{13} = r_{14} = r_{23} = r_{24} = \dots$$

Some sibs are more equal than others !!

Variation in actual relationship as a consequence of Mendelian sampling and linkage

W.G. HILL^{1*} AND B.S. WEIR²

- Only consider two-individual coancestries and not a full pedigree
- Not a general formula but a collection of them
- They consider linkage, we consider one locus
- We provide a general formula for all pairs of individuals in a pedigree

What

Realized relationships = expected relationships + deviations

$$\mathbf{G} = \mathbf{A} + \mathbf{E}$$

What is $Var(\mathbf{E})$?

Let $\mathbf{E} = (\mathbf{E}_{:,1} \quad \mathbf{E}_{:,2} \quad \dots \quad \mathbf{E}_{:,n})$

What is $Cov \begin{pmatrix} \mathbf{E}_{:,1} \\ \mathbf{E}_{:,2} \\ \dots \\ \mathbf{E}_{:,n} \end{pmatrix}$?

$\Delta_i =$ Probability of identity state S_i

Father-son

$$\Delta_8 = 1$$

Full sibs

$$\Delta_7 = 1/4$$

$$\Delta_8 = 1/2$$

$$\Delta_9 = 1/4$$

K. LANGE AND J. S. SINSHEIMER

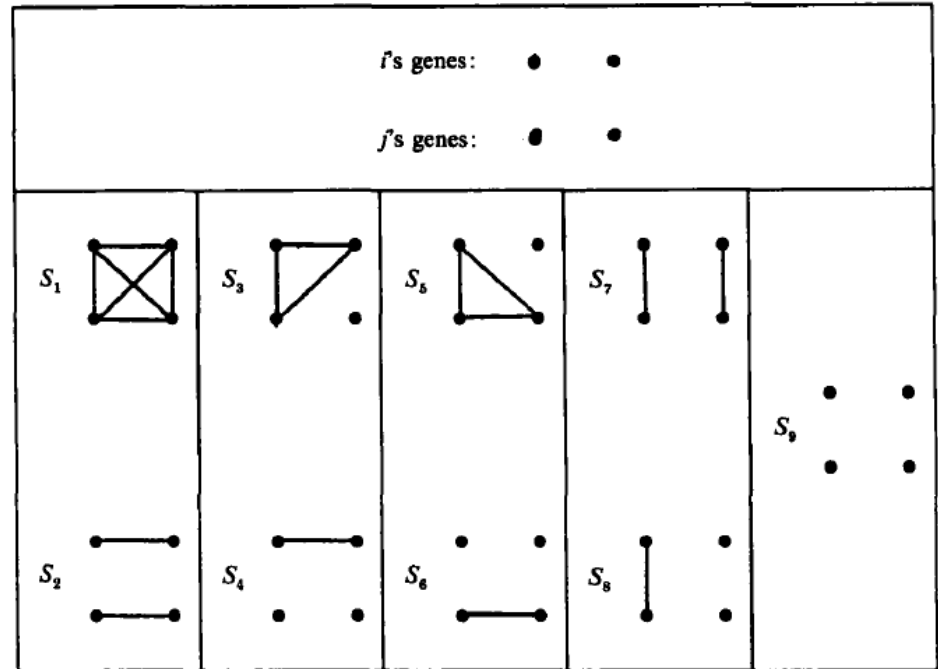


Fig. 1. Condensed identity states.

Coancestry ($= 0.5a_{ij}$) of i and j

$$\phi_{ij} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8$$

$$E(i) = \Delta_i$$

$$\text{Var}(i) = \Delta_i (1 - \Delta_i)$$

$$\text{Cov}(i, j) = -\Delta_i \Delta_j$$

$$\phi_{ij} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8$$

$$\phi_{ij} = \left(1 \quad 0 \quad \frac{1}{2} \quad 0 \quad \frac{1}{2} \quad 0 \quad \frac{1}{2} \quad \frac{1}{4} \quad 0 \right) \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_9 \end{pmatrix}$$

$$\text{Var}(\phi_{ij}) = \text{Var}(W' \Delta)$$

$$= W' \text{Var}(\Delta) W$$

$$= \sum_i^9 w_i^2 \Delta_i (1 - \Delta_i) - 2 \sum_j \sum_{i \neq j} w_i w_j \Delta_i \Delta_j$$

$$\text{Var}(\phi_{ij}) = \sum_i w_i^2 \Delta_i - \left(\sum_i w_i \Delta_i \right)^2$$

$$= \phi_{ij, ij} - (\phi_{ij})^2$$

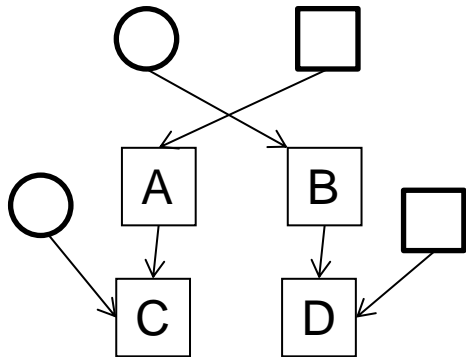
→ Karigl (1981) generalized relationship

Covariance

- The covariance is a simple extension:

$$\text{Cov}(\Phi_{AB}, \Phi_{CD}) = \phi_{AB,CD} - \phi_{AB}\phi_{CD}$$

- This means that the Φ_{CD} is not independent of Φ_{AB} ,
example:



C and D can not share alleles if A and B do not

Generalized relationships

Karigl's 1981 generalized relationships

$$\phi_{AB,CD}$$

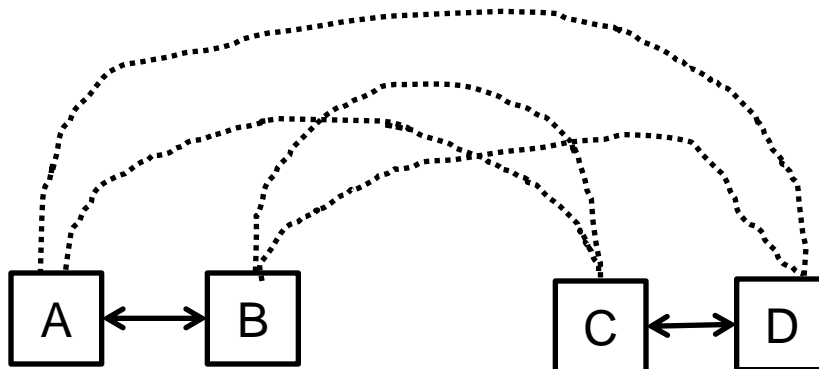
is the probability that

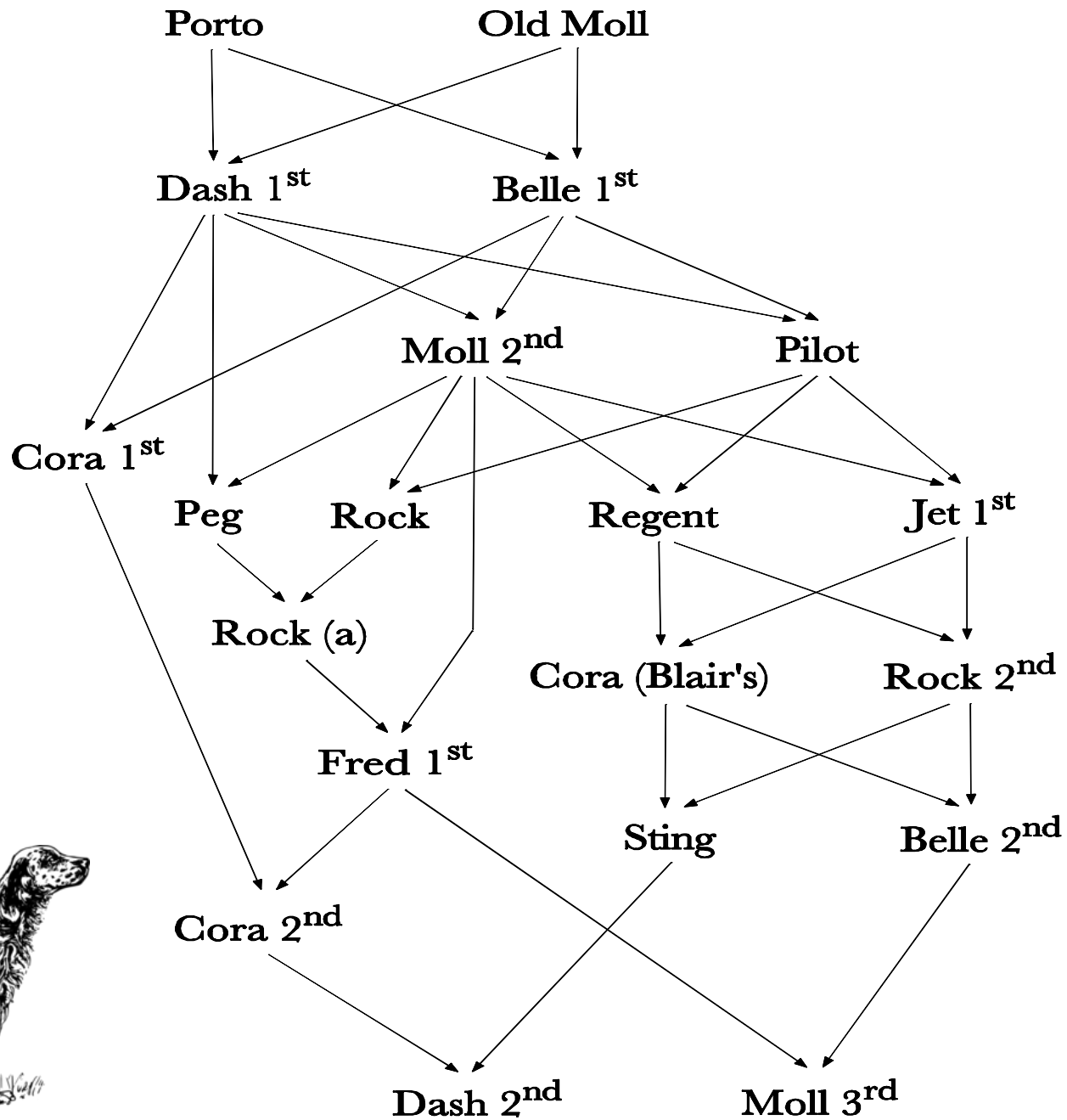
two genes drawn from A and B are IBD

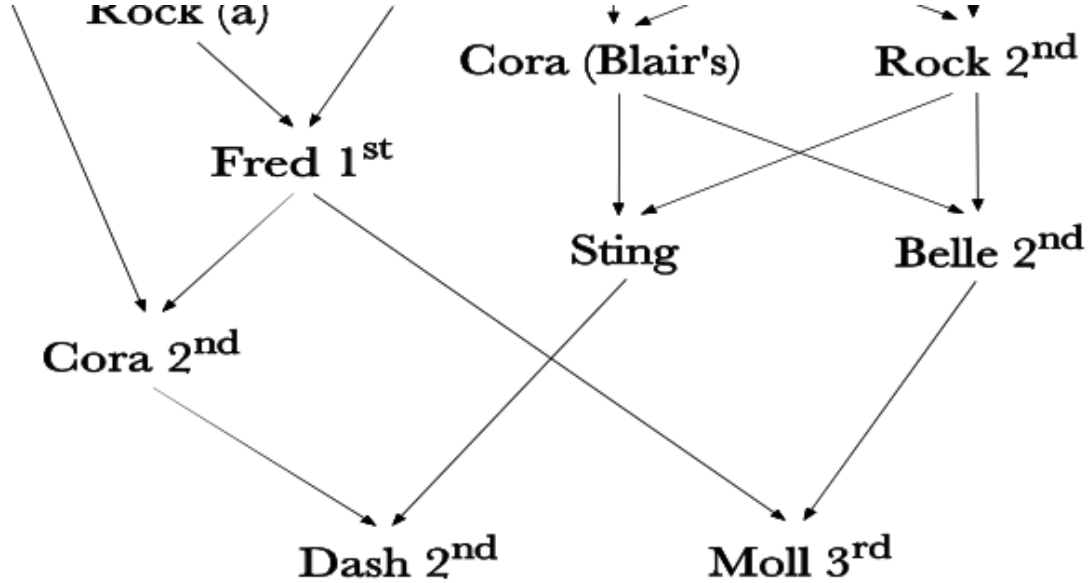
and at the same time

two genes drawn from C and D are IBD

This will depend on the paths linking all four individuals







Inbreeding of Dash 2nd

$$F_D = 0,4297$$

$$V(F_D) = 0,2451$$

$$\phi_{CS} = 0,4297$$

$$V(\phi_{CS}) = 0,0922$$

Coancestry of parents of Dash 2nd

$F_D = \phi_{CS}$, yet the variances are different !!

Extension to several loci

- If there were infinite loci, these variances are divided by ∞ and become 0
- Genome is not infinite, is composed of repetitive “chunks” due to linkage and finite number of animals
- Divide by « M_e = equivalent number of loci » (Goddard 2009 and many others)

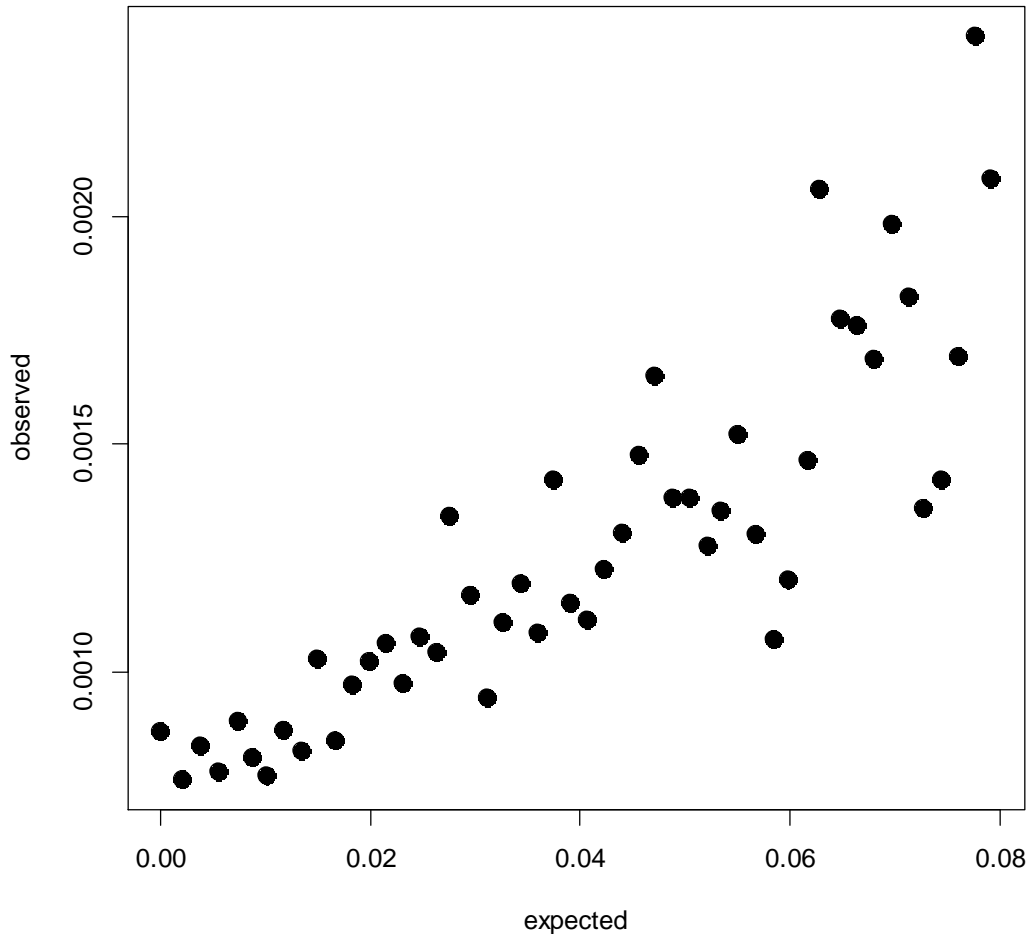
$$\text{Cov}(\Phi_{AB}, \Phi_{CD}) = \frac{1}{M_e} (\phi_{AB,CD} - \phi_{AB}\phi_{CD})$$

Does it work?

- H. Wang, I. Misztal et al., in prep, chicken data, >4000 animals
- **$E=G-A$**
- empirical var(**E**) vs. predictions based on pedigree

Var(**G-A**) vs. our theory

observed vs. expected variance



- Pretty good fit
- From here, number of effective loci ~ 70

Acknowledgments

- project CGL2009-1327-C02-02 of the Ministerio de Ciencia e Innovacion (Spain)
- ANR project Rules & Tools (France)
- action X-Gen of the SelGen metaprogram (INRA)
- partially supported by the platform bioinformatics Toulouse Midi-Pyrenees
- *LAGC made the drawings*